Title
The Solvent Mediated Thermodynamics of Cellulose Deconstruction

Permalink
https://escholarship.org/uc/item/98w8g5ww

Author
Gross, Adam S

Publication Date
2012

Peer reviewed|Thesis/dissertation

The Solvent Mediated Thermodynamics of Cellulose Deconstruction

By

Adam S Gross

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jhih-Wei Chu, Chair
Professor Berend Smit
Professor Phillip Geissler

Fall 2012

Abstract

The Solvent Mediated Thermodynamics of Cellulose Deconstruction

by

Adam S Gross

Doctor of Philosophy in Chemical Engineering

University of California, Berkeley

Professor Jhih-Wei Chu, Chair

Cellulosic biomass has the potential to be used as a sustainable feedstock for the production of liquid transportation fuels and other chemicals. However, this material also possesses an extreme resistance to structural and chemical degradation, known as recalcitrance, which currently prevents it from being used for this purpose. Therefore, an understanding of the origins of this recalcitrance and how to overcome it will be useful in the design of processes and technologies for the production of biofuels. To address these issues, we have performed all-atom molecular dynamics (MD) simulations of cellulose in various conditions and solvents to gain a molecular-level understanding of the forces and interactions that give rise to the macroscopic behaviors related to cellulose recalcitrance and how it is overcome.

Cellulose is a homopolymer of β-glucose connected by β-1,4 bonds found in the plant cell wall in the form of long, semicrystalline fibers called microfibrils. Within the two naturally occurring crystal structures of cellulose, I_β and I_α, glucose chains are arranged into flat sheets that are then stacked up upon each other, the polymerization axis of the chains aligned along the long axis of the microfibril. To better understand the forces that bind the chains of a microfibril together, we divided the interaction network of cellulose into three categories, which we then analyzed. These categories are:

1. Intrachain interactions: between neighboring glucoses within the same polymer chain.
2. Interchain interactions: between neighboring glucoses of different chains within the same sheet
3. Intersheet interactions: between neighboring glucoses of different chains in different sheets

The dominant intermolecular forces in the intrachain and interchain directions are OH—O hydrogen bonds (HBs), while in the intersheet direction, it is CH—O contacts (pseudo HBs) as well as van der Waals (vdW) interactions. We have examined the behavior of these three groups, and found that intersheet forces are most responsible for cellulose recalcitrance. The HBs of the intrachain and interchain interactions are severely disrupted at the microfibril surface by solvent exposure, seen by both HB geometries and HB occupancies. Conversely, by the same metrics, the intersheet CH—O HBs are robust and not weakened to any appreciable extent when at the microfibril surface. Counting up the amount of HBs that exist for each of the three interaction groups, we find that the number

1

of intersheet HBs is far greater than the number of intrachain or interchain HBs. Also, energetically, sheet-sheet interactions are stronger than chain-chain interactions. In the interior of a microfibril, intersheet forces are ~1 kcal/mol-glucose stronger than interchain forces, and at the microfibril surface, the interchain forces weaken considerably, up to 50 %, whereas intersheet forces are only slightly affected.

Solvent water density behavior around the microfibrils is an important aspect of cellulose degradation, both for surface characterization, but more importantly, for its implications in the access of enzymes or chemical catalysts to cellulose surfaces. Cellulose imparts a specific, long-range structure into the surrounding water molecules that solvate it. Because of the amphiphilic nature of glucose, its water hydration layers exhibit the effects of both short and long-length scale hydrophobicity, the CH-presenting cellulose surfaces being the most hydrophobic. At short range, the HB network of water percolates around these CH groups while connecting with the glucose OH groups, while at long-range, the hydration shells of CH-presenting surfaces show the greatest amount of compressibility.

Ionic liquids (ILs), salts molten at or near room temperature, are one of the few solvents that are able to overcome the cellulose interaction network and dissolve the material. However, molecular and thermodynamic knowledge of how these liquids accomplish this is poor. By performing simulations of cellulose in a microfibril state and in a dissociated, dissolved state in a prototypical IL, 1-butyl-3-methylimidazolium (Bmim) chloride (Cl), and contrasting that behavior with the same in water, we are able to explain the solvent abilities of ILs with cellulose. Thermodynamically, the energy of dissolution is favorable in BmimCl but neutral to unfavorable in water, depending on temperature. From analysis of the three-dimensional solvent density distributions around the dissolved chains, chloride anions interact with the equatorial hydrophilic OH groups, forming strong HBs with them, while the cations preferentially solvate from the axial direction of the glucoses, where the hydrophobic CH groups are present, and have favorable contacts with the linker oxygen of the β-1,4 bond as well. Thus, BmimCl can satisfy the amphiphilic nature of glucose. This is in contrast to water, which cannot. In water, the energy of dissolution is not favorable because while water can form HBs with glucose's OH groups, interactions of water with its CH groups are ultimately undesirable.

Entropy also plays an important role in determining the thermodynamics of cellulose dissolution. Because of the long-range, highly ordered solvation structures that form around the sugar chains, solvent entropy favors the undissolved state of cellulose. Cellulose entropy, however, favors the dissolved state. The magnitudes of both of these preferences are less in BmimCl than in water because of the former's longer-range solvent-solvent forces and resulting liquid state structure. Summing the two entropic contributions together, the total entropy of dissolution is positive at all conditions and solvents tested except for in water at room temperature. Thus, in general, entropy changes favor cellulose dissolution.

Another solvent able to dissolve cellulose is lithium chloride (LiCl) in N,N-dimethylacetamide (DMA). Although the mechanism through which this is achieved is generally thought to be direct interaction of dissolved salt with cellulose, the specifics of how this is accomplished, and the relative contributions of the various components of the LiCl/DMA system to dissolving cellulose, are poorly understood. To address these issues, we performed MD simulations and free energy calculations related to the deconstruction

of cellulose in LiCl/DMA, as well as in pure DMA, LiCl/water, and pure water. Calculation of the potential of mean force (PMF) of the deconstruction of a cellulose microfibril confirm that LiCl/DMA is a cellulose solvent, while DMA, LiCl/water, and water are not. Analysis of simulations of dissolved and undissolved cellulose in LiCl/DMA and LiCl/water reveals that solvent-mediated preferential interactions of dissolved ions with sugars are responsible for cellulose dissolution. By computing the three-dimensional density of ions around dissolved cellulose chains, we find that the localization of salt by the sugars is two orders of magnitude greater in DMA than in water. The ions near the sugars are able to disrupt cellulose's internal interaction network by forming $O-Li^+$ and $OH-Cl^-$ interactions, causing dissolution. Because DMA is a poor solvent for both LiCl and cellulose, these two species preferentially interact with each other to the exclusion of DMA, whereas since water is a good solvent for LiCl but a poor one for cellulose, no salt-sugar preferential interactions occur in it. We also find that opposite the case in water, in DMA, the high-density regions of Li^+ cations are nearer to the cellulose atoms than those of the Cl^- anions. Quantification of the salt-sugar interactions via a coarse-graining force-matching analysis confirm that although they both are necessary, attractive sugar-Li^+ interactions, and not sugar-Cl^- interactions, are the most important contributor to the disruption of cellulose's interaction network, and hence, to cellulose dissolution.

Table of Contents

List of Tables

List of Figures

Acknowledgements

It will be difficult to retrace the steps that led me here, to the final days of my graduate school career. It seems to me that my path was created by a series of small decisions, one after the other, built up over time. Now, to finally reach this point of finishing my dissertation, I am amazed that I've come so far. But I know that to get here, I have needed the help, guidance, and support of many, to whom I will always be thankful.

The man most responsible for getting me to graduate school is the supervisor I had at my old internship at Infineum USA L.P., Dr. David Martella. He sparked my interest in scientific research and opened my mind up to the idea of going to graduate school. I'm not sure I would have tried for a Ph.D. had I not met him. Thanks also to my undergraduate research advisor at Cornell, Prof. Elizabeth Fisher. I learned a lot from her on how academic research is conducted. Others to thank from my time as an undergraduate would be my friends. First, my three roommates, Steve Hohwald, Alec Berntson, and Albert Ren, who I lived with almost my entire time at Cornell. We definitely had some fun times together that made those years in Ithaca all the better. I must also mention the other chemical engineers who slogged through four years of hard work with me, Benny Leung, Justin Pike, Erica Schlesinger, Parag Gupta, Adrian Nino de Rivera, Sparsh Bhargava, Bryan Klosiewicz, Danielle Green, and many, many more. I think our senior year we spent more time in the basement of Olin Hall than we did in our own apartments. As it turns out, maybe it was worth it.

And my time here at Berkeley has been enjoyable as well, no doubt thanks to the many friendships I've made. I have to thank the many roommates I've had over the years, Colin Cerretani, Sam Maurer, Joe Swisher and Georges Siddiqi, and Brett Loveless, Dana Nadler, and Wolf Rendall. They always made my home life fun, even when my life in lab was not. In general, I'm lucky that I had a great group of friends here that I really enjoyed being around, who made sure I always had more to do than just lab work, people like Rob Carr, Bryan McCulloch, Joanna MacKay, Keith Beers, Megan Hoarfrost, Sean and Maureen Dee, Ted Amundsen, Bean Getsoian, Maureen Tang, Ting-Ying Chung, and the rest of my class. We made it through a lot, and I was glad we could do it together. And there's the many other friends I've made here too, Will Vining, Raj Gounder, Jarred Ghilarducci, Penny Gunterman, Andrew and Kristen Jones, Will Knaeble, Steph Paquin, Josef Macht, Sebastian Kunz, Michele Sarazen, Anton Mlinar, Frederick de Meyer, Eric and Amy Sacia, and too many more to name. They all made my life here in Berkeley a lot better.

And of course, I must thank my labmates, without whom I never would have been able to graduate. First, my old office mate Dr. Hyung Min Cho, who taught me the practical side of molecular simulations and computational science. How he managed to be so patient with me and humor my questioning in those early days, I do not know. The same can be said of Jason Brokaw, who was always ready to help me when I had problems. Thanks to Barry Shang, Dr. Nikolaos Voulgarakis, and Dr. Yuchun Lin for all the help they gave me as well. And thanks also to Kevin Haas, who gave me useful suggestions, and always listened to my harebrained research ideas whenever I had them. Ritankar Das, my undergraduate researcher, was always enthusiastic and inquisitive. I also need to mention all the support staff, who made everything run as smoothly as

It certainly has been an interesting five-and-a-half years in Berkeley. But what made that time so special to me was not the research I performed, but the friendships I made and the relationships I built or made stronger in my time here. Looking back on it now, I feel truly amazed at all the wonderful people I've met, and I realize how lucky I am to have done so.

Chapter 1

Introduction

Lignocellulose is the only sustainable source of organic carbon large enough to supplement petroleum to any appreciable extent as a feedstock for the production of liquid transportation fuels.[1] The US alone can sustainably produce 1.3×10^9 metric tons of dry biomass per year, equivalent to 3.8×10^9 barrels of oil energy equivalent[2,3] (for comparison, the US consumed 6.5×10^9 barrels of oil in 2010[4]). In order to harness the energy contained within biomass, it must first be broken down into its constituent parts, and then converted into fuel products or other chemicals. Combined commercial viability of these processes is still lacking however, due to biomass's strong resistance to physical and chemical degradation, collectively known as recalcitrance.[5,6] Therefore, pretreatment of biomass prior to its break down and conversion to separate and disaggregate its components is essential to improve the efficiency and decrease the cost of any lignocellulose conversion scheme.[6-9] While inquiry into lignocellulose pretreatment has a long and storied history spanning over a century, understanding of the physics necessary for these processes to be successful is incomplete, and all industrial applications of non-degrading pretreatments have had limited, if any, success.[10]

Lignocellulose, the material of the plant cell wall, consists primarily of three components.[11] Cellulose, the most abundant (35-50 wt %), is a linear homopolymer of β-glucose connected by β-1,4 linkages, and will be the focus of this work. The other two are hemicellulose (25-30 wt %), an amorphous branched heteropolysaccharide, and lignin (15 to 30 wt %), an aromatic polymer of monolignols.[10] Even by itself, cellulose is difficult to degrade. In nature, it exists in the form of long crystalline to semi-crystalline fibers called microfibrils.[12] Typical dimensions of these microfibrils are 20-50 Å in diameter (corresponding to ~36 polymer chains in cross section),[13] with a degree of polymerization of the chains anywhere from the hundreds to the tens of thousands. We show an example microfibril in Figure 1.1. Naturally occurring cellulose, collectively known as cellulose I, has two allomorphs, cellulose I_β and cellulose I_α.[14] Of the two, I_β is more stable, as heating of cellulose I_α will irreversibly transform it into I_β.[15] The structures of the two allomorphs have only been resolved within the last decade.[16,17] Both consist of glucan chains aligned equatorially in parallel to form flat sheets, which are then stacked axially upon each other to form the full microfibril (see Figure 1.1).

Because of both the structure of the individual β-glucoses, and the way these monomers are arranged within crystalline cellulose, a complex, strong, highly redundant interaction network exists within the microfibrils that bind the polymer chains together and render the material highly resistant to degradation of any kind.[5] This network consists of OH—O hydrogen bonds (HBs), CH—O pseudo HBs,[18-20] and van der Waals (vdW) contacts. Interactions within a single chain (intrachain) and between chains within a single sheet (interchain) are dominated by OH—O HBs, while those between chains in different sheets (intersheet) are composed of both CH—O HBs and vdW interactions. No intersheet OH—O HBs exist within either of the cellulose I crystal structures.[16,17] Previous work on the interaction network, both experimental[21-25] and simulation,[26-28] has focused almost entirely on the intrachain and interchain OH—O HBs, while the intersheet forces have received scant attention. Although cellulose's OH—O HBs are an

important part of its interaction network, and have vast implications on the material's stability, their comparative strength to the intersheet CH—O HBs and vdW forces are unclear.

Using all-atom molecular dynamics (MD) simulations of cellulose I_β and I_α microfibrils, we have investigated the behavior of the intrachain, interchain, and intersheet components of cellulose's interaction network to determine which are most responsible for cellulose recalcitrance, which we detail in chapter 2 of this work. MD simulations are especially adept at this task because they give direct and complete access to molecular level information, which can then be used to connect back with macroscopic behavior. We have identified the HBs, both OH—O and CH—O, that exist within the crystal structures of cellulose I, and calculate both their average occupancies over the MD trajectories, and the change in these occupancies throughout the microfibril cross section as a function of solvent exposure. We also enumerate the number of HBs that exists for each of three directional interaction classes to determine the redundancies of each type. In addition, we quantify their strengths via interaction energy calculations for the interchain and intersheet forces, and examine how they depend on proximity to solvent. The ability of interactions to remain stable at the surface or in the presence of solvent is key, as those interactions that are not disrupted will be most responsible for recalcitrance. Finally, we quantify the solvent density and its fluctuations around the microfibrils. Knowledge of solvent density profiles and their compressibility is important as these phenomena have ramifications on the access of enzymes or chemical catalysts to the different cellulose surfaces.

Cellulose pretreatment can speed up conversion processes by disrupting the material's intrinsic interaction network. Cellulose dissolution is one example of a pretreatment process. Unfortunately, cellulose solvents are uncommon.[10] The material does not dissolve in water, organics, and other typical solvents. However, recent work has shown that certain ionic liquids (ILs), salts molten at or slightly above room temperature, are able to dissolve cellulose up to 25 wt %.[29-31] Typical cellulose-dissolving ILs consist of an alkyl-functionalized heterocyclic ring as the cation (e.g. imidazolium, pyridinium), and a small, conjugate base anion (e.g. chloride, acetate).[32,33] The mechanisms by which ILs dissolve cellulose are not completely understood. From both NMR relaxation experiments of model carbohydrates in a prototypical IL, 1-butyl-3-methylimidazolium (Bmim) chloride (Cl),[34,35] as well as MD simulations of model carbohydrates in IL,[36-39] the role of the anion is to form strong HBs with the sugar OH groups to replace the ones lost during disintegration of the crystal structure. Consequently, a strong dependence of solubility on anion identity has been observed.[29,32] The role of the cation is far less clear. No definite cation-sugar interactions have been identified, so its role in the dissolution process has been considered nonspecific or indirect.[34,35,38] However, cation composition and structure does affect cellulose solubility.[29,33] Macroscopically, the thermodynamics of the dissolution process are also unknown.

To investigate the solvent abilities of ILs towards cellulose, in chapter 3 we show the molecular ordering of BmimCl around dissolved cellulose calculated from MD simulations, and how this ordering relates to specific molecular sugar-IL interactions. This behavior is contrasted with the same for cellulose in water. We also calculate the energy of dissolution of cellulose in BmimCl and water over a wide range of

temperatures to find whether there is an energetic driving force for dissolution in IL, and whether there is one for insolubility in water. Because the time scales for dissolution are too large to access directly via all-atom MD, we calculate the energy of dissolution by simulating cellulose in a microfibril state, and also in a hypothetical, dissolved state in which the glucan chains are separated from each other. Thermodynamic quantities of dissolution are then calculated by taking the difference in values between the two states.

Knowledge of entropic driving forces is also important in any thermodynamic understanding of cellulose dissolution. In chapter 4 we detail our calculations of the entropy of dissolution of cellulose in BmimCl and water for the same systems described in chapter 3. Absolute entropies are calculated for each cellulose state/solvent combination using the two phase thermodynamic (2PT) model,[40] which has been successfully applied to calculate the absolute entropies of Lennard-Jones fluids, water, and other molecular fluids previously.[41-44] The entropies of dissolution are then calculated in the same manner as the energies of dissolution. Entropies of dissolution are split up into solvent and cellulose contributions, with the relative magnitudes of each compared to determine if either dominated the other. Finally, we combine the total entropies of dissolution with the energies of dissolution to calculate the free energy of dissolution of cellulose in BmimCl and water, and the signs of these values are checked against the experimentally known phase behavior of cellulose in either solvent. Also, the relative importance of the entropies and energies of dissolution in determining these free energy values are compared to see which is the greater driving force in each cellulose/solvent system.

Lithium chloride (LiCl) dissolved in N,N-dimethylacetamide (DMA) is another solvent capable of dissolving cellulose.[45,46] The molecular mechanism by which this occurs is generally thought to be interaction of the sugar hydroxyl groups with the dissolved salt, particularly the Cl^- anion.[47,48] Results from ^1H, ^{13}C, and ^7Li NMR experiments show that the chemical shift of glucose hydroxyl hydrogen atoms, glucose hydroxyl-connected carbon atoms, and Li^+ cations, respectively, are functions of salt/sugar ratio or salt identity.[49,50] However, the specifics of how and why salt-sugar interactions occur, and a quantification of their strengths, have yet to be determined.

Because LiCl in water or most other solvents is not able to dissolve cellulose, the unique solvation environment created by DMA is key in causing cellulose dissolution. The presence of DMA creates preferential interactions between LiCl and cellulose that do not exist in other solvents. In Chapter 5 we examine this notion of salt-sugar preferential interactions via a combination of free energy and traditional MD simulations of cellulose deconstruction. The potential of mean force (PMF) of removing a single cellulose chain from the top of a microfibril in LiCl/DMA, pure DMA, LiCl/water, and pure water is calculated to estimate the strengths of the thermodynamic driving force for dissolution or insolubility in these four solvents. We then calculate both the three-dimensional density distribution of dissolved ions around cellulose chains in LiCl/DMA and LiCl/water, and quantify the sugar-ion interaction potentials with a coarse-graining force-matching analysis. From these we infer both the origins of the sugar-ion preferential interactions, as well as the relative importance of sugar-Li^+ and sugar-Cl^- interactions in causing cellulose dissolution in DMA.
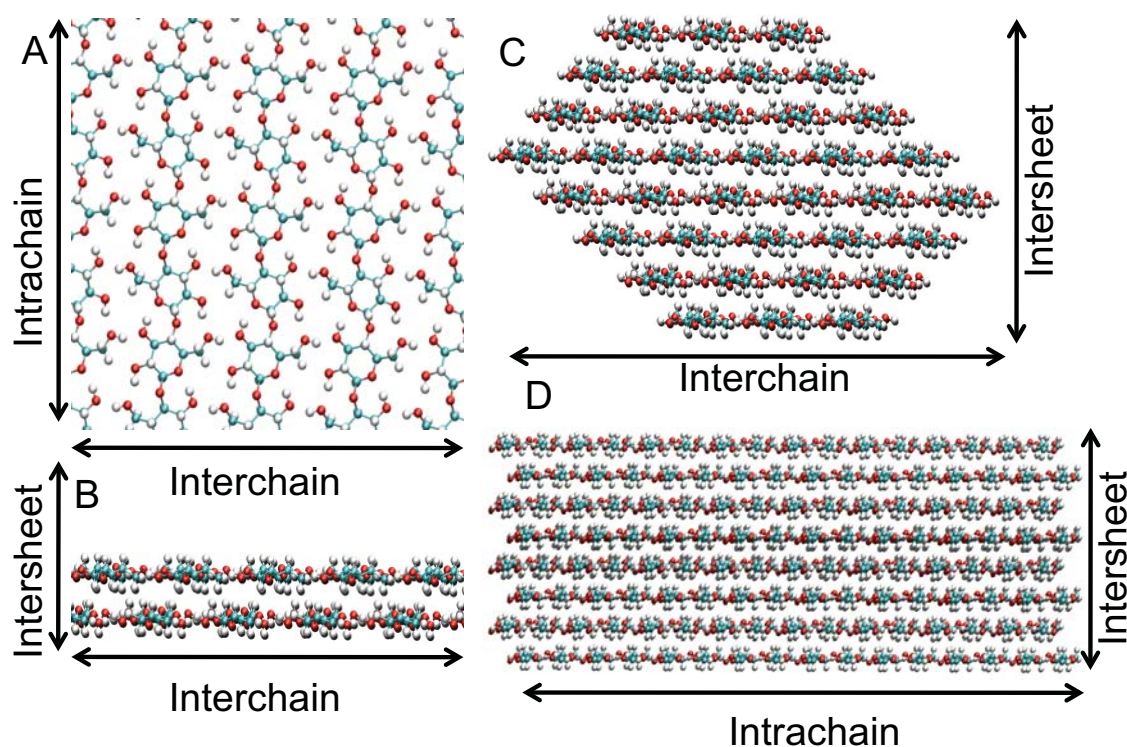
Figure 1.1. Arrangement of cellulose I. The glucan chains are aligned equatorially in flat sheets (A), which are then stacked up upon each other axially (B). The image in (B) is rotated 90° from the one in (A). In (A), the axis of polymerization is vertical, while in (B), it is coming out of the plane of the image. In (C) we show the cross section of a typical microfibril (containing 36 chains), the axis of polymerization coming out of the plane of the image. In (D), the microfibril is rotated 90° from its position in (C), the axis of polymerization running horizontally across the image.

References

(1)   Huber, G. W.; Iborra, S.; Corma, A. *Chem. Rev.* **2006**, *106*, 4044-4098.

(2)   Perlack, R. D. *Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasability of a Billion-Ton Annual Supply*; DOE/GO-102995-2135; Oak Ridge National Laboratory: Oak Ridge, TN, 2005.

(3)   Klass, D. L. *Biomass for enewable Energy, Fules, and Chemicals;*; Academic Press: San Diego, 1998.

(4)   *Annual Energy Review 2010*; DOE/EIA-0384(2010); U. S. Energy Information Administration: Washington, D. C., 2011.

(5)   Himmel, M. E.; Ding, S. Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W.; Foust, T. D. *Science* **2007**, *315*, 804-807.

(6)   Wyman, C. E. *Trends Biotechnol.* **2007**, *25*, 153-157.

(7)   Yang, B.; Wyman, C. E. *Biofuel. Bioprod. Bior.* **2008**, *2*, 26-40.

(8)   Wyman, C. E.; Dale, B. E.; Elander, R. T.; Holtzapple, M.; Ladisch, M. R.; Lee, Y. Y. *Bioresource Technol* **2005**, *96*, 1959-1966.

(9)   Lynd, L. R.; Wyman, C. E.; Gerngross, T. U. *Biotechnol. Progr.* **1999**, *15*, 777-793.

(10)   Wertz, J. L.; Bedue, O.; Mercier, J. P. *Cellulose Science and Technology*; EPFL Press: Lausanne, 2010.

(11)   Ragauskas, A. J.; Williams, C. K.; Davison, B. H.; Britovsek, G.; Cairney, J.; Eckert, C. A.; Frederick, W. J.; Hallett, J. P.; Leak, D. J.; Liotta, C. L.; Mielenz, J. R.; Murphy, R.; Templer, R.; Tschaplinski, T. *Science* **2006**, *311*, 484-489.

(12)   Perez, S.; Mazeau, K. In *Polysaccharides: Structural Diversity and Functional Versatility*, 2nd ed.; Dumitriu, S., Ed.; Marcel Dekker: New York, 2005; pp 41-68.

(13)   Ding, S. Y.; Himmel, M. E. *J. Agr. Food Chem.* **2006**, *54*, 597-606.

(14)   Atalla, R. H.; Vanderhart, D. L. *Science* **1984**, *223*, 283-285.

(15)   Horii, F.; Yamamoto, H.; Kitamaru, R.; Tanahashi, M.; Higuchi, T. *Macromolecules* **1987**, *20*, 2946-2949.

(16)   Nishiyama, Y.; Langan, P.; Chanzy, H. *J. Am. Chem. Soc.* **2002**, *124*, 9074-9082.

(17)   Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P. *J Am Chem Soc* **2003**, *125*, 14300-14306.

(18)   Jeffrey, G. A. *An introduction to hydrogen bonding*; Oxford University Press: New York, 1997.

(19)   Steiner, T.; Saenger, W. *J Am Chem Soc* **1992**, *114*, 10146-10154.

(20)   Steiner, T. *Chem Commun* **1997**, 727-734.

(21)   Kondo, T. In *Polysaccharides: Structural Diversity and Functional Versatility*, 2nd ed.; Dumitriu, S., Ed.; Marcel Dekker: New York, 2005; pp 69-98.

(22)   Marechal, Y.; Chanzy, H. *J Mol Struct* **2000**, *523*, 183-196.

(23)   Watanabe, A.; Morita, S.; Ozaki, Y. *Biomacromolecules* **2006**, *7*, 3164-3170.

(24)   Watanabe, A.; Morita, S.; Ozaki, Y. *Biomacromolecules* **2007**, *8*, 2969-2975.

(25)   Kondo, T. *J Polym Sci Pol Phys* **1997**, *35*, 717-723.

(26)   Tashiro, K.; Kobayashi, M. *Polymer* **1991**, *32*, 1516-1530.

(27)   Bergenstrahle, M.; Berglund, L. A.; Mazeau, K. *J Phys Chem B* **2007**, *111*, 9138-9145.

(28)   Shen, T. Y.; Gnanakaran, S. *Biophys J* **2009**, *96*, 3032-3040.

(29)   Swatloski, R. P.; Spear, S. K.; Holbrey, J. D.; Rogers, R. D. *J. Am. Chem. Soc.* **2002**, *124*, 4974-4975.

(30)   Moulthrop, J. S.; Swatloski, R. P.; Moyna, G.; Rogers, R. D. *Chem. Commun.* **2005**, 1557-1559.

(31)   Singh, S.; Simmons, B. A.; Vogel, K. P. *Biotechnol. Bioeng.* **2009**, *104*, 68-75.

(32)   Pinkert, A.; Marsh, K. N.; Pang, S. S.; Staiger, M. P. *Chem. Rev.* **2009**, *109*, 6712-6728.

(33)   El Seoud, O. A.; Koschella, A.; Fidale, L. C.; Dorn, S.; Heinze, T. *Biomacromolecules* **2007**, *8*, 2629-2647.

(34)   Remsing, R. C.; Swatloski, R. P.; Rogers, R. D.; Moyna, G. *Chem. Commun.* **2006**, 1271-1273.

5

(35)   Remsing, R. C.; Hernandez, G.; Swatloski, R. P.; Massefski, W. W.; Rogers, R. D.; Moyna, G. *J. Phys. Chem. B* **2008**, *112*, 11071-11078.

(36)   Cho, H. M.; Gross, A. S.; Chu, J.-W. *J Am Chem Soc* **2011**, *133*, 14033-14041.

(37)   Youngs, T. G. A.; Holbrey, J. D.; Deetlefs, M.; Nieuwenhuyzen, M.; Gomes, M. F. C.; Hardacre, C. *Chemphyschem* **2006**, *7*, 2279-2281.

(38)   Youngs, T. G. A.; Hardacre, C.; Holbrey, J. D. *J. Phys. Chem. B* **2007**, *111*, 13765-13774.

(39)   Liu, H. B.; Sale, K. L.; Holmes, B. M.; Simmons, B. A.; Singh, S. *J. Phys. Chem. B* **2010**, *114*, 4293-4301.

(40)   Lin, S. T.; Blanco, M.; Goddard, W. A. *J Chem Phys* **2003**, *119*, 11792-11805.

(41)   Lin, S. T.; Maiti, P. K.; Goddard, W. A. *J Phys Chem B* **2010**, *114*, 8191-8198.

(42)   Pascal, T. A.; Abrol, R.; Mittal, R.; Wang, Y.; Prasadarao, N. V.; Goddard, W. A. *J Biol Chem* **2010**, *285*, 37753-37761.

(43)   Pascal, T. A.; Lin, S. T.; Goddard, W. A. *Phys Chem Chem Phys* **2011**, *13*, 169-181.

(44)   Huang, S. N.; Pascal, T. A.; Goddard, W. A.; Maiti, P. K.; Lin, S. T. *J Chem Theory Comput* **2011**, *7*, 1893-1901.

(45)   Mccormick, C. L.; Lichatowich, D. K. *J Polym Sci Pol Lett* **1979**, *17*, 479-484.

(46)   El-Kafrawy, A. *J Appl Polym Sci* **1982**, *27*, 2435-2443.

(47)   Dawsey, T. R.; Mccormick, C. L. *J Macromol Sci R M C* **1990**, *C30*, 405-440.

(48)   Morgenstern, B.; Kammer, H. W. *Trends Polym Sci* **1996**, *4*, 87-92.

(49)   Mccormick, C. L.; Callais, P. A.; Hutchinson, B. H. *Macromolecules* **1985**, *18*, 2394-2401.

(50)   Morgenstern, B.; Kammer, H. W.; Berger, W.; Skrabal, P. *Acta Polym* **1992**, *43*, 356-357.

Chapter 2

On the Molecular Origins of Biomass Recalcitrance: The Interaction Network and Solvation Structures of Cellulose Microfibrils

Abstract

Biomass recalcitrance is a fundamental bottleneck to producing fuels from renewable sources. To understand its molecular origin, we characterize the interaction network and solvation structures of cellulose microfibrils via all-atom molecular dynamics simulations. The network is divided into three components: intrachain, interchain, and intersheet interactions. Analysis of their spatial dependence and interaction energetics indicate that intersheet interactions are the most robust and strongest component and do not display a noticeable dependence on solvent exposure. Conversely, the strength of surface exposed intrachain and interchain hydrogen bonds is significantly reduced. Comparing the interaction networks of I_β and I_α cellulose also shows that the number of intersheet interactions is a clear descriptor that distinguishes the two allomorphs and is consistent with the observation that I_β is the more stable form. These results highlight the dominant role of the often-overlooked intersheet interactions in giving rise to biomass recalcitrance. We also analyze the solvation structures around the surfaces of microfibrils and show that the structural and chemical features at cellulose surfaces constrict water molecules into specific density profiles and pair correlation functions. Calculations of water density and compressibility in the hydration shell show noticeable but not drastic differences. Therefore, specific solvation structures are more prominent signatures of different surfaces.

2.1 Introduction

Decomposition of lignocellulosic biomass followed by catalytic or enzymatic conversion is a potential route for producing liquid fuels from renewable sources.[1-3] However, commercial viability has not yet been achieved due to the recalcitrance of biomass and the concomitant high processing cost.[4] Cellulose is a major component of plant cell walls and a significant contributor to biomass recalcitrance. In plant cell walls, the basic structural unit of cellulose is slender aggregates of cellobiose chains called elementary fibrils or microfibrils.[5] A wide range of diameters (2-20 nm) and lengths (0.1-100 μm) of microfibrils are observed.[6] Such lengths correspond to a degree of polymerization ranging from the hundreds to the tens of thousands of glucose units along the axial direction of the microfibril. The crystalline structures of these microfibrils create a complex network of hydrogen bonding and van der Waals (vdW) interactions that resist deconstruction by solvent or mechanical forces. To develop a molecular basis for advancing the technologies of biomass deconstruction, we aim to characterize in detail the interaction networks of two types of cellulose (I_β and I_α) by using all-atom molecular dynamics (MD) simulations.

Naturally occurring cellulose, cellulose I, has two distinct crystal phases, the monoclinic I_β and the triclinic I_α.[7] I_β is the more stable of the two, as heating of the I_α phase causes irreversible conversion to I_β.[8] The atomic structures of both crystal forms

were only recently resolved,[9,10] and are used as the starting points of our analysis. Both allomorphs consist of polymerized cellobiose chains arranged in parallel to form flat sheets. These sheets are stacked on top of each other to form the full 3-dimensional crystal structure.[9,10] Because of the close proximity of glucose residues in the structures, a complex network of non-bonded interactions permeates the crystals. Specific O-H—O bonds form between neighboring residues of the same chain (intrachain) and between neighboring residues of different chains in the same sheet (interchain). Numerous C-H—O contacts and vdW interactions connect residues on neighboring sheets (intersheet). This interaction network gives rise to the strength and robustness of cellulosic materials.[11] In this work, we quantify the relative strength of different interactions in cellulose microfibrils as well as the solvation structures at different surfaces of cellulose.

Previous studies of the interaction network in cellulose focused mostly on O-H—O hydrogen bonds.[12,13-25] Three O-H—O hydrogen bonds (HB), the O3-H—O5 intrachain HB, the O2-H—O6 intrachain HB, and the O6-H—O3 interchain HB, can be identified from crystal structures and have been shown to bear significant effects on the structures and dynamics of cellulose.[13] For example, intrachain HBs limit the rotation of glucose units about the glycosidic linkages.[19] Of the two intrachain HBs, the O3-H—O5 HB is the dominant form, as indicated by its higher intensity of IR signal in cellulose materials[16] and higher impact in affecting their Young's moduli.[19] Dynamic FTIR measurements on periodically stretched cellulose also show that O3-H—O5 HBs are the major component that responds to the applied stress.[20]

In addition to intrachain HBs, interchain HBs (O6-H—O3) are also found to be important in maintaining the structural integrity of cellulose. Selective etherification of the O6-H group perturbs interchain interactions and leads to a significant disruption in the crystallinity of cellulose and hence increases in solubility.[23] The effects of etherifying O6-H on crystallinity and solubility are much more pronounced than those of etherifying the O3-H and O2-H groups that only participate in intrachain HBs.

Both intrachain and intersheet HBs persist over a wide temperature range. IR measurements show that intrachain HBs (O3-H—O5) are broken after interchain HBs at elevated temperatures, and can still be observed at temperatures (200 ^{O}C) close to the temperature of the melting transition.[21,22] This trend of continued occupancy at high temperatures is also observed in atomic simulations of crystalline cellulose[24] and is suggested from simulation results of a coarse-grain lattice model.[25]

Compared to intrachain and interchain HBs, the nature of intersheet interactions in cellulose and how they measure against those of intrachain and interchain HBs are largely unknown, despite their potentially significant contribution to the recalcitrance of cellulose. Intersheet interactions are expected to involve both C-H—O pseudo-HBs and van der Waals interactions. While C-H—O interactions deviate from the classical definition of hydrogen bonds, the observed geometries in the crystal structures of cellulose have similar characteristics: a close-to-linear C-H—O angle (90°-150°) and a short H—O distance (2.2 Å to 3.2 Å).[26,27] For O-H—O HBs in cellulose, the corresponding values are 130°-180° for O-H—O angles and 1.5 Å to 2.2 Å for H—O bond lengths. As the geometry of C-H—O pairs are observed to have hydrogen bond like properties in the crystal structures of many small organic compounds, including carbohydrates, they are often reported as hydrogen bonds in the literature.[27,28] Furthermore, similar to classical hydrogen bonds, *ab initio* calculations showed that

electrostatics are the dominant contribution to the favorable C-H—O interaction energetics.[28-31] From the results of all-atom MD simulations, we also compute the electrostatic energies of C-H—O HBs and show their dominant contribution (90 %) to individual C-H—O interactions (2 kcal/mol). In the rest of this work, we will refer to intersheet C-H—O interactions as C-H—O HBs.

The high spatial resolution of all-atom simulations provides an ideal framework to distinguish between the different interactions in cellulose microfibrils. By using all-atom MD simulations, we examine how different types of hydrogen bonds in a microfibril differ from each other and how their strengths depend on their proximity to the solvent exposed surfaces. We also quantify the solvation structure and solvent compressibility around the different surfaces of the microfibrils. Knowledge of the solvation of cellulose is especially important for designing degradation processes and for analyzing the adsorption and migration of cellulases on cellulose.

We found that the fluctuations of C-H—O HB angles and H—O (or C—O) distances of intersheet bonds are much smaller than those of intrachain and interchain O-H—O HBs. Therefore, using fluctuations of HB geometries as a metric, intersheet interactions appears to be the most stable among the three types of HBs (intrachain, interchain, and intersheet), even though the electrostatic interaction of each individual HB is weaker. Furthermore, a significant difference between I_β and I_α microfibrils is the number of intersheet HBs. Our results highlight the under-appreciated importance of intersheet interactions as a critical component in causing biomass recalcitrance.

In analyzing the effects of solvent exposure on the interaction network of cellulose, reduction in strength is only observed at the outermost chains. Chains in the second outermost layers and layers further inwards are not affected to a noticeable extent. Such a difference in chain behavior based on distance to microfibril surface has been hypothesized,[5] although here the number of different types is only two. We also compute water compressibility in slabs near different cellulose surfaces and illustrate a clear surface dependence. Water layers near surfaces that have the most exposed CH groups are more compressible, indicating a more hydrophobic nature than other surfaces. Examining water-water hydrogen bonding surrounding different cellulose surfaces reveals that water structures are strongly affected by the topologies and exposed moieties of different surfaces.

2.2 Methods

Two models of a cellulose microfibril are constructed based on the crystal structures of I_β and I_α cellulose.[9,10] Each microfibril contains 36 chains organized into 8 sheets (Figure 2.1). The resulting diameter is 40 Å, in line with the experimentally observed values for microfibrils in plants (typically 30-50 Å).[6] The cross-section was arranged according to the geometries inferred from AFM measurements.[5] Each of the 36 chains in the microfibril contains 16 glucose residues, resulting in a total length of 80 Å. Since cellulose chains in plant cell walls contain hundreds to thousands of glucose units, we connect sugars across the periodic boundary to reduce finite size effects. In conducting all-atom MD simulations, the microfibrils are placed in a hexagonal unit cell with the axis of polymerization aligned with the z-axis. Both the I_β and I_α microfibrils are solvated in explicit TIP3P[32] water. The I_β microfibril was hydrated with 5883 water

molecules, and the I_α microfibril with 5861, both giving a hydration level of 1.13 g H_2O/g cellulose. The resulting models of both microfibrils are shown in Figure 2.1. Each microfibril has hydrophobic surfaces on the top and bottom (exposing CH groups), and four hydrophilic surfaces on the sides.

All cellulose simulations were performed with the molecular simulation software CHARMM[33] under constant temperature (300 K) and pressure (1 atm) conditions via the Nose-Hoover thermostat[34,35] and Langevin piston barostat.[36] The recently optimized CHARMM force field for carbohydrates was employed to describe the interactions between atoms.[37,38] This force field was validated against thermodynamic and structural data for several saccharides in both the aqueous and crystal phases.[37,38] Starting from X-ray structures, a series of energy minimizations was performed followed by a gradual heating to 300 K. The xx and yy components of the stress tensor were controlled to be 1 atm, while the zz component was controlled to have zero stress. Data was taken from production runs of 10 ns with frames saved every 0.5 ps for analysis.

To quantify the strength of hydrogen bond interactions, we compute the averaged occupancies from the all-atom MD simulations. Occupancy was determined using a D-H—A angle cutoff and a H—A distance cutoff. For O-H—O HBs, the angle cutoff is 130° and the distance cutoff is 2.2 Å. For the C-H—O pseudo-HBs we used an angle cutoff of 110° and a distance cutoff of 3.2 Å based on the distribution of H—O distances observed in the crystal structures of carbohydrates.[27] Varying the bond cutoff by up to 0.2 Å and the angle cutoff by up to 20° does not change the relative strength differences between different HBs to any appreciable extent. The calculated statistical uncertainty for all HB occupancies is less than 0.001. Uncertainties were calculated according to standard methods that assume Gaussian statistics of time series.[39] The structural properties and energetics calculated from the 10 ns simulations provide statistically significant assessments of the differences in and between the cellulose interaction networks of the two allomorphs

2.3 Results and Discussion

2.3.1 Fluctuations of Structural Properties of Cellulose Microfibrils

First, we compare the crystalline structures of the microfibrils from the simulation with those of the ideal crystals to quantify the effects of solvation and finite molecular packing. For both the I_β and I_α microfibrils, cellulose atoms remain close to their crystalline positions.[9,10] The averaged root-mean-squared-deviation (RMSD; mass weighted) from the initial configuration (based on X-ray structures) is only 0.71±0.003 Å for the I_β microfibril and 0.77±0.003 Å for the I_α microfibril. The effective unit cell parameters calculated from the simulations of both microfibrils are also very close to those observed in X-ray structures (Tables 2.1 and 2.2) for interior as well as exposed chains. Therefore, finite molecular packing and solvation do not perturb the crystalline structures of cellulose even at the surface.

An important aspect of cellulose recalcitrance is the spatial dependence of the different intermolecular interactions. The unique network of hydrogen bonds and van der Waals interactions is considered to be the main mechanism against deconstruction. As described earlier, most attention has been focused on O-H—O HBs, but not as much on

C-H—O HBs that couple chains in different sheets. In this work, we will compare and contrast the three types of HBs in cellulose. These are:

1. Intrachain HBs - between two glucose residues in the same chain
2. Interchain HBs - between two glucose residues in two different chains but in the same sheet.
3. Intersheet HBs - between two glucose residues in two different chains in two different sheets.

Since cellulose is composed entirely of β-D-glucose, all aliphatic hydrogen atoms occupy the axial position and all hydroxyl groups the equatorial position. This orientation, along with the *tg* conformation of the C_6 exocylic group, prevents intersheet O-H—O connections.[9,10] Tables 2.3 and 2.4 list the hydrogen bonding interactions in the $I_β$ and $I_α$ microfibrils that we identified from the crystal structures of cellulose.[9,10] For a fully coordinated chain, each glucose unit can interact with 14 other nearby residues. In the following, we analyze the properties and spatial dependence of these HBs from the results of all-atom MD simulations.

Tables 2.3 and 2.4 summarize the averaged occupancies (over all 36 chains) of different HBs observed in the all-atom MD simulations of the $I_β$ and $I_α$ microfibrils. It can be seen that intrachain and interchain HBs have similar occupancies in the two allomorphs. The O3-H—O5 intrachain HBs have ~0.9 occupancies and are the most persistent O-H—O HB. This result is consistent with the lower reactivity of the O3-H hydroxyl group toward substituting agents as compared to the O2-H or O6-H groups.[40] The other type of intrachain HB (O2-H—O6) has a lower occupancy, ~0.6, in both the $I_β$ and $I_α$ microfibrils. The smaller occupancy of the O2-H—O6 HB is due to the rotation of the C6 hydroxymethyl group that occurs primarily in the outermost layer of a microfibril.

For interchain HBs, O6-H—O3 is the dominant interaction and has an averaged occupancy of 0.53 for $I_β$ and 0.58 for $I_α$ (Tables 2.3 and 2.4). The involvement of the C6 hydroxymethyl group in interchain HBs leads to lower occupancies than the O3-H—O5 intrachain HBs. The trend that intrachain HBs are more stable than interchain HBs agrees with dynamic FTIR measurements during the slow heating of cellulose which show that intrachain HBs are the last to break.[21] A slightly higher occupancy of interchain HBs in the $I_α$ (0.58) than in the $I_β$ (0.53) microfibril is not in line with the observation that $I_β$ is the more stable crystal form.[8] As described in the following, the differences in intersheet HB occupancy are more pronounced than those of the interchain interactions, and are consistent with the higher stability of $I_β$.

2.3.2 Intersheet HBs in Cellulose are Less Flexible than Intrachain and Interchain HBs

A total of 12 types of intersheet HBs (Tables 2.3 and 2.4) are identified from the X-ray structures of the $I_β$ and $I_α$ crystals via the criteria of a 110° angle cutoff and a 3.2 Å H—O bond length cutoff. The large number of CH groups in a glucose ring give rise to many more participating donor-acceptor pairs than for intrachain and interchain HBs. Due to a (nearly) screw-axis symmetry along a glucose chain and a nearly identical packing of these chains in both types of microfibril,[10] the bonding pairs of intersheet HBs are almost identical between $I_β$ and $I_α$ allomorphs. The only differences are the involvement of a C1-H—O3 HB in the $I_α$ microfibril but not in the $I_β$ microfibril, and the

dual acceptors of the C5-H—O4/O6 HB in the I_β microfibril versus a single acceptor in I_α.

Most of the C-H—O HBs shown in Table 2.3 and 2.4 have high occupancies, around 0.9. Similar to interchain HBs, intersheet HBs that involve the C6 group have lower occupancies due to exocyclic rotation. Furthermore, the occupancies of most C-H—O HBs are higher than the O-H—O HBs (Table 2.3 and 2.4), even though the strength of an individual C-H—O bond (<4 kcal/mol) is less than that of an O-H—O bond (4-15 kcal/mol).[26] With the all-axial configuration of the glucose rings, C-H groups can act together in parallel, leading to a very high occupancy of C-H—O HBs in both the I_β and I_α microfibrils.

To characterize the spatial dependence of intrachain, interchain, and intersheet HBs, we examine and compare the chain-averaged properties of HBs across the microfibril cross-section. In particular, we analyze if surface-exposed chains have weaker HBs, and if so, which types of HBs are most susceptible.

For intrachain HBs, the averaged RMSFs (root-mean-squared-fluctuation) in the O3-H—O5 angle of surface-exposed chains are 50-100% higher than those in the interior. More flexible intrachain HBs at surfaces correlate with the lack of molecular packing and competition with solvent for hydrogen bonding. Corner chains and chains on the hydrophobic surfaces have higher RMSFs in the O3-H—O5 angle than the chains on the surfaces that expose mostly OH groups. These results are shown in Figure 2.8 of the supporting information (SI).

The spatial dependence of interchain HBs is qualitatively similar to that of intrachain bonds: occupancy for chains residing on the solvent exposed surfaces is lower. For the O6-H—O3 interchain HB in the interior of a microfibril, occupancy is ~0.5-0.8 (Figure 2.9 of the SI), and becomes lower at the surfaces. The higher flexibility of surface chains is also observed via ^{13}C NMR measurements which showed that the spin-lattice relaxation rates are higher on the surfaces of cellulose microfibrils compared to the interior.[41]

Since interchain HBs are completely solvent exposed on the hydrophobic surfaces of a cellulose microfibril, the reduction in occupancy is much higher than those of the hydrophilic surfaces. On the hydrophobic surfaces, the averaged occupancy of interchain HBs is less than 0.15. The interchain HBs on the hydrophilic surfaces do not show such pronounced reduction in occupancy. The significant reduction in the strength of interchain HBs on hydrophobic surfaces can also be shown by observing the RMSF of the O-H—O angles; results are shown in Figure 2.10 of the SI.

The C-H—O intersheet HBs that do not involve the C6 hydroxymethyl group (see Tables 2.3 and 2.4) reside in between the cellulose sheets of a microfibril, and thus have no dangling components or direct solvent exposure. For these intersheet HBs, the averaged occupancies of different chains do not show a noticeable spatial dependence. The robustness of intersheet HBs is observed by measuring the RMSFs of the C-H—O angle, which only increase slightly at the outermost layer. Chain-averaged bond occupancy also has a constant value across the microfibril cross-section (Figures 2.11-14 of the SI). The high integrity and low flexibility of the intersheet HBs that do not involve C6 is a significant signature of cellulose recalcitrance. Therefore, deforming a cellulose microfibril via delamination is expected to be difficult due to the requirements of breaking these buried intersheet HBs.

For intersheet C-H—O HBs that involve the C6 group, exocyclic rotation could lead to a more pronounced reduction in strength at the surfaces. However, although exocyclic rotation inevitably breaks the original C-H—O HB, alternative hydrogen bonds form in the new rotameric state to retain the net interaction. As a result, the total number of C6 involved C-H—O HBs is not reduced due to exocyclic rotation. For instance, when the C6 group of a C6-H—O5 HB on the (110) surface of the I_β microfibril rotates from the *tg* to *gt* configuration[42] (*tg* is the rotameric state observed in X-ray structures), the other C6 hydrogen moves in between C6 and O5 to retain the C-H—O HB. This compensating behavior is shown in Figure 2.2. Similar behaviors are observed in both microfibrils, and another example is shown in Figure 2.15 of the SI. The architecture of molecular packing in both the I_β and I_α microfibrils allows for the retention of intersheet HBs upon exocyclic rotation and ensures the integrity of the interaction network of the cellulose microfibrils.

2.3.3 Cellulose I_β Differs from I_α in Having a Higher Number of Intersheet HBs

Combining all three types, the average number of HBs per glucose residue is shown in Figure 2.3. For both the I_β and I_α microfibrils, the total number of HBs per glucose is highest in the interior and decreases toward the surface due to both a lack of bonding partners and decreases in individual bond occupancy from finite packing and solvent effects. The glucose chains on the two hydrophobic surfaces ((100) and $(\bar{2}\,00)$ for I_β and (110) and $(\bar{1}\,\bar{1}\,0)$ for I_α, top and bottom of each microfibril in Figure 2.3) have the fewest number of hydrogen bonds. These results are not sensitive to the criteria used for determining HB occupancy.

Comparing I_β and I_α microfibrils indicates that each glucose residue in I_β has two more HBs in the interior than the residues in I_α (24.5±0.04 versus 22.5±0.04), consistent with the higher stability observed for the I_β phase.[8] Figure 2.4 shows the average number of the three types of hydrogen bonds for the two microfibrils; only interior chains are considered in the calculation. It can be seen that the two allomorphs only differ in the number of intersheet HBs. Therefore, we conclude that intersheet HBs are the main cause of the different stabilities between I_β and I_α cellulose.[8] This result highlights the importance of intersheet interactions in giving rise to the recalcitrance of cellulose.

2.3.4 Interaction Energies Corroborate the Stabilizing Effects of Intersheet Interactions.

To further quantify the roles of intrachain, interchain, and intersheet HBs in contributing to the recalcitrance of cellulose microfibrils, we compute the interaction energies between glucose residues using the configurations sampled in all-atom MD simulations. In short, the coordinates of a pair of neighboring chains or sheets are extracted from each frame of the MD trajectory for computing the interaction energy (E_{12}- E_1- E_2; E is the potential energy); the rest of the atoms in the system are not included in the calculation. The resulting energetics provide only an approximate description of the nature of the interactions since solvent-mediated interactions, long-range electrostatics, and many-body correlations are not fully considered. The results of these calculations, normalized per glucose, are shown in Table 2.5. Intersheet interaction

energies range between -4.9 to -5.8 kcal/mol/glucose across the microfibril, i.e. they have no significant systematic spatial variation. Conversely, for interchain interactions, interaction energies are -4.3 kcal/mol/glucose for interior chains and drop by roughly half for chains on the surface. Comparing the two types in the interior of the microfibrils, intersheet interaction energies are stronger than interchain interactions by ~1.5 kcal/mol/glucose in I_β and ~0.7 kcal/mol/glucose in I_α. On the exterior, the differences are even more pronounced. To the best of our knowledge, stronger intersheet interactions than interchain interactions in cellulose microfibrils have not yet been illustrated.

  In Table 2.5, we also present the contribution of different energy terms to the interaction energetics. Although the calculated values also reflect the specific nature of the force field used, we aim to examine the trends of the spatial dependence of the different energy terms. The decomposition of interaction energies shown in Table 2.5 indicates that both vdW and electrostatics forces have significant contributions to intersheet interactions, with the vdW term contributing a higher percentage (60-70%). Prior simulation work[12] also showed that intersheet interactions are dominated by vdW forces, but to a far greater extent (> 90%). Separate calculations of the interaction energies of individual O-H—O HBs (6 kcal/mol) and C-H—O pseudo HBs (2 kcal/mol) using the configurations of glucose monomers indicate that electrostatics are the dominant contributor to both types of HBs. Thus, the vdW interactions are acting in concert with the weaker C-H—O HBs to compose the intersheet interactions. The sum of these two results in per residue intersheet interaction energies stronger than per residue interchain interactions (Table 2.5). The magnitudes of the electrostatics and vdW components of the intersheet interactions also have very little spatial dependence, consistent with the high occupancies and low angle RMSFs of C-H—O HBs discussed earlier.

  For interchain HBs, the electrostatic component dominates the interaction energy (more than 90% for interior chains). As O-H—O HBs are significantly disrupted on hydrophobic surfaces, the contribution of electrostatics to interchain interactions drops significantly on these surfaces (Table 2.5). For interchain interactions on the $(\bar{2}\,00)$ face of I_β and the $(\bar{1}\,\bar{1}\,0)$ face of I_α the contribution of electrostatics reduces to 82%; on the (100) face of I_β and the (110) face of I_α, the contribution of electrostatics reduces to 68%. While the interchain electrostatic energies reduce on the exterior of the microfibrils, the vdW components remain constant regardless of chain position. This result shows that the crystalline structures of the microfibrils are not affected at the surface despite the reduced strength of the O-H—O HBs there. The intersheet interactions from the next layer are able to retain the structural integrity.

  Figure 2.4 shows that while they both have the same number of intrachain and interchain HBs, cellulose I_β has two more intersheet HBs per glucose than cellulose I_α. Calculations of interaction energies further corroborate that intersheet HBs are the main contributor to the stronger intersheet interactions of cellulose I_β than those of cellulose I_α (Table 2.5). In the interior of microfibril, the magnitude of the intersheet interaction energy is 0.6 kcal/mol/glucose higher in I_β than in I_α. The majority of this difference, 0.44 kcal/mol/glucose, is from electrostatics. While there have been previous attempts to explain the molecular origin of the observed higher stability of cellulose I_β than $I_{\alpha,}$[12,43] in this work, we identify intersheet interactions as the cause, with most of the contribution coming from electrostatics (C-H—O HBs).

2.3.5 Solvation Structures at the Surfaces of Cellulose Microfibrils

In addition to the interaction network, solvation structures around cellulose are also an important consideration in designing processes to deconstruct cellulose. In Figure 2.5, the spatial distribution of water densities near the different surfaces of the microfibrils are shown. Cellulose surfaces are topographically and chemically heterogeneous, with CH and OH rich regions varying over molecular-scale distances (3-5 Å). As a result, the hydration layer water densities shown in Figure 2.5 vary considerably within each microfibril surface. Each surface has a distinct density profile due to the specific patterns of CH and OH exposure.

Since water molecules tend to form HBs with surface OH groups, and the immediately neighboring CH groups constrict the hydrogen bonding partners of water, high-density regions localized around the OH sites can be seen clearly in Figure 2.5. This result is consistent with earlier simulation work on cellulose microfibrils of different shapes, which also suggested that a dense first solvation shell may impede the diffusion of molecules towards the microfibril surface[42] (see the SI for additional details of the density profiles around the microfibril faces).

In addition to regions near the OH groups of a surface, regions of high water density can also be found around the exposed CH groups, particularly on the (110), ($\overline{1}\,\overline{1}$ 0), ($\overline{2}$ 00), and (100) surfaces of the I_β microfibril and the ($0\overline{1}$ 0), (010), (110) and ($\overline{1}\,\overline{1}$ 0) surfaces of the I_α microfibril. On these surfaces, the distances between the OH groups that flank each CH exposed regions are only 5-7 Å, allowing the hydrogen-bonding network of water to permeate across as shown in Figure 2.5C. Instead of forming hydrogen bonds with the CH groups, water molecules form stronger hydrogen bonds with each other due to the imbalanced distribution of bonding partners and thus the spatial correlation of water density is higher, a signature of small-scale hydrophobic effects.[44] On the surfaces that do not expose the CH groups, regions of high water density are only observed around the OH groups (the ($1\overline{1}$ 0) and ($\overline{1}$ 10) surfaces of I_β and the ($\overline{1}$ 00) and (100) of I_α), and there is no second area of higher water density as on hydrophobic surfaces (see the SI for the analysis of the solvent accessible areas of different cellulose surfaces).

Other than the heterogeneous distributions of water density in the radial directions, the solvation structures along the longitudinal direction also have strong surface dependence. The pair correlation functions (PCFs) of the centers of mass of water along the longitudinal direction at the different surfaces of the microfibrils are shown in Figure 2.6. PCFs are calculated only for water molecules that are within 5 Å of a surface. Near the ($1\overline{1}$ 0) and ($\overline{1}$ 10) surfaces of I_β and the ($\overline{1}$ 00) and (100) of I_α, where the water ordering effects are minimal in the radial directions due to the least exposure of CH groups, water molecules have the most prominent ordering in the longitudinal PCFs. The periodicity of peaks in the longitudinal PCF is 2.6 Å, corresponding to the interstitial distances between polar oxygens on these surfaces. On the surfaces that have the most CH exposure, ($\overline{2}$ 00) and (100) of I_β and ($\overline{1}\,\overline{1}$ 0) and (110) of I_α, the periodicity of the longitudinal PCFs is 5.2 Å instead, which corresponds to the interstitial distances between CH-exposed regions, and is roughly the length of a single glucose molecule. On the remaining two surfaces, the effects induced by OH and CH groups appear to interfere, resulting in an OH-type periodicity of 2.6 Å, but with damped oscillation amplitude in the longitudinal PCFs. The analyses of radial and longitudinal correlation functions indicate

that water structures around cellulose are strongly affected by the specific surface patterns.

The averaged number density in a hydration shell of 5 Å above each cellulose surface is shown in Figure 2.7A. The shells with the highest density are the $(110)/(\bar{1}\bar{1}0)$ of I_β and the $(0\bar{1}0)/(010)$ of I_α. The shells with the smallest densities are the $(1\bar{1}0)/(\bar{1}10)$ of I_β and the $(\bar{1}00)/(100)$ of I_α. The CH-rich surfaces are of intermediate density. Therefore, averaged water densities do not distinguish the hydrophobic nature of different surfaces.[45] A useful metric of the hydrophobicity of a surface is the compressibility (or density fluctuations) of water molecules in the hydration shell.[44-47] We show in Figure 2.7B-C the hydration shell compressibility of the six surfaces of the I_β and I_α microfibrils. The hydration shells are most compressible near the surfaces with the largest exposure of CH groups, the $(\bar{2}00)$ and (100) faces of I_β and the (110) and $(\bar{1}\bar{1}0)$ faces of I_α. The $(1\bar{1}0)/(\bar{1}10)$ I_β and $(\bar{1}00)/(100)$ I_α solvation shells are of intermediate compressibility, and water in the $(110)/(\bar{1}\bar{1}0)$ I_β and $(0\bar{1}0)/(010)$ I_α shells is the least compressible. The hydration shells near the CH exposed surfaces do have a larger compressibility, but the difference with the other two surfaces is not dramatic, ~20% higher. Comparing the hydration shell compressibility of water near folded and unfolded protein shows a difference three-and-a-half times as large.[46] Therefore, the specific solvation structures are the more prominent signatures of different cellulose surfaces and are an important consideration for designing effective molecular and chemical processes to deconstruct cellulose.

2.4 Conclusions

In this work we analyzed the interaction network and solvation structures of cellulose microfibrils by all-atom MD simulations. The major finding is that intersheet interactions, which involve C-H—O pseudo hydrogen bonds and vdW interactions acting in concert, are the strongest and most robust component in the interaction network. Structural fluctuations of intersheet interactions are spatially homogeneous, with no variation with respect to closeness to solvent exposed surfaces. Although the interaction energy of each individual C-H—O HB is less than that of an O-H—O HB (both of which are electrostatic in nature), the addition of vdW interactions makes intersheet interactions stronger than interchain interactions (on a per glucose basis, Table 2.5). Consistent with our calculations, early simulation studies on isolated glucose chains also showed that it is energetically more favorable for glucose rings to associate via staggered hydrophobic stacking than by hydrogen bonding through hydrophilic groups.[48] The other two types of HBs, the intrachain and interchain HBs, show significant reduction in strength in solvent-exposed chains as they can also form hydrogen bonds with water molecules. Furthermore, the number of intersheet C-H—O HBs per glucose is much larger than those of the other two components (Figure 2.4). Our simulation results thus point to intersheet interactions as a focal point for the engineering via molecular interactions of processes such as pretreatment for the deconstruction of crystalline cellulose.

We also found that water structures around cellulose are strongly affected by the specific patterns of individual surfaces. We showed that water molecules near the CH-presenting $(\bar{2}00)$ and (100) surfaces of I_β and (110) and $(\bar{1}\bar{1}0)$ surfaces of I_α have a higher compressibility than those in other solvation shells, but not to a large extent

(~20% higher). We hypothesize that surface-dependent solvation structures would affect the reactivity of different cellulose surfaces toward catalytic and enzymatic reactions. The coupling between solvation structure and reactivity would thus be an important consideration in the development of viable processes for producing fuels from biomass. In addition to the chemistry step of breaking glycosidic bonds, the access of catalysts to cellulose surfaces and desorption of soluble products are also likely to be strong functions of solvation structure.

2.5 Acknowledgements

2.6 Supporting Information

2.6.1 O3-H—O5 Intrachain HB

As seen in Figure 2.8, the variation of the O3-H—O5 angle RMSF from the interior of the microfibril to its surface is similar in both the I_β and I_α forms and has also been observed in earlier work.[12] RMSF is low in the interior, and rises on the surface. Intrachain O2-H—O6 bond trends are comparable (results not shown), although it is complicated by hydroxymethyl group rotation, as the O6 exists as part of rotationally mobile exocyclic hydroxymethyl group. Rotation to other conformers destroys the O2-H—O6 bond.

2.6.2 O6-H—O3 Interchain HB

The O6-H—O3 HB angle RMSFs (Figure 2.10) behave in line with the O6-H—O3 occupancies (Figure 2.9, discussed in the main text). They are lowest in the microfibril interior, and rise on the four lateral surfaces. Fluctuations are largest on the top and bottom surfaces of each microfibril, the ($\bar{2}$ 00) and (100) surfaces of the I_β microfibril and the (110) and ($\bar{1}$ $\bar{1}$ 0) surfaces of the I_α microfibril.

The sheet-by-sheet alternating behavior of both the HB occupancies and angle RMSFs in the I_β microfibril are due to each sheet being composed of a single 'type' of glucose. There are two geometrically distinct glucose molecules in each allomorph's unit cell. In I_β's case, each sheet is constructed with only one of these two glucoses, the identity of which alternates between sheets. This gives rise to the sheet-by-sheet alternating behavior, and is not seen in I_α because for I_α, each chain is composed of alternating glucose 'types', making each sheet an equal mixture of the two.

2.6.3 C2-H2-OX Intersheet HB

Intersheet bonds that do not involve the C_6 exocyclic hydroxymethyl group all behave similarly. They are stable in the interior, and do not have any significant reduction of occupancy on the microfibril exterior. As an example, we show the average HB angle RMSF and occupancy of the C2-H2—O3 interchain HB of the I_β microfibril in Figures 2.11-12, and the same for the C2-H2-O4 of the I_α microfibril in Figures 2.13-14, respectively. HB angle fluctuations are constant throughout the microfibril, and are actually smaller than the fluctuations of the stronger O-H—O intrachain and interchain bonds (Figures 2.8-10). Bond occupancies decrease slightly, if at all, on the surface, and the amount is far less than the reduction of the same of intrachain or interchain bonds on the surface, especially on the hydrophobic faces.

2.6.4 HB Angle RMSF Error Analysis

The statistical uncertainties in HB angle RMSFs are no more than 0.1% of the averages.

2.6.5 Hydrogen Bonding Switching Behavior

We have shown that intersheet bonds that involve the C_6 group display a switching behavior on the microfibril surfaces. As another example, we show in Figure 2.15 such a switching process on the (100) surface of the I_α microfibril, where the C_6 group is acting as a HB acceptor. In the crystalline tg state[42] on the surface, two bonds exist, the C1-H—O6 HB and the C3-H—O6 HB. However, after exocyclic rotation to the gt position, the C4-H—O6 HB replaces the C3-H—O6 HB, while the C1-H—O6 HB remains intact. This is seen in Figure 2.15, as the net HB occupancy is always constant near two despite individual HB occupancies changing. The exocyclic group is allowed to rotate, but rotation does not cause a net loss in favorable interactions. The number of hydrogen bonds remains the same, as one C-H—O HB is switched for another. Similar processes occur on all microfibril surfaces, and would be partially responsible for cellulose's recalcitrance

2.6.6 Surface Chains Contain Fewer HBs: Disentangling Lack of Neighboring Chains with Solvent Interactions and Finite Packing Effects

We have shown that glucose residues on the microfibril surfaces have fewer HBs than those in the interior (Figure 2.3 of the main text). Moving from chains in the interior to those on the four lateral surfaces, the number of HBs per glucose is reduced by ~25%, and going from the interior to the two hydrophobic surfaces and surface corners, the reduction is ~50%.

To decouple the effects of neighbor coordination number and chain flexibility, we examine the ratio of HBs per glucose per chain to the number expected from the number of nearest neighbors to that glucose in Figure 2.16. The results show that deviation of this ratio from one occurs only on the hydrophobic surfaces and the surface corners of

each microfibril. The results of all-atom MD simulations thus indicate that the (100) and ($\bar{2}$00) surfaces of I$_\beta$ and (110) and ($\bar{1}$ $\bar{1}$ 0) surfaces of I$_\alpha$, or any other more hydrophobic surface of a cellulose microfibril, as well as surfaces corners, would be most susceptible to structural distortion. The glucoses at these positions have the overall least number of HBs, and that number is actually less than expected from packing considerations.

2.6.7 Solvent Density Profiles

In Figures 2.17-18 we show normalized density profiles of water above the various microfibril surfaces of cellulose I$_\beta$ and I$_\alpha$. Profiles, and thus densities, are similar between the two allomorphs. All show the presence of a first peak, and decay to a normalized density of 1 by 10 Å. Also, solvent density profiles of diametrically opposed faces of a microfibril mirror each other, except for the bottom right and top left faces of both microfibrils. As mentioned in the main text, we see that macroscopic expectations of behavior do not necessarily match microscopic density profiles: of the six surfaces, waters near the hydrophobic faces are the most strongly coordinated. Because surfaces are typographically rough, classification of interfaces as wet or dry based on first peak position are not viable.   Although to compute the density profiles, the surfaces are treated as flat planes, they are in fact heterogeneous, with side groups protruding irregularly into the solvent. Therefore, the 'zero' of the surface is only approximate, as local surface boundaries vary considerably over the course of a given face.

2.6.8 Solvent Accessible Surface Area

In Figures 2.19-20 we show the average solvent accessible surface area per glucose for the six surfaces of each microfibril. Both show the same trend. The hydrophobic surfaces have by far the most surface area. Of the four lateral surfaces, the bottom right and top left faces of each microfibril have the lowest area, and the other two are somewhat larger. This trend follows that of intersheet chain spacing for the four lateral surfaces, 6 Å for the lower area faces, and 7 Å for the higher ones. The increase in chain spacing opens up the interior of the surface glucoses, exposing more of the aliphatic CH groups. In fact, the bulk of the difference in surface area among the lateral sides is due to an increase in water-accessible aliphatic hydrogen area (Data not shown).

Table 2.1. Unit cell parameters from simulation and diffraction studies[9] for cellulose I$_\beta$.

Parameter	Simulation	Experimental
A (Å)	7.957±0.019	6. 784
B (Å)	8.350±0.022	8.201
C (Å)	10.444±0.008	10.38
α (°)	90.18±0.27	90.00
β (°)	90.00±0.21	90.00
γ (°)	98.31±0.25	96.50

Table 2.2. Unit cell parameters from simulation and diffraction studies[10] for cellulose I$_\alpha$.

Parameter	Simulation	Experimental
A (Å)	7.008±0.018	6.717
B (Å)	5.944±0.018	5.962
C (Å)	10.365±0.008	10.400
α (°)	118.54±0.20	118.08
β (°)	114.32±0.22	114.80
γ (°)	80.35±0.21	80.37

Table 2.3. Hydrogen bonds that exist in the crystal structure and are occupied for a significant time in the simulation of the cellulose I$_\beta$ microfibril.

Class	Interaction (D-H—A)	H—A Distance (Å)[a]	D-H—A Angle (°)[a]	Occupancy[b]
Intrachain	O3-H—O5	1.75/1.97	162.2/137.0	0.850
Intrachain	O2-H—O6	1.90/1.83	165.1/158.7	0.589
Interchain	O6-H—O2	2.54/3.37	125.0/89.4	0.033
Interchain	O6-H—O3	1.78/2.04	156.6/144.3	0.537
Intersheet	C1-H—O6	2.55	144.2	0.697
Intersheet	C2-H—O3	2.49	172.4	0.965
Intersheet	C2-H—O4	2.61	157.8	0.997
Intersheet	C2-H—O6	2.83	137.7	0.731
Intersheet	C3-H—O2	2.52	162.0	0.911
Intersheet	C3-H—O6	3.19	133.8	0.187
Intersheet	C4-H—O2	2.84	144.6	0.928
Intersheet	C5-H—O3	2.80	166.6	0.963
Intersheet	C5-H—O4	2.59	145.0	0.988
Intersheet	C5-H—O6	2.66	141.4	0.833
Intersheet	C6-H—O2	2.36	172.8	0.768
Intersheet	C6-H—O5	2.75	152.1	0.874

[a]Distances and angles are those in the crystal structure.
[b]For O-H—O bonds, a HB cutoff of 2.2 Å and 130° was used. For C-H—O bonds, a cutoff of 3.2 Å and 110° was used.

Table 2.4. Hydrogen bonds that exist in the crystal structure and are occupied for a significant time in the simulation of the cellulose I_α microfibril.

Class	Interaction (D-H—A)	H—A Distance (Å)[a]	D-H—A Angle (°)[a]	Occupancy[b]
Intrachain	O3-H—O5	2.08/1.96	137.9/164.5	0.858
Intrachain	O2-H—O6	1.70/1.77	134.6/125.9	0.618
Interchain	O6-H—O2	2.88/2.79	135.2/141.1	0.037
Interchain	O6-H—O3	1.87/2.18	153.0/151.7	0.584
Intersheet	C1-H—O3	3.11	145.0	0.682
Intersheet	C1-H—O6	2.90	151.7	0.968
Intersheet	C2-H—O3	2.62	170.6	0.760
Intersheet	C2-H—O4	2.51	149.9	0.999
Intersheet	C2-H—O6	2.94	144.6	0.561
Intersheet	C4-H—O2	3.10	159.8	0.735
Intersheet	C3-H—O2	2.59	160.3	0.847
Intersheet	C3-H—O6	3.12	146.2	0.663
Intersheet	C5-H—O3	2.85	156.6	0.917
Intersheet	C5-H—O4	2.71	155.9	0.996
Intersheet	C6-H—O2	2.48	176.5	0.836
Intersheet	C6-H—O5	2.46	134.0	0.857

[a]Distances and angles are those in the crystal structure.
[b]For O-H—O bonds, a HB cutoff of 2.2 Å and 130° was used. For C-H—O bonds, a cutoff of 3.2 Å and 110° was used.

Table 2.5. Sheet-sheet and chain-chain interaction energies for cellulose I$_\beta$ and I$_\alpha$

Type[a]		Interaction Energy (kcal/mol-glucose)			f_{vdw}[b]	f_{elec}[b]
		Total	van der Waals	Electrostatics		
Sheet	I$_\beta$	-5.66±0.02	-3.83±0.01	-1.83±0.03	0.677	0.323
Middle	I$_\alpha$	-5.07±0.02	-3.68±0.01	-1.39±0.03	0.726	0.274
Sheet	I$_\beta$	-5.71±0.02	-3.46±0.01	-2.24±0.02	0.607	0.393
Top	I$_\alpha$	-4.91±0.01	-3.46±0.00	-1.44±0.01	0.706	0.294
Sheet	I$_\beta$	-5.21±0.01	-3.46±0.00	-1.75±0.02	0.664	0.336
Bottom	I$_\alpha$	-5.04±0.04	-3.40±0.01	-1.64±0.04	0.674	0.326
Chain	I$_\beta$	-4.18±0.00	-0.40±0.00	-3.79±0.05	0.095	0.905
Middle	I$_\alpha$	-4.39±0.01	-0.41±0.01	-3.99±0.01	0.093	0.907
Chain	I$_\beta$	-2.43±0.19	-0.44±0.01	-1.99±0.19	0.179	0.821
Top	I$_\alpha$	-2.25±0.03	-0.37±0.02	-1.87±0.03	0.166	0.834
Chain	I$_\beta$	-1.94±0.03	-0.62±0.01	-1.33±0.04	0.317	0.683
Bottom	I$_\alpha$	-1.85±0.02	-0.60±0.01	-1.26±0.02	0.321	0.679

[a]For sheet-sheet interactions, top refers to the two uppermost sheets, middle the two middle sheets, and bottom the two lower most sheets. For chain-chain interactions, top refers to two chains on the top sheet, middle, chains from the middle sheet, and bottom, chains from the bottom most sheet.

[b]F$_{vdw}$ and f$_{elec}$ represent the fraction of total interaction energies from van der Waals and electrostatic interactions respectively.
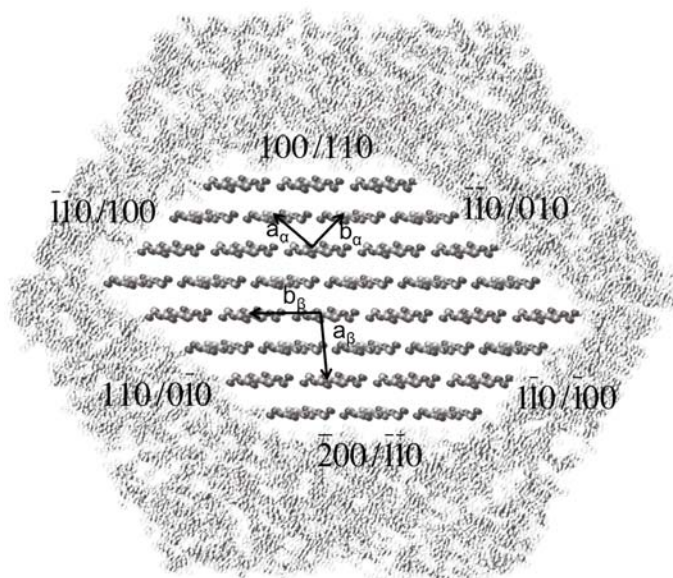
Figure 2.1. Microfibril cross-section with solvent water. Miller indices for the I_β and I_α microfibrils are shown next to their corresponding faces. Unit cell vectors are also shown. For both allomorphs, the c vector is orthogonal to the plane of the image.
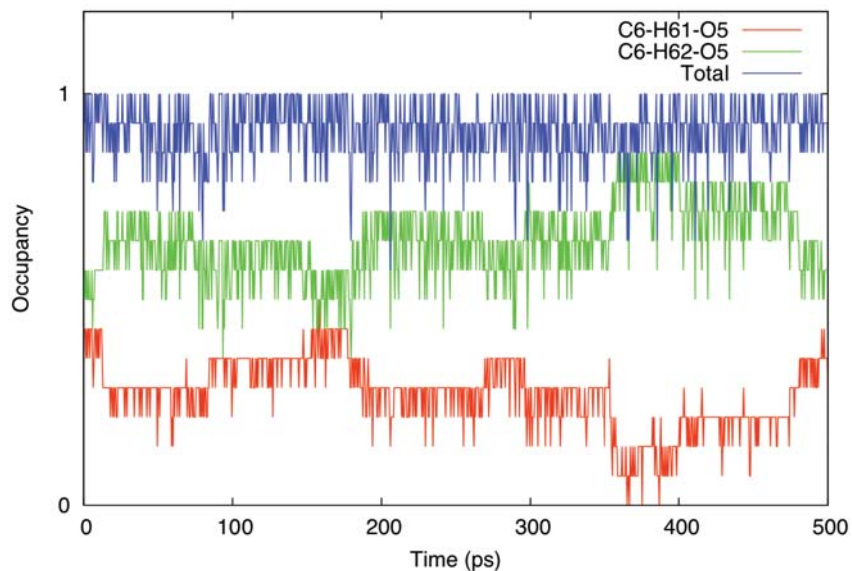
Figure 2.2. Alternative intersheet hydrogen bonds on the (110) surface of the I_β microfibril upon exocyclic rotation of the C6 group. A HB cutoff of 3.2 Å and 110° was employed. The C6-H61-O5 hydrogen bond is the native bond. Results are shown only for the first 500 ps, but behavior is similar for the entire simulation.

(A)

12.3 14.1 13.7

17.7 22.6 22.8 17.6

18.5 24.5 24.4 24.6 17.9

12.3 24.5 24.4 24.4 24.7 17.0

17.6 24.8 24.8 25.1 24.6 13.6

17.5 24.3 24.5 24.6 17.8

18.3 24.3 24.2 19.0

12.9 13.9 12.4

(B)

11.6 12.9 11.9

17.1 22.1 21.8 16.7

16.9 23.0 22.3 22.9 16.2

12.6 22.9 22.7 22.6 22.9 15.8

16.2 22.6 22.6 22.8 22.7 12.5

16.3 22.4 22.2 22.7 17.5

16.2 22.3 22.4 17.9

12.3 13.3 11.9

Figure 2.3. Average number of hydrogen bonds per glucose per chain for the I_β (A) and I_α (B) microfibrils. A HB cutoff of 2.2 Å and 130° was used for O-H—O bonds, and 3.2 Å and 110° for the C-H—O bonds. Error are all less than 0.1.
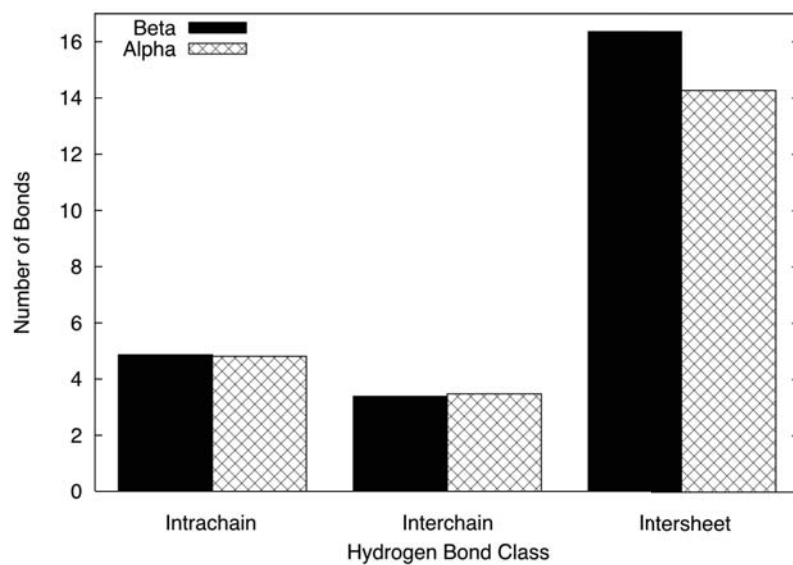
Figure 2.4. Average number of hydrogen bonds per glucose for each HB class in the interior of the cellulose microfibrils.

(A)
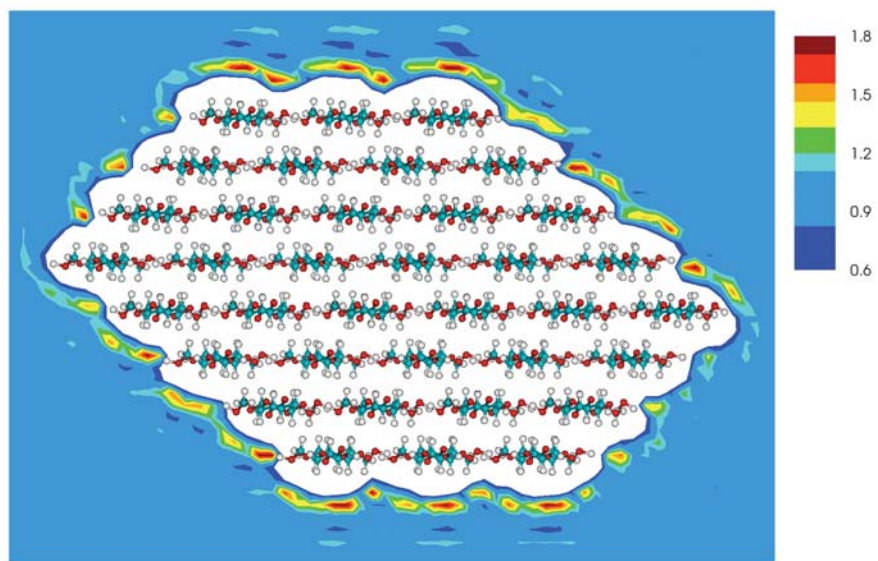
(B)
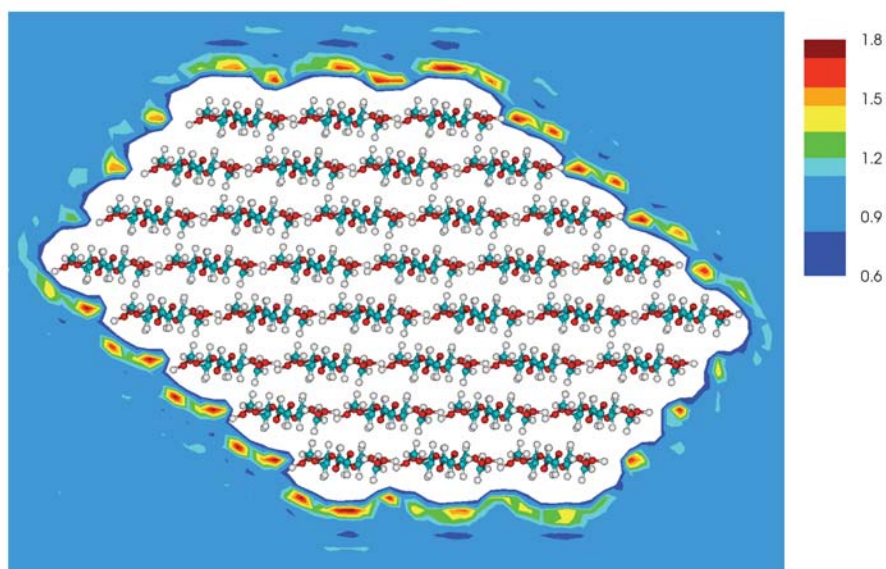
(C)
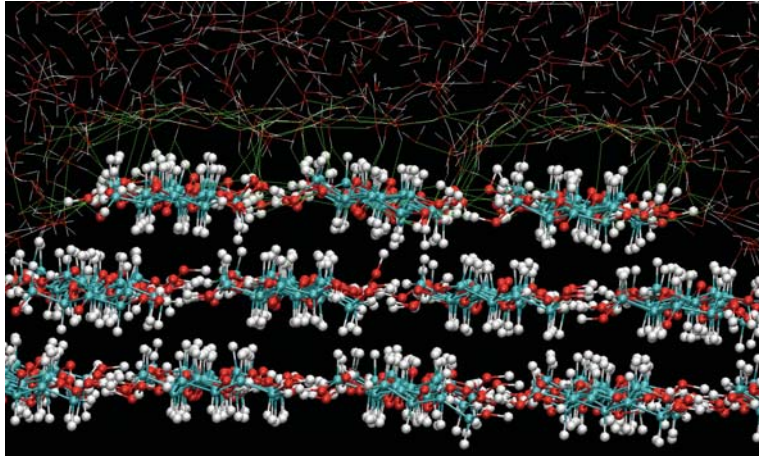
Figure 2.5. Water density relative to the bulk, averaged along the polymerization axis direction for the length of the simulations of the I_β (A) and I_α (B) microfibrils, with microfibrils represented in the interior. (C) A snapshot of the water-cellulose hydrogen bond network (in green) near the (110) surface of cellulose I_α.

(A)
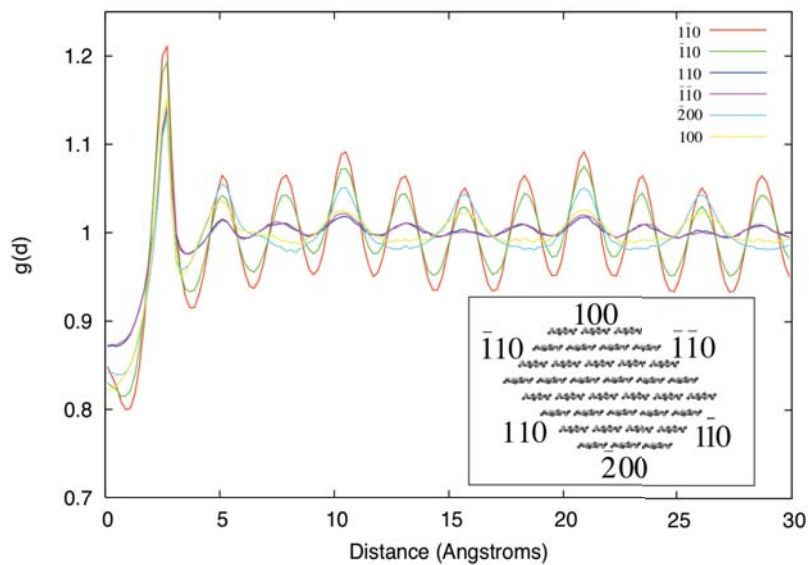
(B)
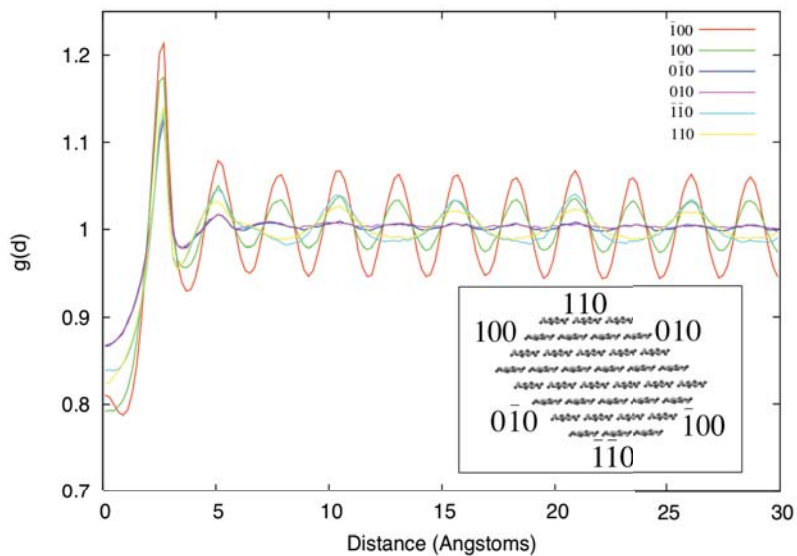
Figure 2.6. Water-water density correlation along the polymerization direction for each of the hydration layers of the six surfaces of each microfibril, I_β (A) and I_α (B). The hydration layer was defined to be a rectangular box 5 Å in height about each surface, beginning at the initial rise in water density perpendicular to the surface.

(A)
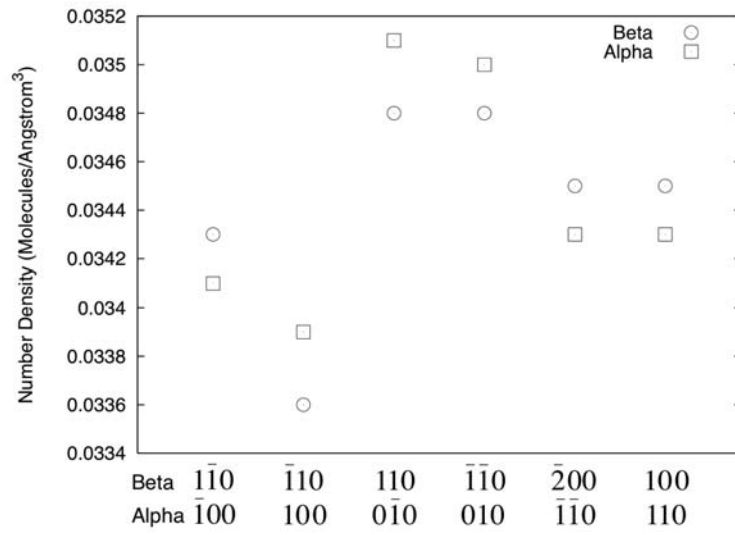
(B)
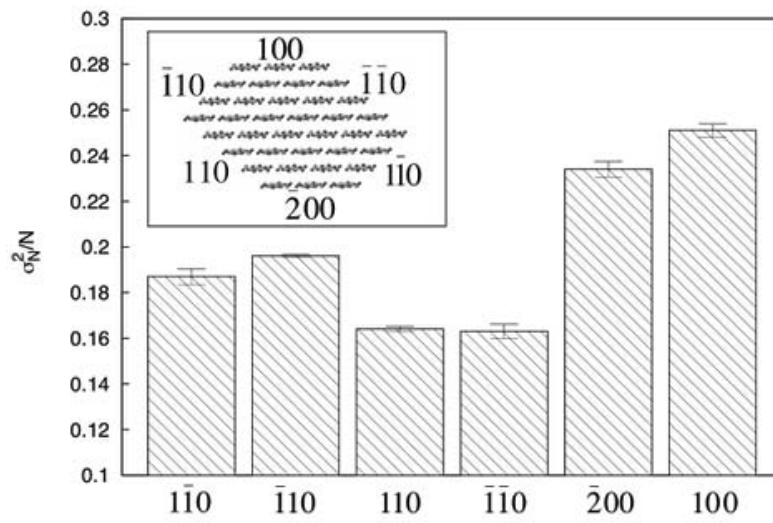
(C)
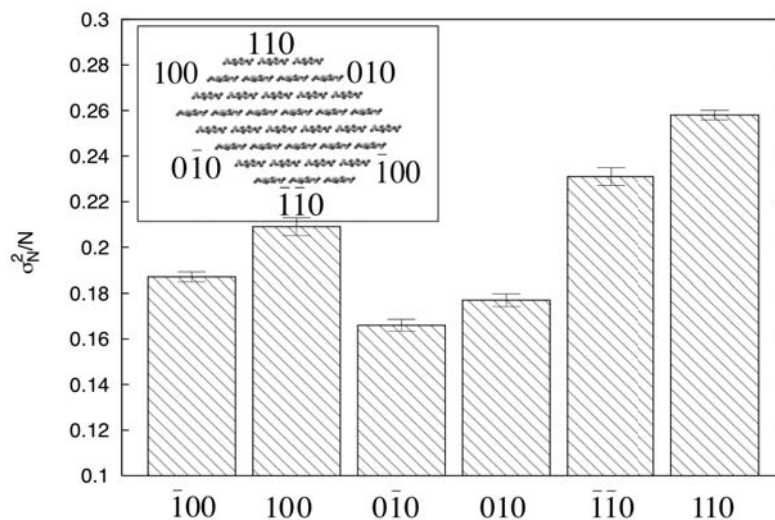
Figure 2.7. Densities of the solvation layers of both microfibrils (A). Error are of the size of the symbols. Compressibility of the solvation layers of the microfibril surfaces of cellulose I_β (B) and I_α (C). Compressibility was measured as the fluctuation in water number contained in a box 5 Å in height above each surface, normalized by the average number of coordinating waters.

(A)

36.1   30.1   34.2

34.7   41.3   34.7   26.5

20.3   15.3   14.3   14.0   21.4

41.1   21.5   21.2   22.6   24.5   33.6

25.1   13.8   13.6   13.7   14.9   30.1

28.4   24.5   24.5   23.5   30.0

22.0   13.9   15.5   19.8

33.8   30.7   26.3

(B)

29.8   33.1   27.7

26.7   17.9   18.3   21.2

26.9   15.2   14.5   14.7   22.7

32.4   15.0   14.2   13.7   13.8   23.9

20.5   15.7   13.9   14.1   14.2   35.4

19.5   16.3   15.8   14.7   21.8

21.0   18.0   13.4   22.7

31.1   31.5   27.4

Figure 2.8. Average O3-H—O5 HB angle RMSF per chain for the I$_\beta$ (A) and I$_\alpha$ (B) microfibrils. Values were calculated by averaging the angle RMSFs per interaction down a single chain.

(A)

0.09 0.14

0.39 0.31 0.18

0.82 0.83 0.83 0.81

0.27 0.55 0.54 0.51 0.36

0.82 0.83 0.83 0.84 0.77

0.33 0.52 0.53 0.47

0.81 0.81 0.82

0.02 0.01

(B)

0.09 0.05

0.64 0.65 0.64

0.68 0.70 0.71 0.67

0.68 0.70 0.71 0.71 0.66

0.59 0.70 0.71 71 0.60

0.69 0.69 0.69 0.68

0.67 0.62 0.68

0.02 0.01
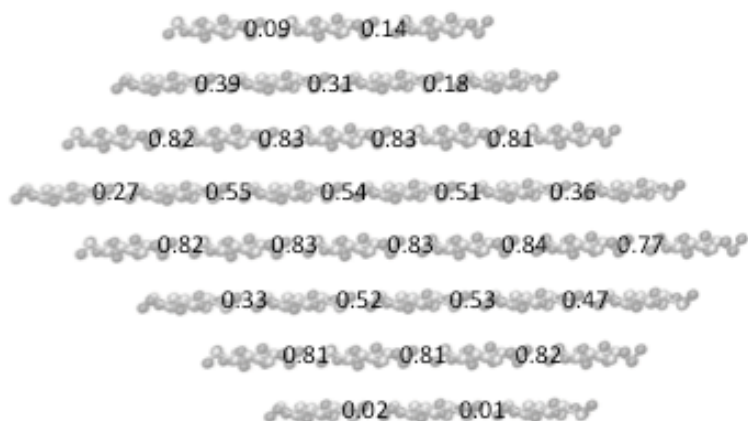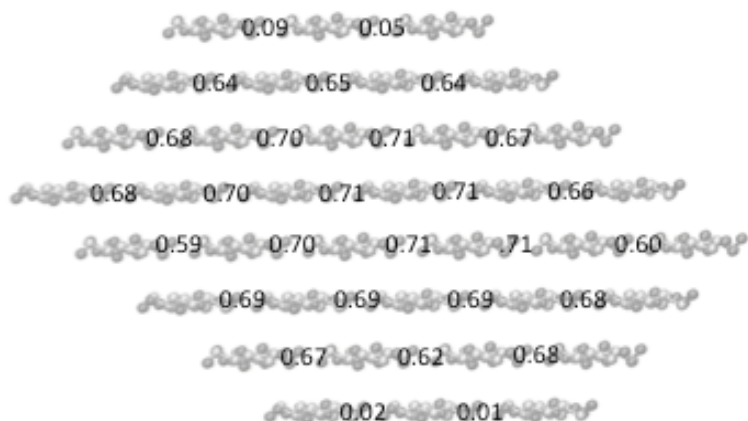
Figure 2.9. Average O6-H—O3 occupancy per chain for the I_β (A) and I_α (B) microfibrils. Individual HB occupancies were calculated with a bond length cutoff of 2.2 Å and a bond angle cutoff of 130°. Values were calculated by averaging the occupancies per interaction down a single chain.

(A)
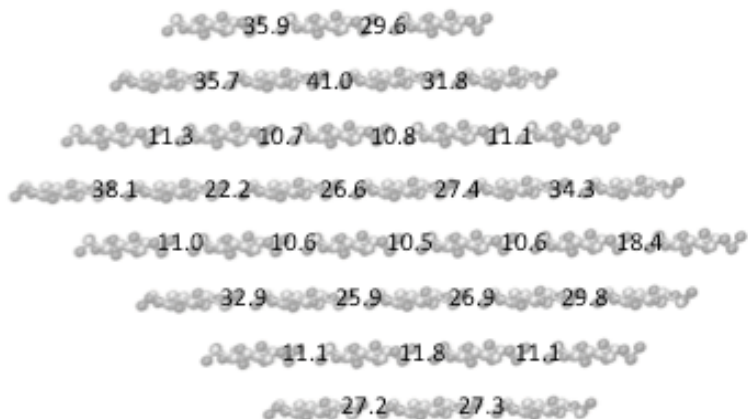
(B)
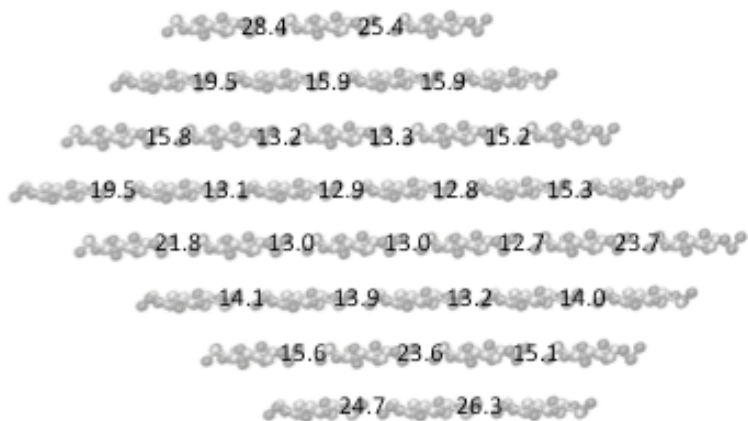
Figure 2.10. Average O6-H—O3 HB angle RMSF per chain for the I$_\beta$ (A) and I$_\alpha$ (B) microfibrils. Values were calculated by averaging the angle RMSFs per interaction down a single chain.
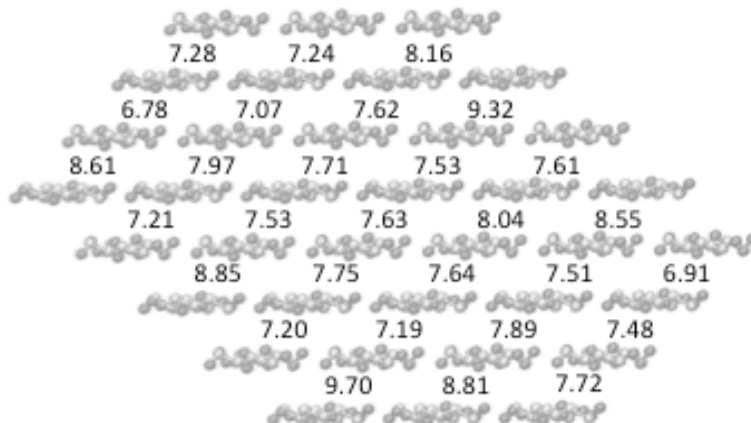
Figure 2.11. Average C2-H2—O3 HB angle RMSF per chain pair for the I$_\beta$ microfibril. Values were calculated by averaging the angle RMSFs per interaction down a single chain pair.
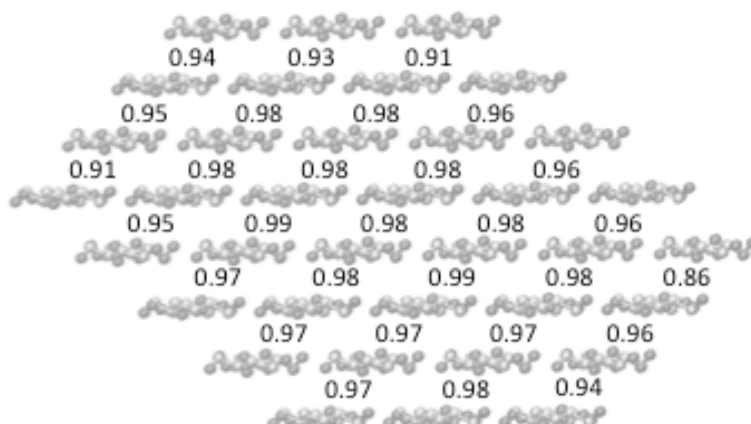
Figure 2.12. Average C2-H2—O3 HB occupancy per chain pair for the I$_\beta$ microfibril. Individual HB occupancies were calculated with a bond length cutoff of 3.2 Å and a bond angle cutoff of 110°. Values were calculated by averaging the occupancy per interaction down a single chain pair.
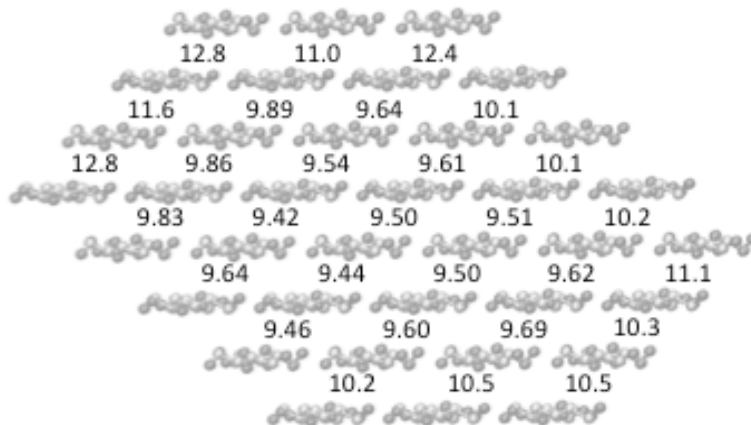
Figure 2.13. Average C2-H2—O4 HB angle RMSF per chain pair for the I_α microfibril. Values were calculated by averaging the angle RMSFs per interaction down a single chain pair.
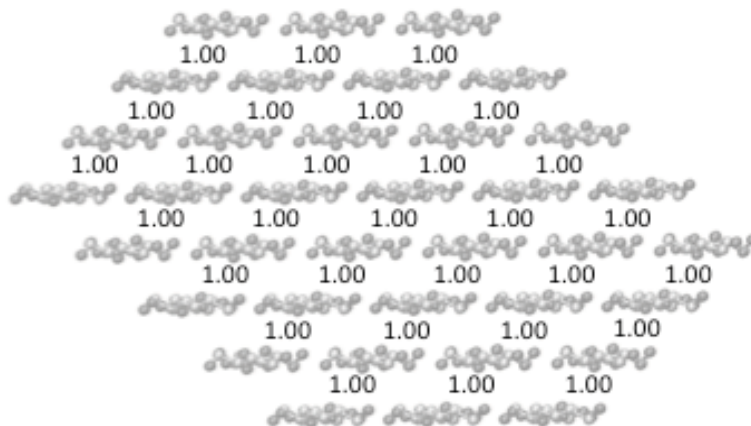
Figure 2.14. Average C2-H2—O4 HB occupancy per chain pair for the I_α microfibril. Individual HB occupancies were calculated with a bond length cutoff of 3.2 Å and a bond angle cutoff of 110°. Values were calculated by averaging the occupancy per interaction down a single chain pair.

Figure 2.15. Hydrogen bond switching behavior on the (100) surface of the I$_\alpha$ microfibril. A HB cutoff of 3.2 Å and 110° was employed. The C1-H-O6 and C3-H-O6 hydrogen bonds are the native contacts. Results are shown only for the first 500 ps, but behavior is similar for the entire simulation.

(A)

0.84 0.86 0.93

0.91 0.92 0.93 0.97

0.94 0.99 0.99 1.00 0.99

0.83 0.99 0.99 0.99 1.00 0.94

0.97 1.01 1.01 1.02 1.00 0.92

0.97 0.99 0.99 1.00 0.91

1.00 0.99 0.98 0.97

0.87 0.85 0.85

(B)

0.85 0.84 0.87

0.94 1.00 0.97 1.02

0.93 1.02 0.99 1.02 0.99

0.92 1.02 1.01 1.00 1.02 0.97

0.99 1.00 1.00 1.01 1.00 0.91

1.00 0.99 0.99 1.00 0.97

0.99 0.99 0.99 0.99

0.90 0.86 0.87
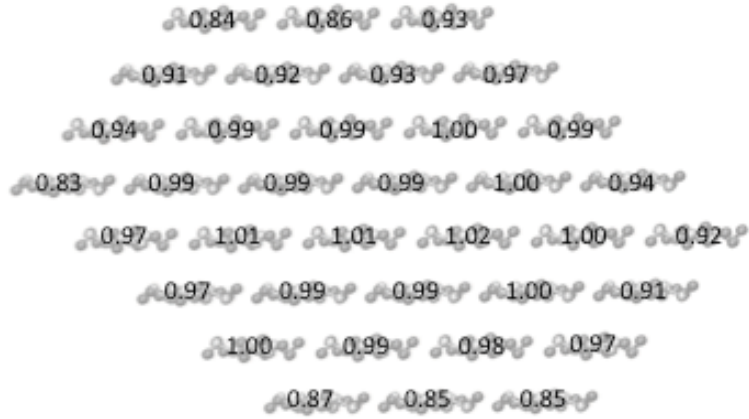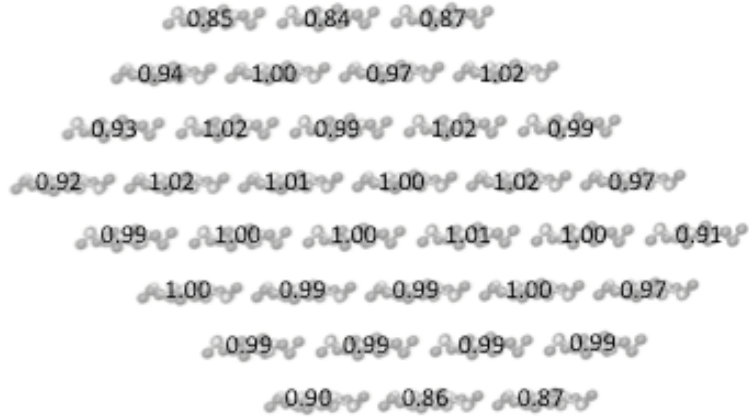
Figure 2.16. Ratio of actual number of hydrogen bonds found to the amount expected per glucose, based on glucose location in the microfibril cross-section, for both cellulose I_β (A) and cellulose I_α (B) microfibrils. Expected number of bonds was calculated using geometrical considerations. Hydrogen bond cutoffs were the same as those used in Figure 2.3 of the main text.
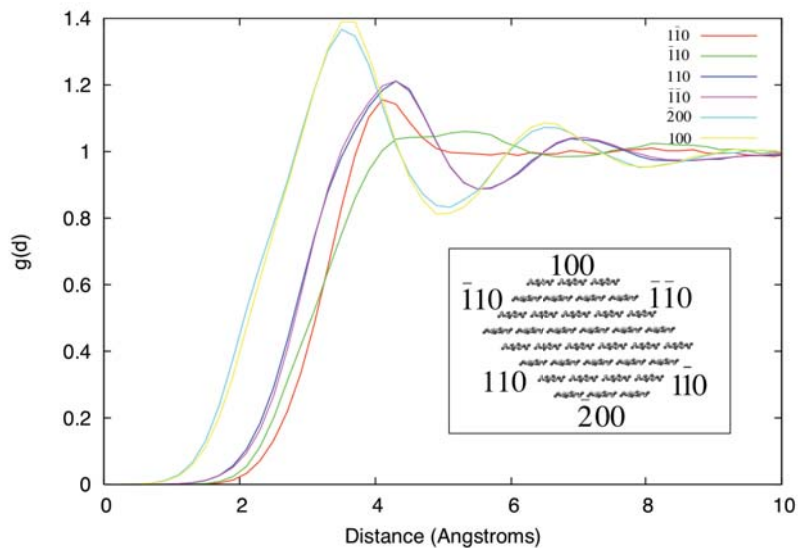
Figure 2.17. Normalized solvent water density profiles above the 6 microfibril surfaces of the I$_\beta$ microfibril.
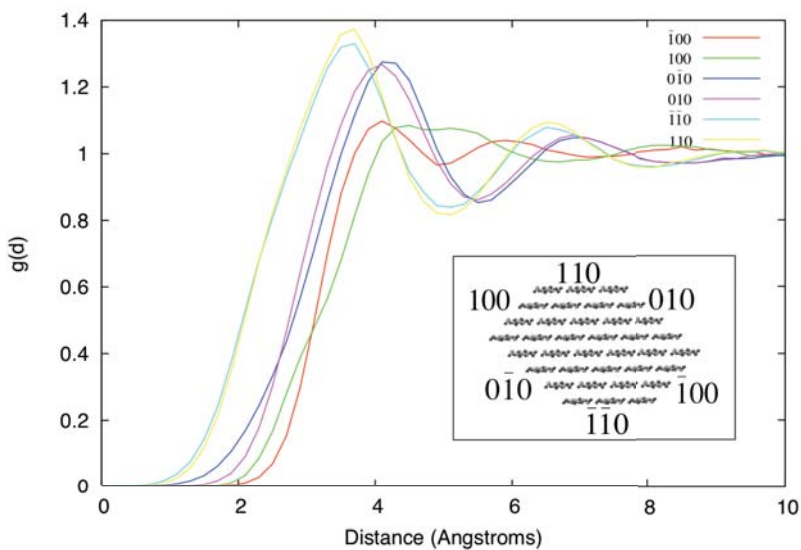
Figure 2.18. Normalized solvent water density profiles above the 6 microfibril surfaces of the I$_\alpha$ microfibril.
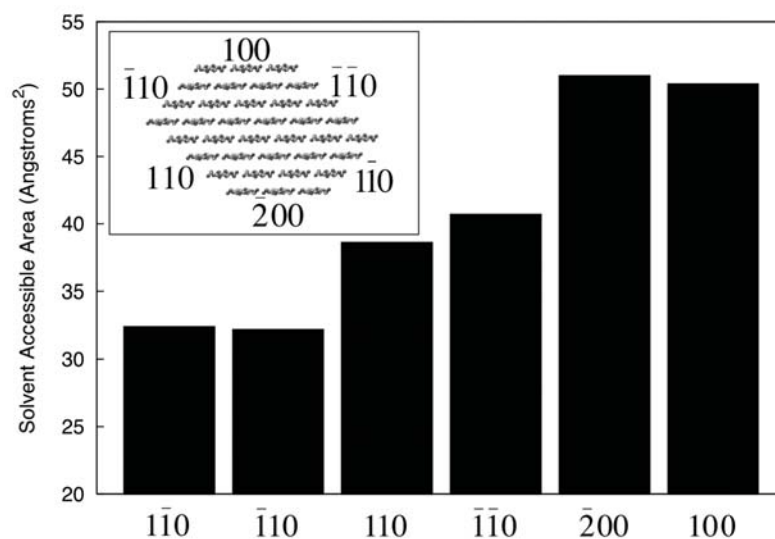
Figure 2.19. Average solvent accessible surface area per glucose for the six surfaces of the I$_\beta$ microfibril. Errors are all less than 0.25 Å^2.
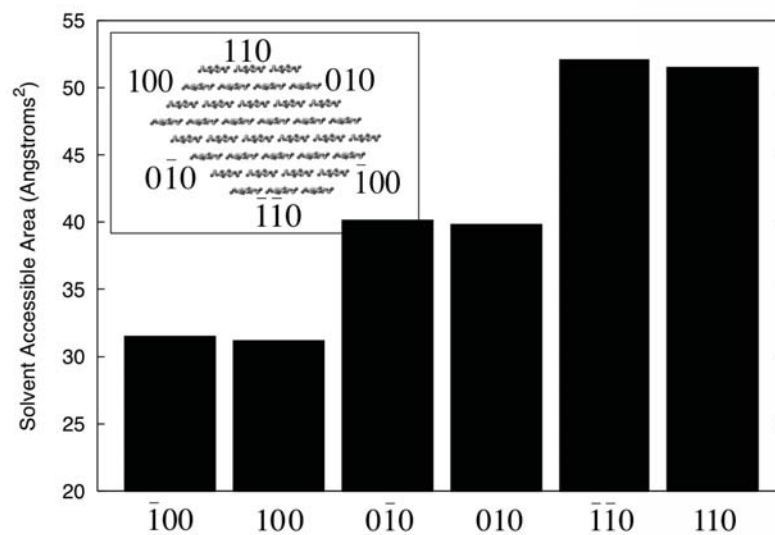
Figure 2.20. Average solvent accessible surface area per glucose for the six surfaces of the I$_\alpha$ microfibril. Errors are all less than 0.25 Å^2.

References

(1)   Perlack, R. D. *Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasability of a Billion-Ton Annual Supply*; DOE/GO-102995-2135; Oak Ridge National Laboratory: Oak Ridge, TN, 2005.

(2)   Lynd, L. R.; Wyman, C. E.; Gerngross, T. U. *Biotechnol. Progr.* **1999**, *15*, 777-793.

(3)   Sheehan, J.; Himmel, M. *Biotechnol Progr* **1999**, *15*, 817-827.

(4)   Wyman, C. E. *Trends Biotechnol.* **2007**, *25*, 153-157.

(5)   Ding, S. Y.; Himmel, M. E. *J. Agr. Food Chem.* **2006**, *54*, 597-606.

(6)   Perez, S.; Mazeau, K. In *Polysaccharides: Structural Diversity and Functional Versatility*, 2nd ed.; Dumitriu, S., Ed.; Marcel Dekker: New York, 2005; pp 41-68.

(7)   Atalla, R. H.; Vanderhart, D. L. *Science* **1984**, *223*, 283-285.

(8)   Horii, F.; Yamamoto, H.; Kitamaru, R.; Tanahashi, M.; Higuchi, T. *Macromolecules* **1987**, *20*, 2946-2949.

(9)   Nishiyama, Y.; Langan, P.; Chanzy, H. *J. Am. Chem. Soc.* **2002**, *124*, 9074-9082.

(10)   Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P. *J Am Chem Soc* **2003**, *125*, 14300-14306.

(11)   Himmel, M. E.; Ding, S. Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W.; Foust, T. D. *Science* **2007**, *315*, 804-807.

(12)   Heiner, A. P.; Sugiyama, J.; Teleman, O. *Carbohyd Res* **1995**, *273*, 207-223.

(13)   Kondo, T. In *Polysaccharides: Structural Diversity and Functional Versatility*, 2nd ed.; Dumitriu, S., Ed.; Marcel Dekker: New York, 2005; pp 69-98.

(14)   Liang, C. Y.; Marchessault, R. H. *J Polym Sci* **1959**, *37*, 385-395.

(15)   Marchessault, R. H.; Liang, C. Y. *J Polym Sci* **1960**, *43*, 71-84.

(16)   Marechal, Y.; Chanzy, H. *J Mol Struct* **2000**, *523*, 183-196.

(17)   Michell, A. J. *Carbohyd Res* **1993**, *241*, 47-54.

(18)   Sugiyama, J.; Persson, J.; Chanzy, H. *Macromolecules* **1991**, *24*, 2461-2466.

(19)   Tashiro, K.; Kobayashi, M. *Polymer* **1991**, *32*, 1516-1530.

(20)   Hinterstoisser, B.; Akerholm, M.; Salmen, L. *Biomacromolecules* **2003**, *4*, 1232-1237.

(21)   Watanabe, A.; Morita, S.; Ozaki, Y. *Biomacromolecules* **2006**, *7*, 3164-3170.

(22)   Watanabe, A.; Morita, S.; Ozaki, Y. *Biomacromolecules* **2007**, *8*, 2969-2975.

(23)   Kondo, T. *J Polym Sci Pol Phys* **1997**, *35*, 717-723.

(24)   Bergenstrahle, M.; Berglund, L. A.; Mazeau, K. *J Phys Chem B* **2007**, *111*, 9138-9145.

(25)   Shen, T. Y.; Gnanakaran, S. *Biophys J* **2009**, *96*, 3032-3040.

(26)   Jeffrey, G. A. *An introduction to hydrogen bonding*; Oxford University Press: New York, 1997.

(27)   Steiner, T.; Saenger, W. *J Am Chem Soc* **1992**, *114*, 10146-10154.

(28)   Taylor, R.; Kennard, O. *J Am Chem Soc* **1982**, *104*, 5063-5070.

(29)   Steiner, T. *Chem Commun* **1997**, 727-734.

(30)   Calhorda, M. J. *Chem Commun* **2000**, 801-809.

(31)   Koch, U.; Popelier, P. L. A. *J Phys Chem-Us* **1995**, *99*, 9747-9754.

(32)   Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-935.

(33)   Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187-217.

(34)   Hoover, W. G. *Phys Rev A* **1985**, *31*, 1695-1697.

(35)   Nose, S.; Klein, M. L. *Mol Phys* **1983**, *50*, 1055-1076.

(36)   Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613-4621.

(37)   Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; Mackerell, A. D. *J. Comput. Chem.* **2008**, *29*, 2543-2564.

(38)   Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. *J. Chem. Theory Comput.* **2009**, *5*, 2353-2370.

(39)   Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press; Oxford University Press: Oxford England; New York, 1987.

(40)   Rowland, S. P.; Roberts, E. J. *J Polym Sci A1* **1972**, *10*, 867-879.

(41)   Vietor, R. J.; Newman, R. H.; Ha, M. A.; Apperley, D. C.; Jarvis, M. C. *Plant. J.* **2002**, *30*, 721-731.

(42)   Matthews, J. F.; Skopec, C. E.; Mason, P. E.; Zuccato, P.; Torget, R. W.; Sugiyama, J.; Himmel, M. E.; Brady, J. W. *Carbohyd Res* **2006**, *341*, 138-152.

(43)   French, A. D.; Miller, D. P.; Aabloo, A. *Int J Biol Macromol* **1993**, *15*, 30-36.

(44)   Chandler, D. *Nature* **2005**, *437*, 640-647.

(45)   Godawat, R.; Jamadagni, S. N.; Garde, S. *P Natl Acad Sci USA* **2009**, *106*, 15119-15124.

(46)   Sarupria, S.; Garde, S. *Phys Rev Lett* **2009**, *103*, 037803.

(47)   Patel, A. J.; Varilly, P.; Chandler, D. *J Phys Chem B* **2010**, *114*, 1632-1637.

(48)   Cousins, S. K.; Brown, R. M. *Polymer* **1995**, *36*, 3885-3888.

Chapter 3

The Thermodynamics of Cellulose Solvation in Water and the Ionic Liquid 1-Butyl-3-Methylimidazolim Chloride

Abstract

Cellulose is present in biomass as crystalline microfibrils held together by a complex network of intermolecular interactions, making it difficult to initiate its hydrolysis and conversion to fuels. While cellulose is insoluble in water and most organic solvents, complete dissolution of cellulose can be achieved in certain classes of ionic liquids (ILs). The present study was undertaken to analyze the thermodynamic driving forces of this process and understand how the anions and cations comprising an IL interact with the different moieties of glucose residues to cause dissolution. All-atom molecular dynamics (MD) simulations were performed at two extreme states of cellulose dissolution, a crystalline microfibril and a dissociated state in which all the glucan chains of the microfibril are fully separated from each other by at least four solvation shells. MD simulations of the two states were carried out in water and in the IL 1-butyl-3-methylimidazolium chloride (BmimCl) to provide a comprehensive analysis of solvent effects on cellulose dissolution. The results reveal two important molecular aspects of the mechanism of cellulose dissolution. The first is that the perturbation of solvent structures by the dissolved glucan chains can be a crucial factor in determining solubility, particularly for the insolubility of cellulose in water at 300 K. Second, both Cl^- and the $Bmim^+$ ions of BmimCl interact with the moieties of glucose residues that form intersheet contacts, the most robust component in the interaction network of crystalline cellulose. Cl^- anions can form hydrogen bonds (HBs) with the hydroxyl groups of glucan chains from either the equatorial or axial directions. For $Bmim^+$ cations, the calculated density profiles reveal that the contacts with glucan chains along the axial directions are closer than those along the equatorial directions. Based on the results of atomistic MD simulations, we propose that interacting with glucan chains along axial directions and disrupting the intersheet contacts of cellulose is an important ability of cellulose pretreatment solvents.

3.1 Introduction

Lignocellulosic biomass has the potential to serve as a renewable feedstock for the production of transportation fuels.[1-8] One of the primary routes to fuels under consideration involves the isolation of the cellulosic and hemicellulosic components of biomass, hydrolysis of these biopolymers to produce glucose and xylose, and subsequent fermentation of these sugars to produce ethanol or butanol. Alternatively, the sugars can be processed with chemical catalysis to produce a broad spectrum of products compatible with gasoline and diesel fuels. Releasing the carbohydrate fraction of biomass remains a major challenge since the cellulosic component is present in the form of microfibrils surrounded by amorphous hemicellulose, which in turn is encased in a wax-like lignin sheath.[9] While hemicelluloses and lignin are amorphous materials, cellulose is crystalline, adding to the difficulty of processing this component. As a result, all current approaches for biomass conversion require harsh and expensive physical and/or chemical

pretreatment methods,[5,8,10,11] and consequently there is considerable incentive to find alternative pretreatment technologies that could reduce the cost of converting lignocellulosic biomass to fuels.[10,12,13]

Recent studies have shown that all of the components of biomass can be dissolved in ionic liquids (ILs),[14-16] salts that have melting points below 373 K. Of particular interest is the fact that crystalline cellulose can be dissolved in ILs, since the glucan strands of cellulose are held together by a robust interaction network that includes specific interactions consisting of classical OH—O hydrogen bonds (HBs) and nonconventional CH—O contacts that together help to render the material insoluble in water and most organic solvents.[17,18] The intermolecular OH—O HBs, which are part of cellulose's interchain interactions, act in the glucans' equatorial directions and serve to bind neighboring sugar chains together to form flat sheets.[18-28] The CH—O contacts form along the axial directions of glucose residues and contribute to the intersheet interactions holding glucan sheets together in crystalline cellulose microfibrils. In molecular dynamics (MD) simulations of cellulose microfibrils, the CH—O contacts of intersheet interactions were observed to be highly intransigent to surface exposure[18] or microfibril twist,[29] opposite the behaviors of the OH—O HB components of interchain interactions. The calculated glucan-glucan interaction energies in cellulose microfibrils are stronger for intersheet interactions than for interchain interactions,[18] and path optimization and free energy simulations of cellulose deconstruction further affirm that intersheet interactions are the main cause of recalcitrance.[30] The signature of stronger intersheet than interchain interactions are also revealed in restrained dissociations of small, soluble model carbohydrates[31-33] which showed that association energy or free energy via axial stacking is greater than that via equatorial pairing.

While there are innumerable anion-cation pairs that might be considered as solvents for cellulose, the best ILs are those that contain a heterocyclic ring (imidazole or pyridine) substituted with alkyl side-chains as the cation (e.g., 1-butyl-3-methylimidazolium) and a conjugate inorganic base anion (e.g., chloride, acetate).[14,34,35] With such ILs, cellulose solubilities up to 25 wt % have been achieved at moderate temperatures.[14] However, little is known about the means by which ILs promote the dissolution of cellulose. The predominant explanation is that the anions of the IL form hydrogen bonds with the hydroxyl groups of the glucan units, thereby weakening the hydrogen bonding between glucan strands. This theory is supported by several experimental and computational investigations that have used water-soluble substrates as model solutes. NMR relaxation experiments carried out with solutions of glucose or cellobiose in 1-butlyl-3-methylimidazolium chloride (BmimCl) show a strong stoichiometric correspondence between IL anions and glucan hydroxyl groups.[36,37] The formation of OH—anion HBs has also been observed in MD simulation studies,[30,38-40] and the interaction energy between a glucan residue and an IL anion has been found to be stronger than that between a glucan residue and a water or methanol molecule in the liquid phase.[41] While a strong dependence of cellulose solubility on the hydrogen-bonding properties of anions is observed,[14,35] OH—anion HBs alone are not sufficient to explain the solubility of cellulose in different ILs. The cation identity also affects the solubility. The length of the alkyl chains and the structure of the charge-bearing ring of the cation (e.g. imidazolium, pyridinium) both change cellulose solubility.[14,34] Since a definite set of cation-mediated modes of interaction has not yet been identified, the roles

of cations are considered mostly non-specific or indirect.[36,37,39] Furthermore, the strong electrostatic coupling between anions and cations indicate that their interactions with cellulose are likely collective. The multifaceted nature of IL molecules complicates the characterization of their interactions with cellulose and an integrated approach is thus needed.

The present investigation was undertaken in order to elucidate the IL-mediated interactions leading to cellulose dissolution to establish a molecular basis for improving IL-based pretreatment strategies or designing novel pretreatment solvents. All-atom MD simulations of cellulose were carried out for two different states, a crystalline microfibril and a state in which the glucan chains of the microfibril are fully dissociated. These two extremes were taken to represent crystalline and dissolved cellulose, and are used to understand the modes by which the anions and cations of the IL interact with the glucan strands. Each state was solvated in water or BmimCl to compare and contrast how the two solvents modulate the relative energies of the microfibril and dissociated states. The solvent structures around glucan chains and on the surfaces of the microfibril were examined to study the specific interactions mediated by the cations and anions of the IL.

3.2 Methods

An all-atom force field of BmimCl was developed using a standardized parameterization prodecure[42-44] for compatibility with the existing CHARMM[45] carbohydrate force fields,[46,47] which have been validated against thermodynamic and structural data for several monosaccharides[46] and cellulose.[18] In short, atomic charges and intramolecular potentials were obtained by fitting against the interaction energies and geometries obtained from *ab initio* calculations at the HF/6-31G* level. The resulting force field was then validated by comparing the calculated thermodynamic properties with experimental measurements. The reported thermodynamic data of BmimCl include liquid density at 360 K and 1 atm[48] and heat capacities at different temperatures,[49] all of which were used for force field development. The atomic and intramolecular parameters were adjusted iteratively to ensure accuracy of both molecular structures and thermodynamic properties. A comparison of the results obtained using our optimized force field with the available experimental data is shown in Figure 3.1. Although the density is predicted accurately by our BmimCl force field, the heat capacities are consistently overestimated by ~10%. Classical force fields often overestimate heat capacity compared to *ab initio* calculations due to the use of harmonic potentials.[50] Additionally, since the anion and cation retain their formal charges in our model, the lack of polarizability may lead to enhanced cage effects and energy variations, causing the small overestimation of heat capacity.[51,52] However, in this work, MD simulation results are used only for energy comparisons between states of cellulose dissolution, which mitigates the biases of any single trajectory caused by the use of fixed-charge force fields, or for structure characterizations. Previous work has shown that the use of fixed-charge models in MD simulations is able to capture IL structures in the liquid state accurately as compared to those measured by neutron diffraction.[53,54] More details of the development and validation of the force field are provided in the supporting information (SI).

Two model states of cellulose were constructed to contrast the thermodynamic and solvation properties of crystalline and dissolved cellulose. Both states contained 36 glucan chains, each with a length of 16 glucan units. The crystalline state was taken to be a cellulose I_β[55] microfibril.[9,18,56] The I_β allomorph is the most stable form of naturally occurring cellulose.[57,58] In the fully dissociated state, the 36 chains were arranged into a 6-by-6 grid with a spacing of at least 4 solvation shells between chains (16 Å in water and 30 Å in IL). Cross sections of the two model states are shown in Figure 3.2A-B. Both states were solvated in either TIP3P[59] water or BmimCl and have the same number of total atoms and composition for a given solvent choice. The cellulose/water simulations contained 28443 waters, and the cellulose/IL simulations contained 12180 BmimCl pairs. These correspond to a mass fraction of cellulose of 0.15 in water and 0.04 in BmimCl, well within the solubility range of the latter system.[14,35]

Simulations were performed with the NAMD software[60] at constant temperature and pressure (1 atm) conditions using the Langevin thermostat and Nose-Hoover Langevin barostat.[61,62] The SHAKE algorithm was used to constrain the length of all covalent bonds to hydrogen.[63] Cellulose/water trajectories were generated at 300 K, 325 K, and 350 K, and cellulose/IL trajectories at 425 K, 450 K, and 500 K. We simulated the BmimCl systems at temperatures elevated compared to the melting point (340 K[64]) for accelerated dynamics. At the start of all simulations, harmonic restraint potentials were placed on all C1 sugar carbons (force constant 5 kcal/mol/Å^2 for water systems and 10 kcal/mol/Å^2 for IL systems) to keep the glucans around their initial positions. After 15000 steps of energy minimization, the systems were heated gradually to their target temperatures and equilibrated for at least 5 ns for the cellulose/water systems and 15 ns for the cellulose/IL systems. After equilibration, the harmonic restraints were removed except for those on the terminal glucans of each chain to retain the system in the microfibril or fully dissociated state. Since cellodextrins in water exist in extended chain conformations,[32,65,66] the effects of the remaining restraints on biasing the conformations of the individual glucan chains are likely insignificant. A comparison of the glycosidic linkage torsion angles sampled in our trajectories with those from the simulation of free short-chain cellooligomers in water[65] shows a similar distribution (See the SI for further details). We also conducted additional simulations in which all harmonic restraints were removed after equilibration, i.e., 'free' simulations, to examine chain collapse behaviors in water and in BmimCl. For all simulations, a time step of 2 fs was employed, and the particle mesh Ewald (PME) sum was used to calculate long-range electrostatic interactions. Data was taken from production runs of at least 15 ns with frames saved every 1 ps for analysis from the cellulose/water trajectories and from runs of at least 30 ns with frames saved every 2 ps from the cellulose/IL trajectories. The simulation lengths of both system types were examined to ensure the convergence of solvent structures by comparing the solvent density profiles calculated from sub-portions of the trajectories. See the SI for further details.

Three-dimensional plots for the number density of solvent molecules around dissociated glucan chains were generated in the following manner. First, a local Cartesian coordinate system was created for each glucan in a trajectory frame. For a given glucan, the vector that connects the two O4 ether oxygens that bound it defines the z-axis. A pseudo-x-axis was defined as the average directions of the C2-O2 and C3-O3 bonds. The cross product of the z- and pseudo-x-axes then defines the y-axis, and the cross product

of the y- and z-axes defines the x-axis. This methodology assigns the x-axis along the equatorial direction and the y-axis along the axial direction of the selected glucan. The origin of the coordinate system was chosen to be the center-of-mass of the glucan ring. A schematic of the process for creating the glucan-centered coordinate system is shown in the SI (Figure 3.17A). After creating a local Cartesian coordinate system for a glucan, the space around the glucan was discretized into small cubic volumes of side length 0.25 Å, and the occupancy of each grid point was calculated. This process was repeated and averaged over all glucans and frames in a trajectory to determine the three-dimensional number density distributions, which are normalized by the density of pure solvent at the same temperature and pressure. To determine locations of the non-monatomic solvent molecules, the center of mass of the whole molecule was used for water, while the center of mass of the imidazolium ring, the solitary methyl group, and the first methyl group off the imidazolium ring in the butyl group was used to represent the charge-carrying portion of the Bmim$^+$ cation, which we name the ring site. The center of mass of the remaining atoms of the Bmim$^+$ cation is then the tail portion. The distribution of Bmim$^+$ tails is presented in the SI (Figure 3.20).

To calculate the solvent density profile as a function of cylindrical radius around a glucan chain, i.e. the cylindrical radial distribution function (RDF), the center of mass of each glucan ring was determined. The vector connecting the centers of masses of sequential glucans is defined as the z-axis of the cylinder associated with a pair of neighboring sugar units. The length of the inter-center-of-mass vector defines the length of the cylinder within which the density profile along the radial direction was calculated. The distance from each solvent molecule to the local z-vector was then calculated, binned, and averaged over all neighboring glucan-glucan pairs and frames to determine the cylindrical RDF. A schematic representation of the construction of a cylindrical coordinate system for a glucan pair is shown in the SI (Figure 3.17B).

3.3 Results and Discussion

3.3.1 The Insolubility of Cellulose in Water

Figure 3.3 shows the differences in the average total potential energies between the dissociated and microfibril states of cellulose in water and BmimCl at several temperatures, normalized by the number of glucans. In water, the two states are nearly isoenergetic at 300 K. At higher temperatures, the dissociated state has increasingly higher potential energies in water, as the heat capacity of the more disordered state is larger.[67] The calculated energy differences from simulating these polymeric materials agree semi-quantitatively with the measured solvation enthalpies of monosaccharides, ~1 kcal/mol, over the investigated temperature range.[68,69] The increase with temperature of the potential energy of the dissociated state relative to the microfibril state is also in line with the temperature dependence of the solvation enthalpy of glucose.[68] Higher potential energies of the dissociated state at elevated temperatures represent part of the thermodynamic penalty of dissolving cellulose in water.

Although the microfibril and dissociated states of cellulose are isoenergetic at 300 K in water, collapse of glucan chains was observed in the free simulation after removing the restraint potentials at the chain ends used for keeping the glucan chains in a

dissociated state. The amorphous aggregate at the end of the free simulation is shown in Figure 3.4A. The number of glucan-glucan contacts, defined as the number of glucan atoms within 3 Å of any other glucan atom in any other chain, increased by a factor of two over the length of the free simulation (Figure 3.4B). Analyzing the different components of interaction energies in the system during chain collapse revealed that the loss of glucan-water contacts increased the glucan-water interaction energy (Figure 3.4B), but was compensated by enhanced water-water and glucan-glucan interactions (Figure 3.4C). As a result, the total potential energy does not show any significant drift during the free simulation (Inset, Figure 3.4C). The magnitude of the change in energy of the system, ~40 kcal/mol, as calculated from the differences in the potential energies of the first and last 10 ns of the trajectory, is small compared to the size of a typical potential energy fluctuation, 270 kcal/mol (The running average of the potential energy of the system is included in the SI, Figure 3.22). Thus chain collapse is considered nearly isoenergetic at 300 K. This result is in line with the data in Figure 3.3, which show that the microfibril and dissociated states of cellulose are also nearly isoenergetic at 300 K. Therefore, the chain collapse in water near room temperature observed in our atomistic MD simulations is considered entropy driven, and solvent structures near different states and conformations of cellulose are expected to play an essential role in the insolubility of cellulose in water.[18] A similar 'free' simulation performed at 300 K with a smaller system, 18 10-mer glucan chains in water, showed identical results, i.e. a collapse of the chains with no drift in total potential energy (See the SI for further information). To analyze how glucan chains affect solvent structures, we discuss next the number density of water molecules around the dissociated glucan chains and on the surfaces of a microfibril.

In Figure 3.5A-B, the three-dimensional distribution of water density around the dissolved glucan chains is plotted as a contour map. Regions of high water density can be observed around the interchain OH—O HB forming OH groups on both the equatorial and axial sides of the glucan ring.[18,70] In addition to regions near the HB donors and acceptors of the glucan units however, high water density was also observed above and below the planes of the glucan rings near the CH groups protruding outward in the axial direction (Figure 3.5A-B). Inspection of water structures in this region reveals that the lack of hydrogen bonding ability with the aliphatic CH groups gives rise to a denser water-water interaction network nearby, a signature of small-scale hydrophobic effects.[71] The amphiphilicity of glucan chains thus results in a peculiar ordering of water around the dissociated cellulose. We also calculated the average density profiles of water as a function of the cylindrical radius around a dissociated chain, and the results are shown in Figure 3.5C. The two peaks in the first shell at 3.8 and 4.8 Å correspond to the high water density regions in the axial and equatorial directions of a glucan chain, respectively.

The amphiphilicity of glucan units and the distribution pattern of exposed groups on the surfaces of a cellulose microfibril give rise to distinctive water structures at different solid-liquid interfaces. On the hydrophobic surfaces that expose CH groups, significant water ordering was observed, whereas on the hydrophilic surfaces of a microfibril, which expose mostly hydroxyl groups, solvent ordering was much less prominent.[18] Since the water-ordering CH groups in the dissolved state of cellulose are solvent accessible, chain collapse occurred spontaneously in the free simulation to reduce their exposure. This entropy-based mechanism is supported by the observation that the total potential energy

does not drift during chain collapse (Figure 3.4C). Our results indicate that the amphiphilicity of glucan chains and their tendency to restrict available water network structures are dominant causes of the insolubility of cellulose in water.

3.3.2 Dissolution of Cellulose in BmimCl

Figure 3.3 shows that the potential energy of the dissociated state of cellulose in BmimCl is lower than that of the microfibril state, indicating more favorable interactions of BmimCl molecules with dissociated glucan chains than with the microfibril. Increasing temperature reduces the difference in potential energy since the more disordered state has a higher heat capacity, but for all investigated temperatures, the dissociated state has a lower potential energy in BmimCl, indicating the enthalpic driving force for cellulose dissolution in the IL. In the free simulation in BmimCl at 450 K, the dissociated glucan chains showed no sign of collapse and remained separated over a 45 ns MD simulation. These results are consistent with the observation that BmimCl can dissolve cellulose at the concentration (4 wt %) used in the present simulations.

To analyze the interactions of BmimCl molecules with glucan chains in the dissociated state, we calculated the three-dimensional density distribution of Cl^- anions and $Bmim^+$ cations around the dissolved glucan chains. The average Cl^- density around a dissolved glucan unit, normalized by the bulk value, is shown in Figure 3.6A,C. As expected from NMR measurements and MD simulations of glucose or cellobiose molecules dissolved in IL,[36-39] Cl^- anions form HBs with the OH groups of dissolved glucan chains and regions of high Cl^- density are evident around the hydroxyl groups. Well-defined regions of high Cl^- density can be observed around the O2-H and O3-H groups as well as in between them (Figure 3.6A,C). Near the C6 hydroxymethyl group, which samples different rotameric states in solution, discrete regions of high Cl^- density can also be found about the circumference of the C6-O6-H rotation. We found that Cl^- anions can form individual OH—Cl^- HBs with each hydroxyl group in a glucan unit as well as shared HBs between neighboring OHs in the same residue or across two adjacent residues along a glucan chain. Calculation of the coordination number of Cl^- up to the first minimum in the sugar-anion radial distribution function gives two anions per glucan and an average anion-to-OH ratio of 2:3. This value is similar to the reported 4:5 anion-to-OH ratio from MD simulations of glucose in 1,3-dimethylimidazolium chloride[39] and the 1:1 ratio derived from NMR measurements of cellobiose solvated in BmimCl[36] in that it shows a strong stoichiometric association between the sugar OH groups and the Cl^- anion. The anion-to-OH ratio is lower here than for those with glucose or cellobiose due to the use of polymeric solutes in our simulations (More discussion is provided in the SI).

While the density distribution of Cl^- anions establishes their formation of HBs with the equatorial hydroxyl groups of glucose residues, the density distribution of the centers of masses of $Bmim^+$ cation rings shown in Figures 3.6B,D reveals closer contacts to the axial moieties of glucan chains than to the equatorial ones. In Figure 3.6D, large volumes of high-density regions of $Bmim^+$ rings are observed over the glucan rings of the chains. Next to the axial moieties of glucose residues, the high-density regions of Cl^- and $Bmim^+$ rings have similar distances to the glucans, whereas near the equatorial groups, the high-density zones of the cations are further away from the sugars than those of the anions (more details are provided in the SI). Therefore, both Cl^- and $Bmim^+$ rings interact

with glucan chains along axial directions, but close equatorial contacts are dominated by Cl$^-$ anions. In the MD simulations of the dissociated and microfibril states, the observed signatures of ring-ring stacking between the Bmim$^+$ cations and sugar rings in dissociated glucan chains or those packed on a cellulose microfibril are weak, in agreement with earlier simulations of dissolved glucose molecules in BmimCl[39] (see the SI for further discussion). It is also clear from Figure 3.6 that the Bmim$^+$ and Cl$^-$ density are arranged in a highly complementary manner, each ion occupying specific regions near a glucan chain while maintaining local charge neutrality. This result, combined with the lower potential energy of the dissociated state of cellulose compared to the microfibril state in BmimCl, indicates that the specific interactions between BmimCl and the amphiphilic glucan chains are compatible with the electrostatic coupling between cations and anions.

The liquid-state structures of BmimCl around a glucan chain can be further illustrated by cylindrical density profiles. In Figure 3.7, we plot the cylindrical density profiles of the Cl$^-$ anion and the three electrophilic hydrogen atoms of the Bmim$^+$ ring. It is clear that the density distribution around a glucan chain of BmimCl atoms with opposite partial charges oscillate and plateau after around ~14 Å from the chain. Converting the structural distributions of BmimCl atoms around a glucan chain into a charge distribution based on their partial charges clearly shows that electrostatic coupling strongly influences the liquid-state structures of BmimCl. The first layer of high positive charge density around a glucan chain in Figure 3.7 results from the three electrophilic hydrogen atoms on the Bmim$^+$ ring. This result is also in line with the observation shown in Figure 3.6B,D that the centers-of-mass of the Bmim$^+$ rings are positioned near the ether oxygen atoms, which carry a negative partial charge. The second layer of high charge density around a glucan chain is negative and comes from Cl$^-$ anions as expected. Further away from the glucan chains, alternative layers of positive and negative charges can be observed and are coherent with the density profiles of Cl$^-$.

In both water and BmimCl, the molecular structures of the amphiphilic glucan chains modulate the nearby solvent structures. The structure of the hydrogen-bonding network of water molecules depends on their proximity to the polar hydroxyl groups or the aliphatic CH groups of the glucan chains. In BmimCl, anions form HBs with hydroxyl groups whereas the Bmim$^+$ rings of cations have closer contacts with the ether oxygen atoms and CH groups along the axial directions of glucose residues than with those groups in equatorial positions. The tail portions of the cations have a similar pattern of contacts with glucan chains as the Bmim$^+$ rings. The electrostatic coupling of BmimCl also leads to oscillatory charge density profiles around the glucan chains. Such spatial distributions of solvent molecules are expected to affect the diffusion of catalyst molecules to the ether linkages as well as the complexation of cellulase enzymes with glucan chains. Therefore, the reactivity and stability of molecular catalysts or enzymes with glucan strands can be expected to depend on the solvation structures and charge distributions around the dissolved chains. Thus, in addition to the dissolution of cellulose, the amphiphilicity of glucan chains and the surrounding arrangement of solvation structures also have significant impacts on the hydrolysis of the dissolved polymer chains.

3.4 Conclusions

We have performed all-atom MD simulations with the aim of understanding the molecular origin of the insolubility of cellulose in water and its solubility in BmimCl. Two extreme states of cellulose were considered, a crystalline microfibril state and a dissociated state in which the glucan chains are fully detached from each other (Figure 3.2). Comparing the average potential energies of the two states reveals a thermodynamic driving force for the dissolution process: the dissociated state has an equal or higher potential energy in water and a lower potential energy in BmimCl (Figure 3.3). Furthermore, starting from the dissociated state, chain collapse occurs spontaneously in water at 300 K, during which the potential energy does not drift away from its initial value. This result, combined with the analysis of water structures near the glucan chains (Figure 3.5), indicates that the insolubility of cellulose in water originates mostly from reduction in solvent entropy.

The Cl$^-$ anions of BmimCl form strong HBs with the hydroxyl groups of the glucan chains of cellulose in both the equatorial and axial orientations, which would help to replace the interchain as well as intersheet interactions of cellulose lost upon dissolution. The Bmim$^+$ cation rings also interact with glucan chains, with the contact distances to glucose residues along axial directions smaller than those along equatorial directions. Therefore, cation-glucan interactions also appear to be an important compensator for lost cellulose intersheet interactions. Furthermore, a clear complementarity in the density profiles of Bmim$^+$ and Cl$^-$ is observed around glucan chains. This result and the observation that the dissociated state of cellulose has a lower potential energy than the microfibril state in BmimCl indicate that the electrostatic coupling between Bmim$^+$ and Cl$^-$ is compatible with the presence of aliphatic glucan chains.

Together, the results presented in this work highlight two mechanistic insights for the dissolution of cellulose by solvent-mediated interactions. First, in water, the perturbation of solvent structures by dissolved glucan chains is a dominant factor in limiting solubility by leading to the entropic penalty of the free energy cost of dissolution. Second, in BmimCl, in addition to interacting with the equatorial moieties of glucose residues that form interchain interactions in cellulose, both the Cl$^-$ and the Bmim$^+$ ions were found to interact with glucose residues along their axial directions, along which the intersheet interactions in cellulose are formed. Cl$^-$ anions can form HBs with hydroxyl groups from the axial directions of sugar rings while Bmim$^+$ cations make contacts with the donors and acceptors of intersheet CH—O HBs. Solvent molecules contacting glucose moieties along axial directions is important since intersheet interactions are the most intransigent component in the interaction network of cellulose.[18] Therefore, we propose that the formation of solvation environments that are effective in interacting with glucan chains along both equatorial and axial directions, as in BmimCl, is likely an indicative signature of an effective pretreatment solvent.

3.5 Acknowledgements

3.6 Supporting Information

3.6.1 Force Field Validation

Equation 3.1 gives the basic energy function for the CHARMM force field. Initially, partial charges and Lennard-Jones, bond, angle, and dihedral parameters were taken from established CHARMM force fields for similar molecular moieties. To develop our BmimCl force field, we followed the methodology outlined in the CHARMM literature,[42-44] which involves matching atomic, molecular, and thermodynamic properties with *ab initio* or experimental results in an iterative manner. Atomic properties were varied in order to match the geometries and interaction energies of several minimum energy TIP3P water/BMIM$^+$ complexes found from *ab initio* calculations at the HF/6-31G* level of theory with Q-CHEM.[72] Figure 3.8 shows the five BMIM$^+$ interaction regions we probed via water complexation. In Table 3.1 we compare the results of the quantum calculations with those from CHARMM, which show good agreement overall. Next, we adjusted harmonic bond and angle constants to match the vibrational structure of the BMIM$^+$ cation in CHARMM with the same for BMIM$^+$ from HF/6-31G*/MP2 *ab initio* calculations, shown in Figure 3.10. Last, we adjusted the force field dihedral parameters to match the minimum energy torsion angle profiles calculated from HF/6-31G*/MP2 quantum calculations. We show two examples in Figures 3.11A-B. Once the force field was constructed, we validated it against thermodynamic density and heat capacity data, as discussed in the main text. If our simulation results did not agree with the experimental data, we repeated through the entire process again until agreement was reached. Also, we only adjusted the parameters of the Bmim$^+$ cation. The parameters of the Cl$^-$ anion were kept the same as in the standard CHARMM force field. The final force field is found in Table 3.2 and atomic labeling of BMIM$^+$ in Figure 3.9.

3.6.2 Cl$^-$ Density Around the Dissolved Glucan Chains

In Figure 3.12 we plot the average Cl$^-$ cylindrical g(r) and coordination number about the dissociated chains in BmimCl at 450 K. The Cl$^-$ coordination number reaches two at the first minimum in the anion-glucan g(r). Since there are three OH groups per glucan in the cellodextrin chains, this gives an anion:OH ratio of 2:3. On average, each of these Cl$^-$'s is involved in one glucan-anion HB. In Figure 3.13 we plot the average number of OH—Cl$^-$ HBs formed per glucan for an increasing H—Cl$^-$ HB distance cutoff for the dissociated chains at 425 K, 450 K, and 500 K (a constant OH—Cl$^-$ angle cutoff

of 130° was used for all calculations). At all temperatures, the number of HBs plateaus at a value of approximately 2 at a distance cutoff of around 2.5 to 3 Å (This distance is larger than the typical values of ~2 Å for traditional OH—O HBs due to the larger size of Cl^- compared to oxygen). This is exactly the same number of anions in the coordination shell of the glucans, thus each glucan-coordinated Cl^- is also hydrogen-bonded to a glucan. Fluctuations away from this value occur when an OH does not have a Cl^- to HB with, or when a single Cl^- is shared between two neighboring OH groups. Earlier MD work[39] and NMR experiments[36,37] in imidazolium-based ILs showed a greater number of HBs to each glucan per OH group. However, these were for either glucose or cellobiose. Here, the presence of an extended chain changes the ability of the glucan and solvent to form HBs.

3.6.3 Bmim$^+$ Ring Stacking

Glucan ring-Bmim$^+$ ring stacking is not a prominent packing motif for both the dissociated and microfibril states of cellulose. The order parameter we used to quantify ring-ring stacking was the absolute value of the dot product of two unit vectors, one defining the average plane of a glucan ring or microfibril surface and the other the average plane of a Bmim$^+$ ring. This parameter ranges from zero to one, zero meaning the two rings are orthogonal, and one, the two are stacked. For the dissociated chains, this parameter was calculated locally about each glucan after the local solvent volume was divided up into a three-dimensional grid, and then averaged, much like in the calculation of the three-dimensional density in the main text. For the microfibril, the glucan vector was taken to be the vector perpendicular to the microfibril surface closest to the selected cation ring. A second calculation was also done for both cases, wherein the order parameter value for each grid point was set to zero if the Bmim$^+$ density for that point was less than the bulk value, and was left untouched otherwise. We used this as a filter to test if ring-ring stacking coincided with density localization. In Figures 3.14A-B for the dissociated chains, and Figure 3.15A for the microfibril we show that per grid point, there is a high amount of ring-ring ordering for both states of cellulose. However, after applying our filter (Figure 3.14C-D, 15B), we see that the amount of stacking is actually low, as many of the regions of high order have been removed. For the dissociated chains, the volume of prominent stacking shrinks considerably, and the order parameters values of those points that remain are low, around 0.7. For the microfibril, the only surface that still exhibits stacking is the top face. Ring-ring stacking is not prominent in regions of high Bmim$^+$ density. If stacking were energetically favorable, it would also cause density localization. However, after filtering, only small volumes of both cation localization and ring-ring stacking were found centered directly above or below the glucan rings or near the microfibril faces. We conclude that the limited stacking that does occur occurs by excluded volume effects and hindered rotational orientations of the Bmim$^+$ near the glucan rings.

3.6.4 Effects of Terminal Carbon Restraints on the Flexibility of Glucan Chains

The main goal of using restraints on the C1 carbons of the terminal glucan rings of the dissociated chains is to prevent their overall translation and rotation and maintain a dissociated cellulose conformation. Here we show that the effects of the restraint potentials on chain flexibility are small. The densities of backbone torsion angles are almost identical to those observed in the simulations of free cellodextrins in water at the same temperatures.[65] In Figure 3.16 we show the distribution of the two torsion angles around the glycosidic linkages (φ: O5'-C1'-O4-C4, ψ: C1'-O4-C4-C5) at 300 K for the dissociated chains. The glucans effectively sample from only a single region of parameter space, around (-90°, -150°). Studies of free cellodextrins in water using the AMBER/GLYCAM force field show a similar distribution, but with secondary regions of torsion angle space more populated.[65] However, the differences in the relative population of the region around (-90°, -150°) between our studies of those of the AMBER work are small, 99% compared to ~96-99% depending on chain length. Also, differences can be attributed in part to the use of different force fields between the two studies. A comparison of high temperature cellulose conformations generated with either the AMBER/GLYCAM or CHARMM/C35 force field showed an overall agreement between the two force fields but with small differences in relative populations of different conformers,[73] much like we see here.

With regards to chain length and end-to-end distance, while the end restraints do force the chains into an extended conformation, earlier work[32,65,66] has shown that isolated small length cellodextrins also adopt a locally-linear extended geometry in water, and do not form coiled or helical structures. Our own simulations fall under this small length regime, as the 16-glucan long chains lengths, ~8 nm, are less than persistence length estimated for aqueous solutions, 10.6 nm,[65] and those found in aqueous metal complexes, 10-15 nm.[74] Furthermore, although a single atom of each chain end is constrained, the remaining portions of the glucans are free to move, and the chains as a whole still retain a high degree of conformational flexibility. Therefore, placing restraints on the C1 carbons of the terminal rings does not result in a significant conformational deviation from the behaviors of isolated cellodextrin chains.

3.6.5 Creation of Local Coordinate Systems for the Three-Dimensional Density Plots and Cylindrical RDFs

In Figure 3.17A, we show a schematic for the creation of the local Cartesian coordinate system for a given glucan used in finding the average three-dimensional solvent number densities. As mentioned in the main text, the z-axis is the unit vector in the direction of the vector that connects the two oxygens that make up the glycosidic linkages fore and aft of the selected glucan in the chain. A pseudo-x-axis, called the x'-axis, is defined as the unit vector that is the average of the unit vectors along the C2-O2 and C3-O3 bond vectors. The cross product of the z- and x'-axes is then the y-axis, and the cross product of the y-and z-axes then defines the x-axis. An axes triad is shown in the lower right of the figure. The y-axis is coming out of the plane of the image. The center-of-mass of the glucan ring (The C1, C2, C3, C4, C5, and O5 atoms) is defined as the origin of the coordinate system.

In Figure 3.17B we show an example of the cylindrical coordinate system method we used to calculate the cylindrical RDFs. For a given pair of neighboring glucan units in a chain, a vector is created that connects the centers of mass of their two rings (ring centers-of-mass are defined as above for the three-dimensional number density calculations). This vector is taken to be the z-axis of a cylindrical coordinate system. Then, two planes are constructed, which represent the ends of the cylinder. The planes are orthogonal to the cylinder's z-axis and each passes through one of the two glucan rings' centers-of-mass, one for each end. Only solvent molecules that lie between these two planes are considered when constructing the cylindrical RDF; in the figure, of the two hypothetical water positions, only the lower one would be included in the RDF for that cylinder (or glucan-glucan pair). This process is repeated over all consecutive glucan ring pairs to construct the average RDF.

3.6.6 Chain Collapse in the 'Free' 18 Chain System

An additional 'free' simulation of glucan chains in water was carried out. This system contained 18 chains, each with 10 glucose residues, instead of the 36 chain, 16 residue per chain system presented in the main text. The chains were solvated in 10304 TIP3P waters, giving a sugar mass fraction of 0.14. The simulation was performed at 300 K in the same manner as the larger free glucan/water system described in the main text. The only differences in methodology between the two was that for this smaller system, time zero was taken as the time of release of the restraints on the glucan atoms, and that the simulation was run for a total of 100 ns, longer than the 40 ns duration of the larger system.

We show the results of this simulation in Figure 3.18. The total potential energy of the system also shows no overall drift over the course of the simulation, while the number of contacts between chains (same as defined as in the main text) increases steadily during the first 60 ns and then varies about a plateau value of ~1100 for the remainder of the simulation. As in the case of the larger system, glucan chain collapse occurred spontaneously without an accompanying change in total potential energy, and therefore the collapse is entropically driven.

3.6.7 Convergence of Solvent Structures

In Figure 3.19 we show the cylindrical g(r)'s for water at 300 K, and Cl$^-$ anion and total Bmim$^+$ cation at 450 K, around the dissociated glucan chains, calculated using either the entire simulation trajectory, its first half, or its second half. In water, the solvent structure profiles are indistinguishable. In IL, the two density profiles are also effectively identical. Overall, the two figures show that the there is no appreciable difference in solvent structures of the two systems between the two halves of their simulation trajectories, and that the overall solvent structures have indeed converged over the length of our simulations.

3.6.8 Three-Dimensional Solvent Structures

We show the coarse-grained distribution of the alkyl tail group of $Bmim^+$ in Figure 3.20. This 'tail' site is defined as the center of mass of the last three methyl groups of the butyl chain of $Bmim^+$ (or alternatively, the terminal propyl group of that butyl chain). This site incorporates all $Bmim^+$ atoms not included in the $Bmim^+$ ring site described in the main text. In the axial direction, locations of high tail number density are present by the glycosidic linkages (Figure 3.20B), like in the case of the $Bmim^+$ rings. In the equatorial direction however, there is a comparative lack of tail number density. This is different than the case of the $Bmim^+$ rings, where there were still high number density regions in the equatorial direction, albeit at positions further away from the glucans than in the axial direction. So, while both $Bmim^+$ groups favor axial localization over equatorial localization, the way in which this preference exists differs.

We show the combined densities of all three BmimCl groups (Cl^-, $Bmim^+$ ring, $Bmim^+$ tail) in Figure 3.21. In the axial direction, all three sites have zones of high number density approximately equal in distance to the glucan, while in the equatorial direction, the high density zones are staggered, with the Cl^- regions closer to the sugar than the $Bmim^+$ ring regions. This staggering of anion and cation localizations may explain in part the lack of tail density in this direction, as this space is needed for the excess-charge carrying portions of the $Bmim^+$ cations to maintain local neutrality.

Equation 3.1

$$U(\mathbf{R}) = \sum_{bonds} K_b (r - r_o)^2 + \sum_{angles} K_\theta (\theta - \theta_o)^2 + \sum_{dihedrals} K_\chi [1 + \cos(n\chi - \delta)] +$$

$$\sum_{UB} K_{UB} (S - S_o)^2 + \sum_{impropers} K_{imp} (\varphi - \varphi_o)^2 + \sum_{nonbonded} \left(\varepsilon_{ij} \left[\left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{12} - \left(\frac{R_{min_{ij}}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{\varepsilon_l r_{ij}} \right)$$

Table 3.1. Comparison of the water/Bmim$^+$ minimum energy complexes' interaction energies and geometries between the HF/6-31G* quantum calculations and the CHARMM force field.

Site	Interaction Energy (kcal/mol)		Hydrogen Bond Distance (Å)	
	HF/6-31G*	CHARMM	HF/6-31G*	CHARMM
1	-8.77	-8.81	2.72/2.54[a]	2.58/2.38[a]
2	-8.28	-8.68	2.54	2.22
3	-11.01	-10.94	2.30	1.99
4	-11.68	-11.40	2.10	2.02
5	-9.12	-9.52	2.38	2.26

[a]Water at Site 1 has two water-Bmim$^+$ HBs in its minimum energy geometry, to hydrogens HE1 and HE2 (See Figure 3.9 for Bmim$^+$ labeling).

Table 3.2A. CHARMM topology file for Bmim$^+$.

MASS   100 CC32A   12.01100 C
MASS   101 CC33A   12.01100 C
MASS   102 CPH1   12.01100 C
MASS   103 CPH2   12.01100 C
MASS   104 HP   1.00800 H
MASS   105 NR3   14.00700 N
MASS   107 HG   1.00800 H
MASS   108 HAC2   1.00800 H
MASS   109 HAC3   1.00800 H

RESI BMIM   1.00
GROUP
ATOM CG   CPH2   0.231
ATOM NE2 NR3   -0.328
ATOM HE2 HP   0.167
ATOM ND1 NR3   -0.328
ATOM HT   HG   0.176
ATOM CE1 CPH1   0.051
ATOM HE1 HP   0.159
ATOM CE2 CPH1   0.046
GROUP
ATOM C1   CC32A -0.154
ATOM H1   HAC2   0.211
ATOM H2   HAC2   0.211
GROUP
ATOM C2   CC32A -0.150
ATOM H3   HAC2   0.110
ATOM H4   HAC2   0.110
GROUP
ATOM C3   CC32A -0.160
ATOM H5   HAC2   0.095
ATOM H6   HAC2   0.095
GROUP
ATOM C4   CC33A -0.260
ATOM H7   HAC3   0.090
ATOM H8   HAC3   0.090
ATOM H9   HAC3   0.090
GROUP
ATOM C5   CC33A -0.254
ATOM H10 HAC3   0.234
ATOM H11 HAC3   0.234
ATOM H12 HAC3   0.234

BOND CG HT   CG ND1 CG NE2 ND1 C1

BOND ND1 CE1 CE1 HE1 CE1 CE2 CE2 HE2
BOND CE2 NE2 NE2 C5   C1 H1   C1 H2
BOND C1 C2   C2 H3   C2 H4   C2 C3
BOND C3 H5   C3 H6   C3 C4   C4 H7
BOND C4 H8   C4 H9   C5 H10 C5 H11
BOND C5 H12

Table 3.2B. CHARMM parameter file for Bmim$^+$.

Bonds
CC32A   HCA2   309.00   1.100
CC33A   HCA3   322.00   1.090
CC32A   CC32A   222.50   1.530
CC32A   CC33A   222.50   1.520
NR3 CPH2 380.000   1.3370
CPH1 CPH1 410.000   1.3700
NR3 CPH1 380.000   1.3700
HP   CPH1 375.000   1.0800
HG   CPH2 333.000   1.0800
NR3 CC32A 250.000   1.4830
NR3 CC33A 250.000   1.4700

Angles
HCA2   CC32A   CC32A   32.500   110.10 22.53 2.179
HCA2   CC32A   CC33A   34.600   110.10 22.53 2.179
HCA2   CC32A   HCA2   35.50   107.00   5.40 1.802
HCA3   CC33A   HCA3   35.50   108.40   5.40 1.802
HCA3   CC33A   CC32A   34.600   110.10 22.53 2.179
CC32A   CC32A   CC32A   58.350   111.60 11.16 2.561
CC32A   CC32A   CC33A   58.000   112.00   8.00 2.561
CPH2 NR3   CPH1 145.000   108.0000
NR3 CPH1 CPH1 145.000 108.0000
NR3 CPH1 HP 22.000   122.30 15.00 2.14000
NR3 CPH2 HG 32.000   125.7 25.00 2.14000
HP CPH1 CPH1 22.000   130.00 15.00 2.21500
CC33A   NR3 CPH1 45.000   125.00 ! 15.00 2.13000
CC33A   NR3 CPH2 45.000   127.00 ! 15.00 2.13000
CC32A   NR3 CPH1 45.000   125.00 ! 15.00 2.13000
CC32A   NR3 CPH2 45.000   127.00 ! 15.00 2.13000
HCA3 CC33A NR3 50.000   109.5000
HCA2 CC32A NR3 50.000   105.5000
NR3 CC32A CC32A 140.000   112.6000
NR3 CPH2 NR3   145.000   109.1000

Dihedrals
HAC2   CC32A   CC32A   HAC2   0.19000   3   0.0
CC32A   CC32A   CC32A   HAC2   0.19000   3   0.00
CC33A   CC32A   CC32A   HAC2   0.19000   3   0.00
HAC2   CC32A   CC33A   HAC3   0.17000   3   0.0
CC32A   CC32A   CC33A   HAC3   0.17000   3   0.0
CC33A   CC32A   CC32A   CC32A   0.1500   1   0.0
CPH2 NR3 CC32A CC32A   0.000 4   0.00
CC32A CC32A NR3 CPH1   0.450 2   0.00

CC32A CC32A NR3 CPH1   0.120 1   180.00
HAC2   CC32A NR3 CPH2   0.0000 3   00.00
CPH1 NR3 CC32A HAC2   0.0650 3   0.00
CC32A NR3 CPH1 CPH1 5.000 2 180.00
HP CPH1 NR3 CC32A 1.0000 2   180.00
HG CPH2 NR3 CC32A 1.0000 2   180.00
NR3 CPH2 NR3 CC32A 1.000 2   180.00
HAC3   CC33A NR3 CPH2   0.0000 3   000.00
CPH1 NR3 CC33A HAC3 0.0650 3   0.00
CC33A NR3 CPH1 CPH1 3.500 2   180.00
HP CPH1 NR3 CC33A 1.0000 2   180.00
HG CPH2 NR3 CC33A 1.0000 2   180.00
NR3 CPH2 NR3 CC33A 1.000 2   180.00
CPH2 NR3 CPH1 CPH1   12.0000 2   180.00
HP CPH1 NR3 CPH2   2.500 2   180.00
HG CPH2 NR3 CPH1   3.0000 2 180.00
NR3 CPH1 CPH1 NR3   12.0000 2 180.00
NR3 CPH2 NR3 CPH1   12.0000 2   180.00
NR3 CPH1 CPH1 HP   2.500 2   180.00
HP CPH1 CPH1 HP   1.0000 2   180.00
CC32A CC32A CC32A NR3 0.3 3   0.00
CC32A CC32A CC32A NR3 0.70 1   180.00
HAC2 CC32A CC32A NR3 0.19 3   0.00

Improper
HG NR3 NR3 CPH2 1.0 0 0.00
NR3 CPH1 CPH2 CC33A 2.00 0 0.00
NR3 CPH1 CPH2 CC32A 2.00 0 0.00
CPH1 CPH1 NR3 HP 1.00 0 0.00
CPH1 NR3 CPH1 HP 1.00 0 0.00

Nonbonded
HCA2   0.0 -0.0220   1.320
HCA3   0.0 -0.0220   1.320
HP   0.000000 -0.046000   0.900000
HG   0.000000 -0.046000   0.700000
CPH1   0.000000 -0.050000   1.800000
CPH2   0.000000 -0.050000   1.800000
NR3   0.000000 -0.200000   1.850000
CC32A   0.0   -0.0550   2.17500 0.0 -0.01 1.9
CC33A   0.0   -0.0800   2.06000 0.0 -0.01 1.9
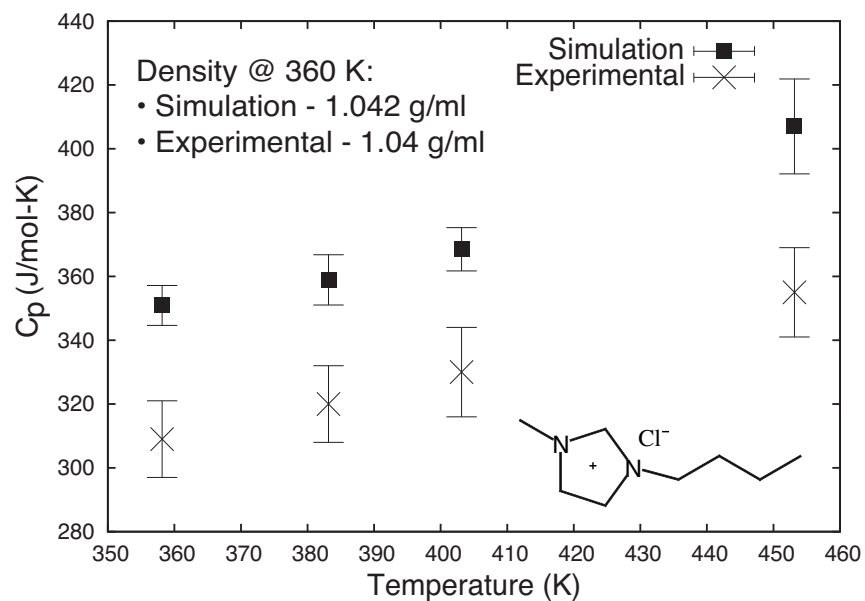
Figure 3.1. Comparison of the calculated heat capacity and density of BmimCl (shown at bottom right) from the newly developed atomistic force field with the experimental data.[48,49]

(A)
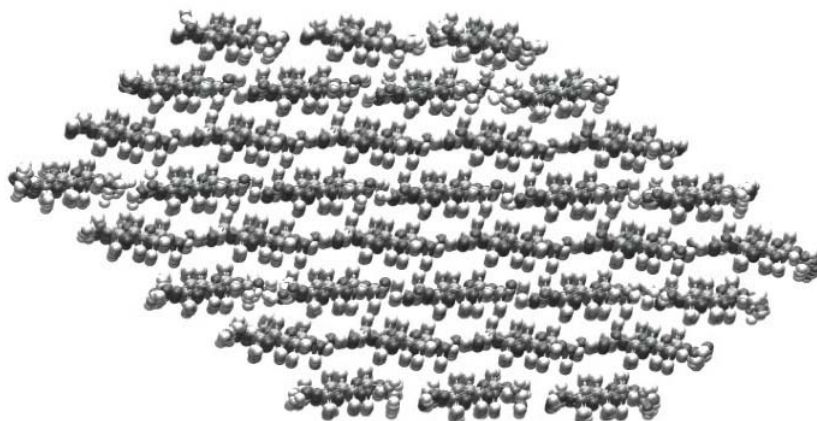
(B)

Figure 3.2. Cross-sections of the two atomistic models of cellulose simulated in this work, the cellulose microfibril (A) and dissociated glucan chains (B). Both contain 36 glucan chains, each 16 residues long.
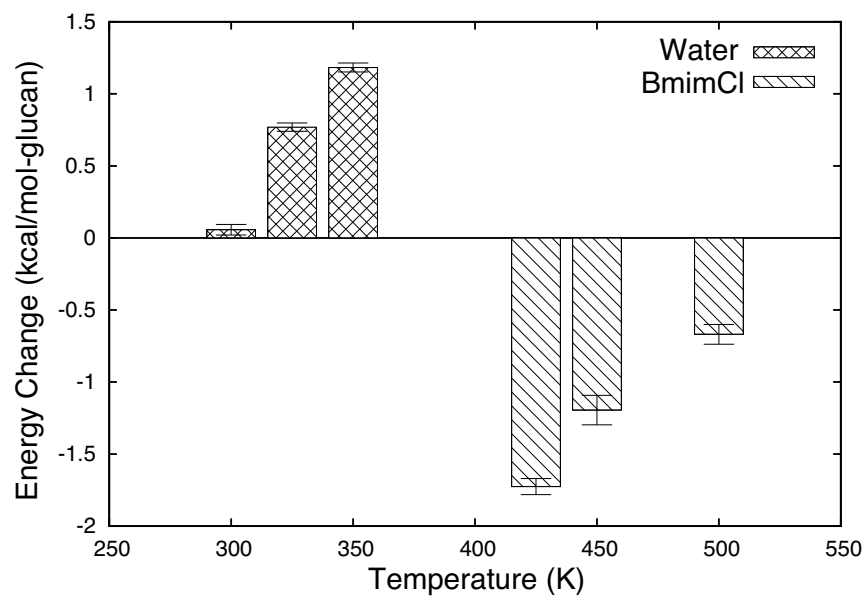
Figure 3.3. Difference in energy between the dissociated and microfibril states of cellulose (Dissociated-Microfibril) in both water and BmimCl, normalized per glucan unit, from the all-atom MD simulations.

(A)
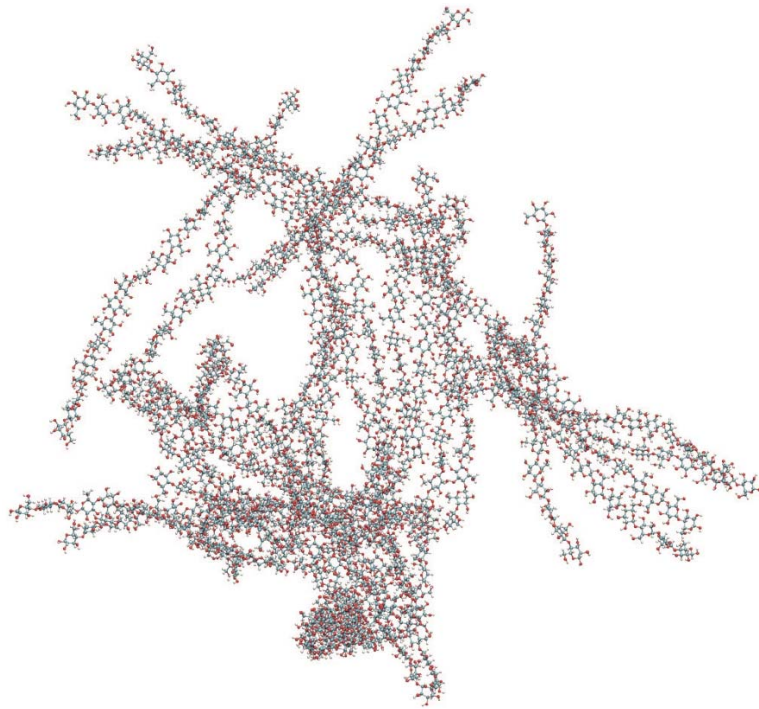
(B)
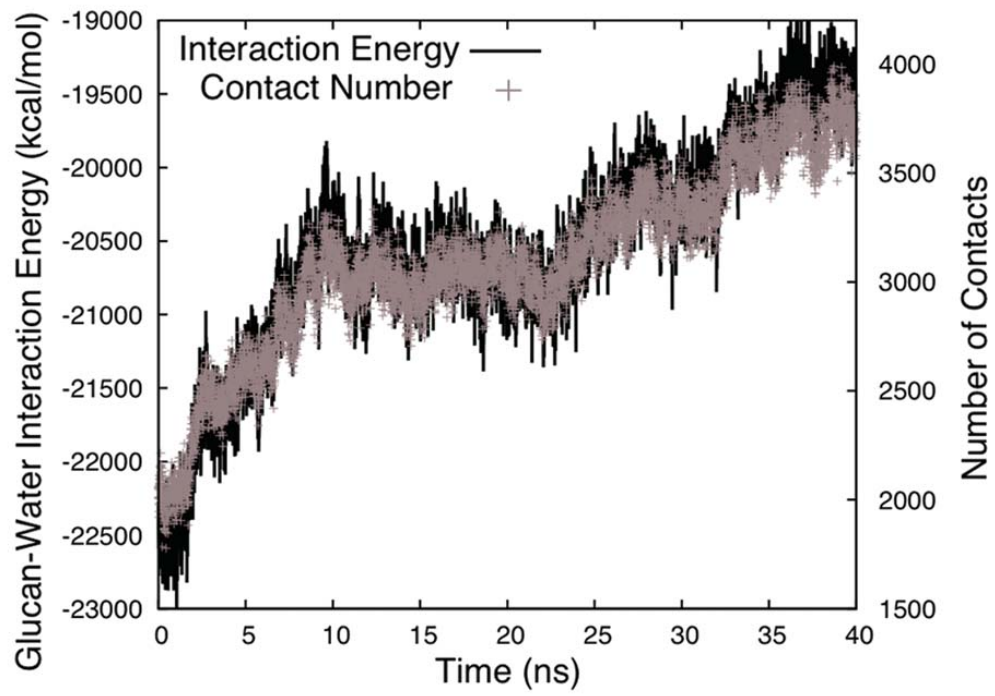
(C)
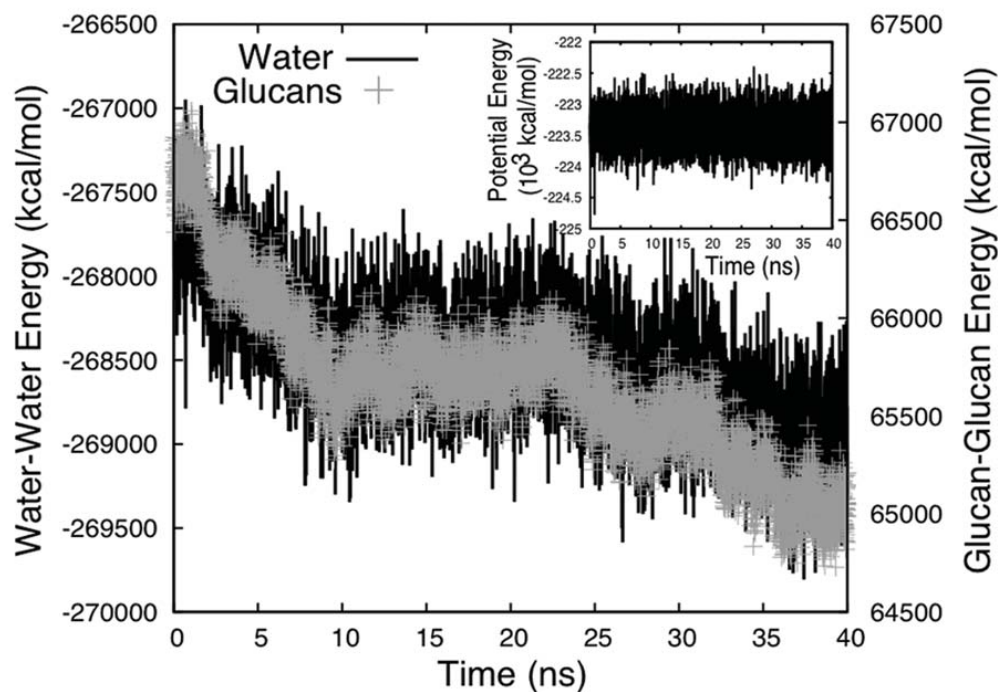
Figure 3.4. Results of the collapse of glucan chains in water at 300 K during the 'free' simulation. **(A)** Structure of the collapsed aggregate of glucan chains at the end of the simulation. (B) The temporal evolution of the total glucan-water interaction energy and the number of glucan-glucan contacts within 3 Å during the 'free' simulation in water. Only the contacts between glucan atoms of different chains are included. (C) The temporal evolution of the total water-water and glucan-glucan interaction energies during the 'free' simulation. The inset shows the temporal evolution of the total potential energy of the entire system.

(A)
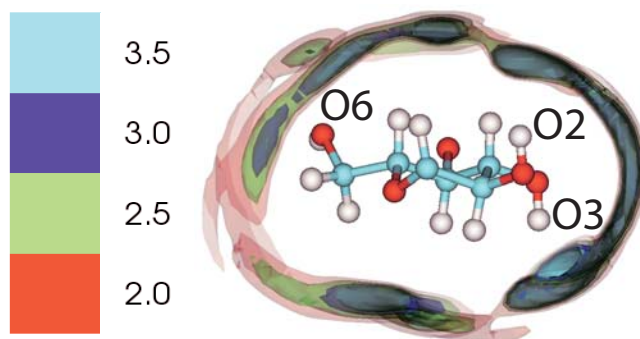
(B)
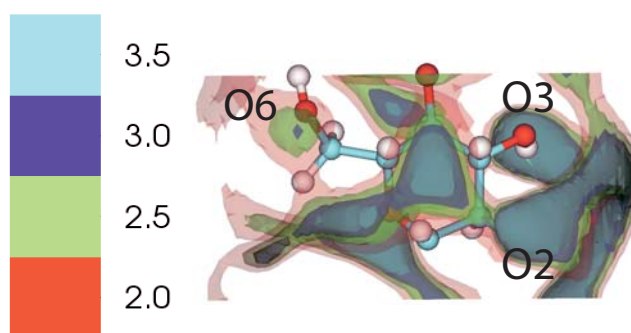
(C)
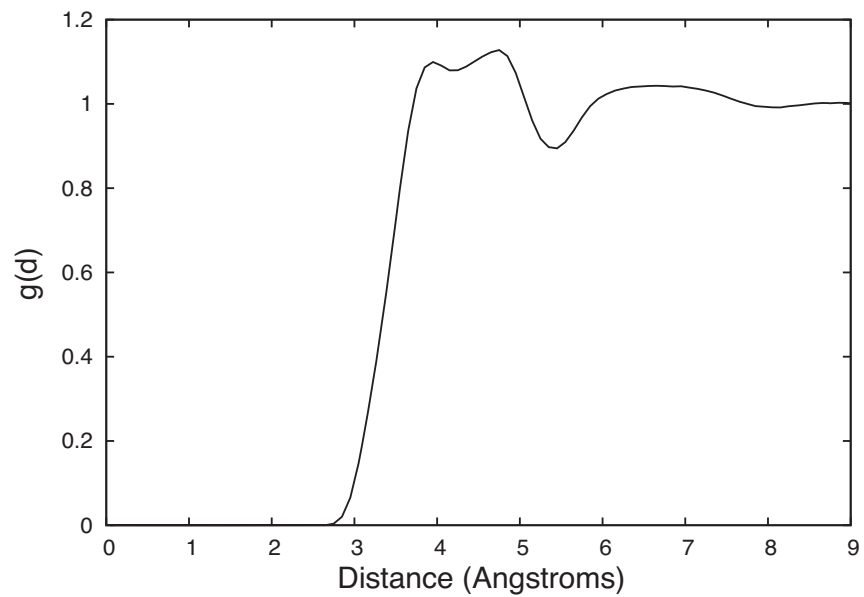
Figure 3.5. The calculated profile of the temporally and spatially averaged number density of water around a glucan chain normalized by the bulk value viewed from along the chain axis (A) and above the plane of the glucan ring (B) at 300 K and 1 atm. (C) The cylindrical radial distribution function of water, by center of mass, around the dissociated chains.

(A)
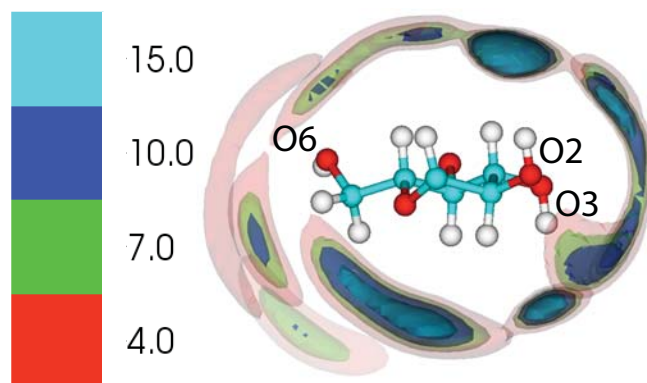
(B)
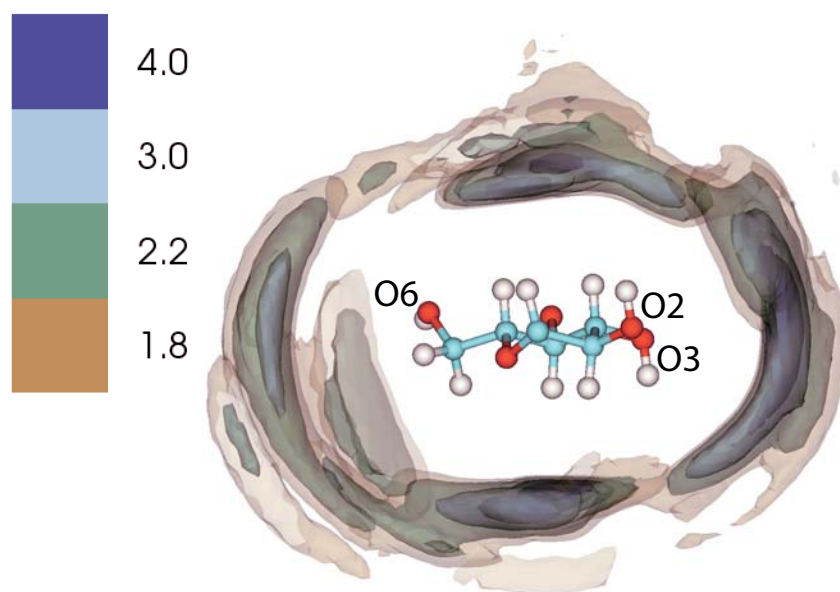
(C)
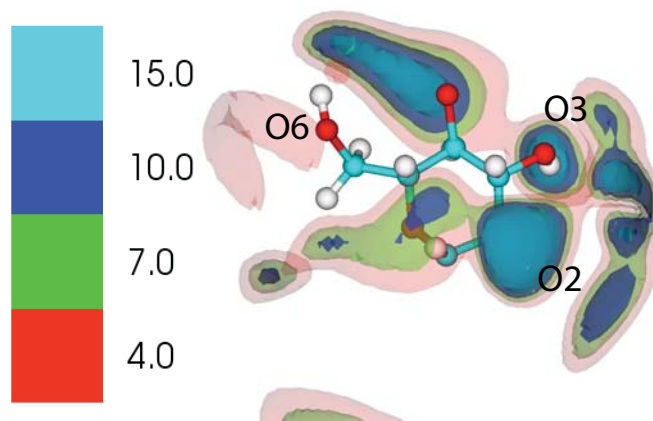
(D)
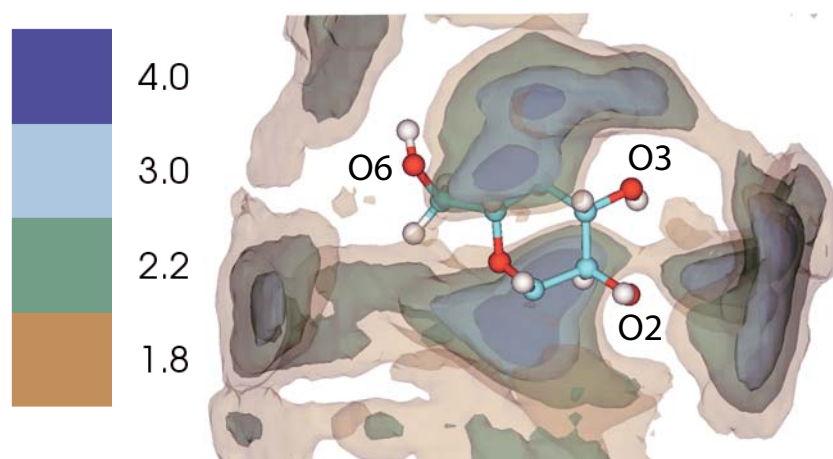
Figure 3.6. The calculated profiles of the temporally and spatially averaged number densities of the centers of masses of anions and cation rings around a glucan chain normalized by bulk values at 450 K and 1 atm. The density profiles viewed from along the chain axis: Cl⁻ (A) and Bmim⁺ (B). The density profiles viewed from the top of the glucan ring: Cl⁻ (C) and Bmim⁺ (D).
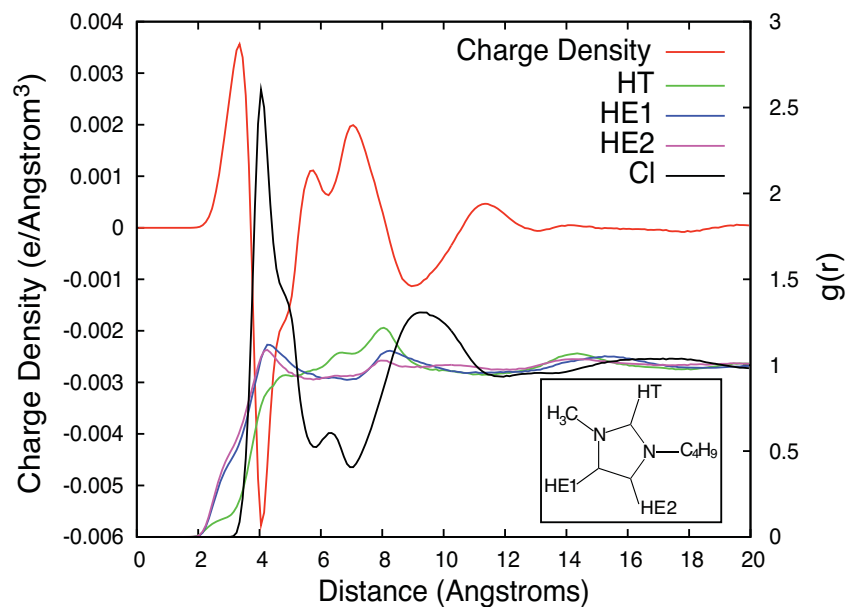
Figure 3.7. The calculated cylindrical radial distribution profile of charge density of BmimCl, based on the partial charges of the IL atoms, and the cylindrical radial density profiles of Cl⁻ and three electrophilic protons of the imidazolium ring, around the dissociated glucan chains, at 450 K and 1 atm. The inset shows the labeling of the three hydrogens in the imidazolium ring.
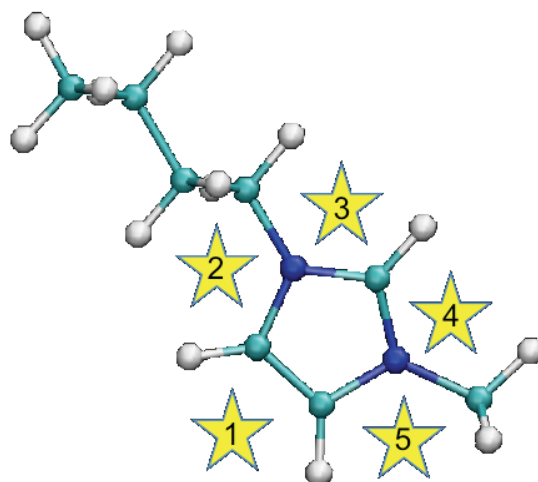
Figure 3.8. Sites of water for creating the five water/Bmim$^+$ complexes used in creating and validating the CHARMM force field.
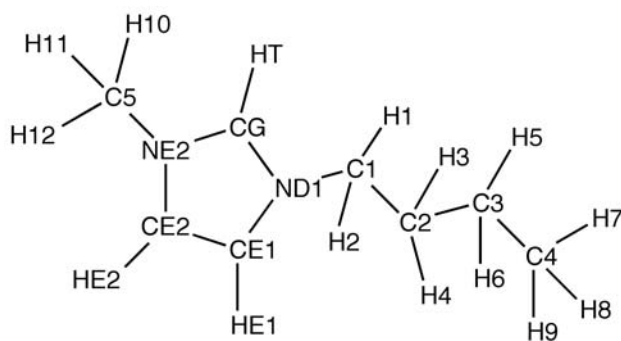
Figure 3.9. Labeling of the Bmim$^+$ atoms in the CHARMM force field.
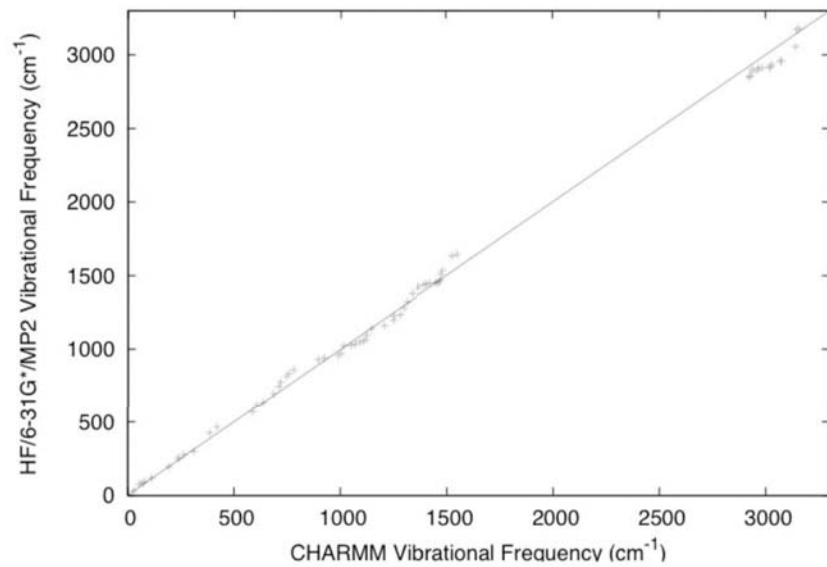
Figure 3.10. Comparison of the vibrational spectrum of Bmim$^+$ from *ab initio* HF/6-31G*/MP2 quantum calculations and from the CHARMM force field.

(A)
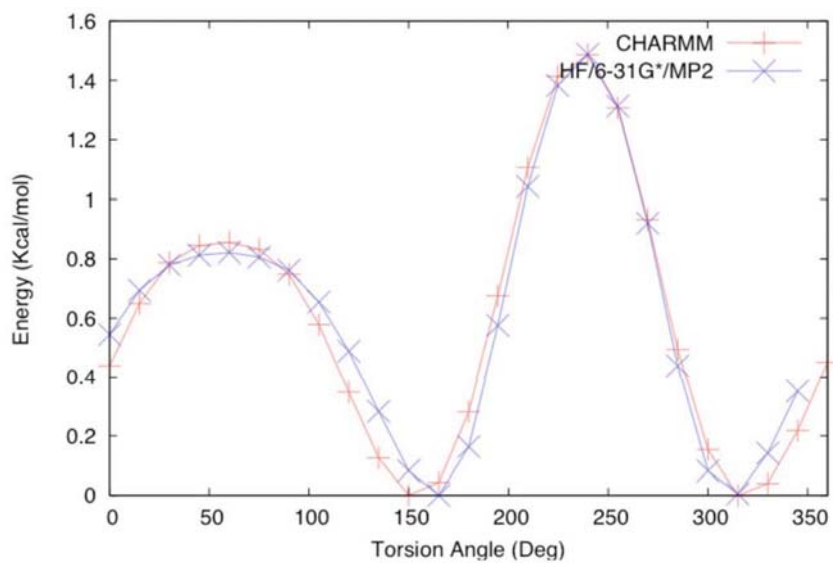
(B)
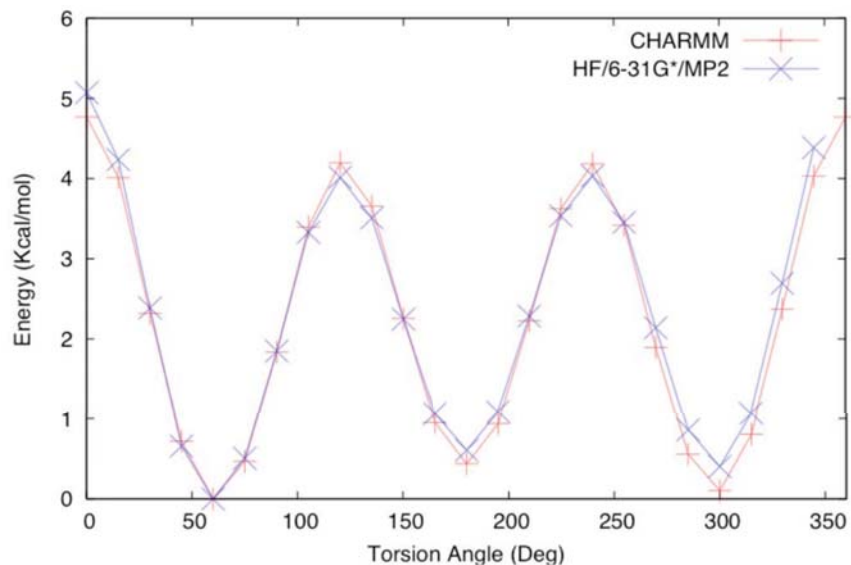
Figure 3.11. Torsion energy profiles of the CE1-ND1-C1-H1 dihedral (A) and ND1-C1-C2-C3 dihedral (B) for both the CHARMM force field and HF/6-31G*/MP2 quantum calculations (See Figure 3.9 for Bmim[+] labeling).
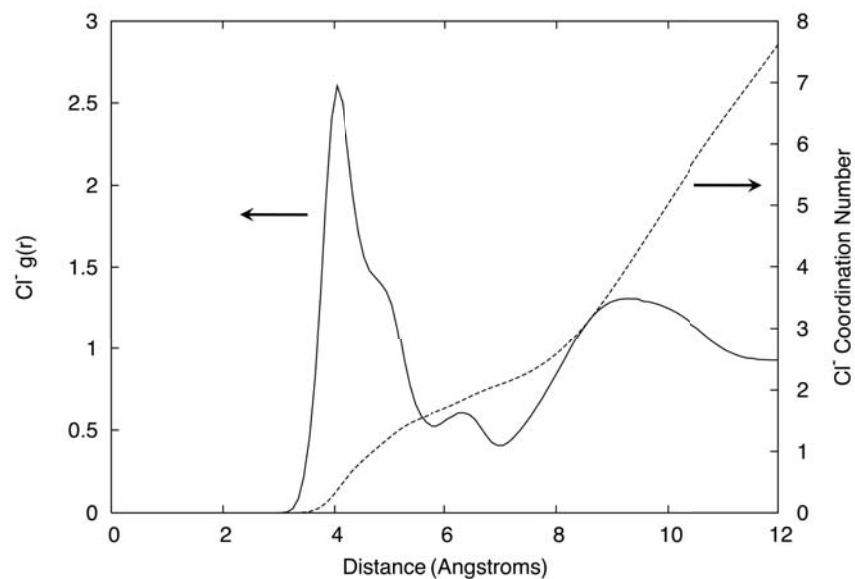
Figure 3.12. Cl⁻ cylindrical g(r) and coordination number per glucan as a function of distance for the dissociated glucan chains at 450 K.
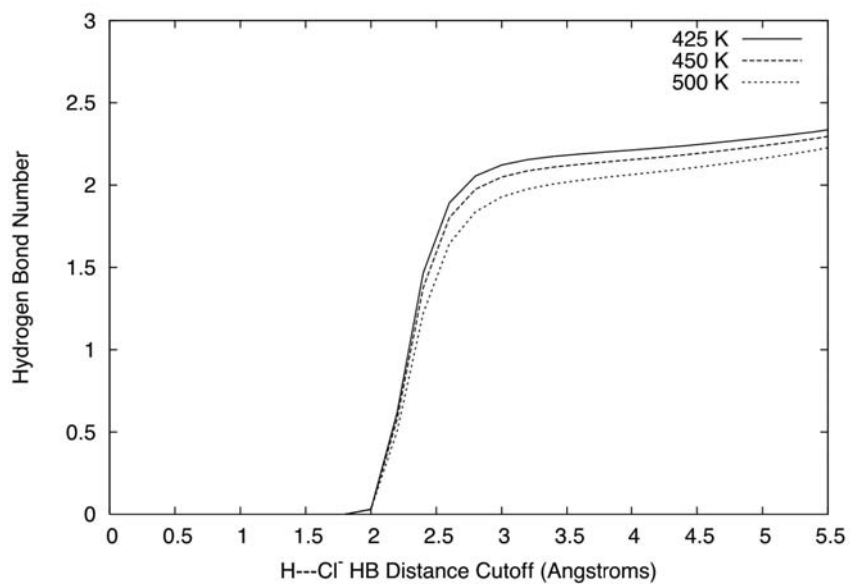
Figure 3.13. Average number of OH—Cl⁻ HBs per glucan for the dissociated chains as a function of the H—Cl⁻ HB distance cutoff at 425 K, 450 K, and 500 K. A constant OH—Cl⁻ angle cutoff of 130° was employed.

(A)
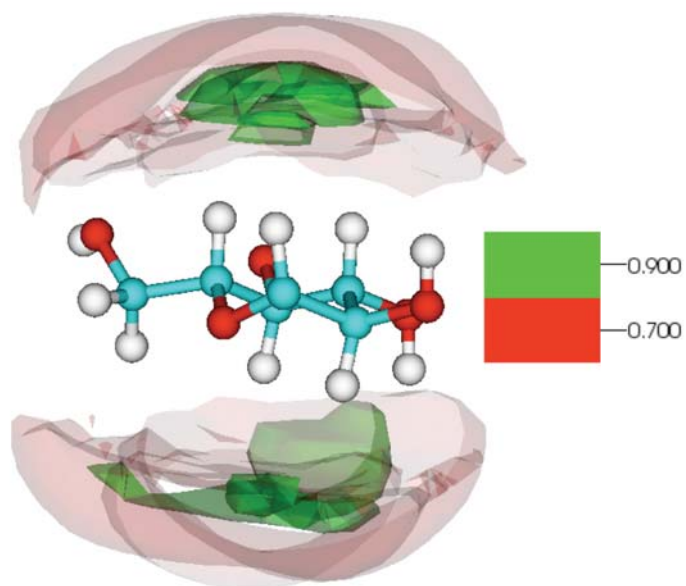
(B)
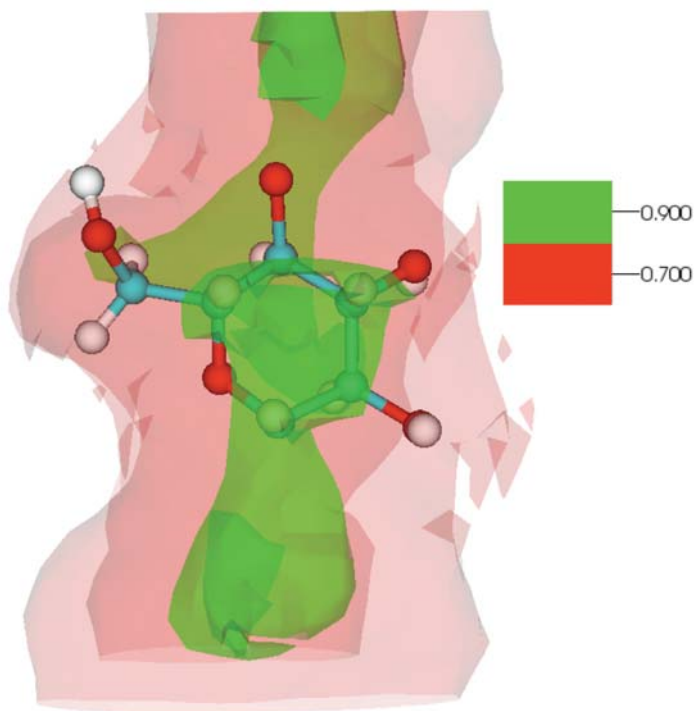
(C)
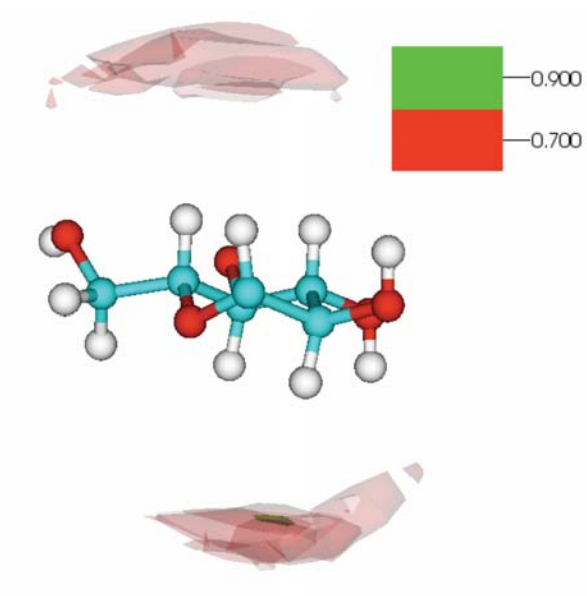
(D)
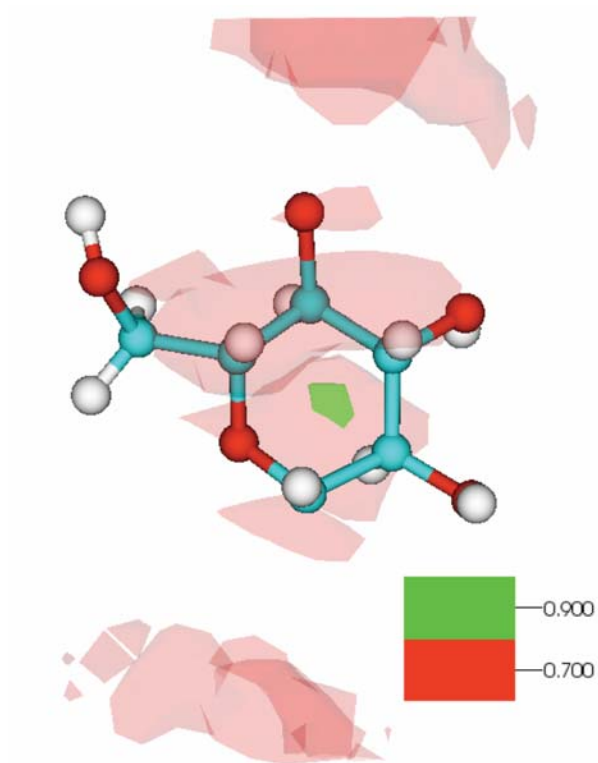
Figure 3.14. Average glucan-Bmim$^+$ ring-ring stacking order parameter about the glucans for the dissociated chains seen from along the chain axis (A) and from above the glucan ring plane (B) at 450 K. (C) and (D) show the same after the values have been set to zero for all regions which have a Bmim$^+$ density less than the bulk.

(A)
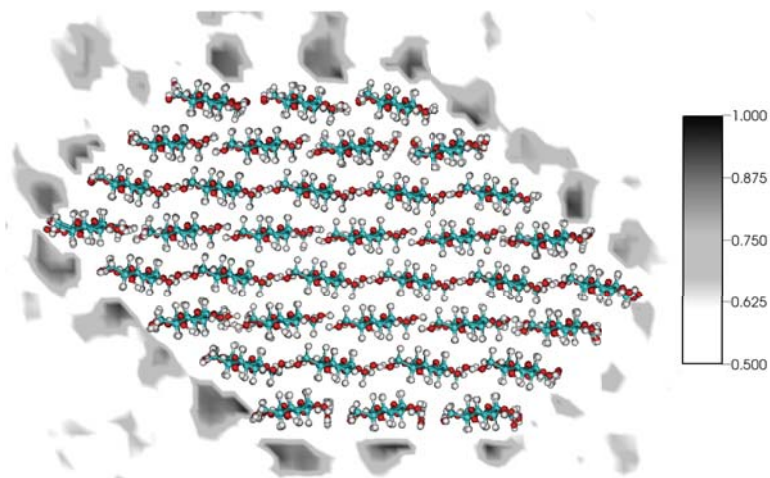
(B)
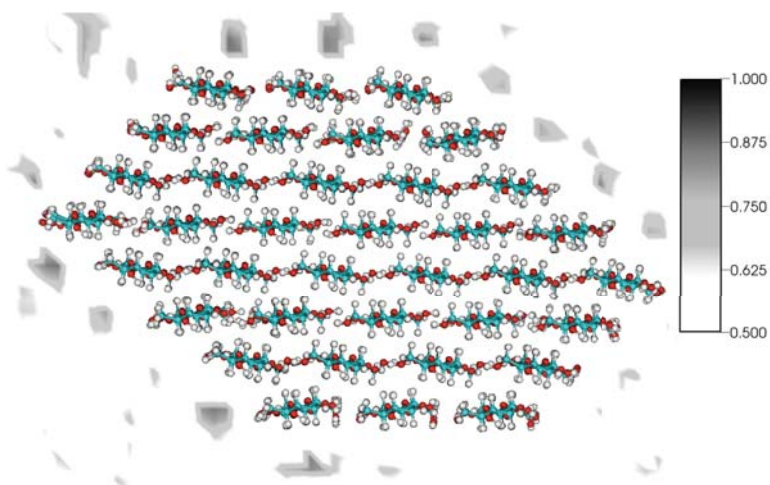
Figure 3.15. Glucan-Bmim$^+$ ring-ring order parameter averaged down the length of the fibril about a cellulose I$_\beta$ microfibril at 450 K (A). (B) shows the same after the values have been set to zero for all regions which have a Bmim$^+$ density less than the bulk.
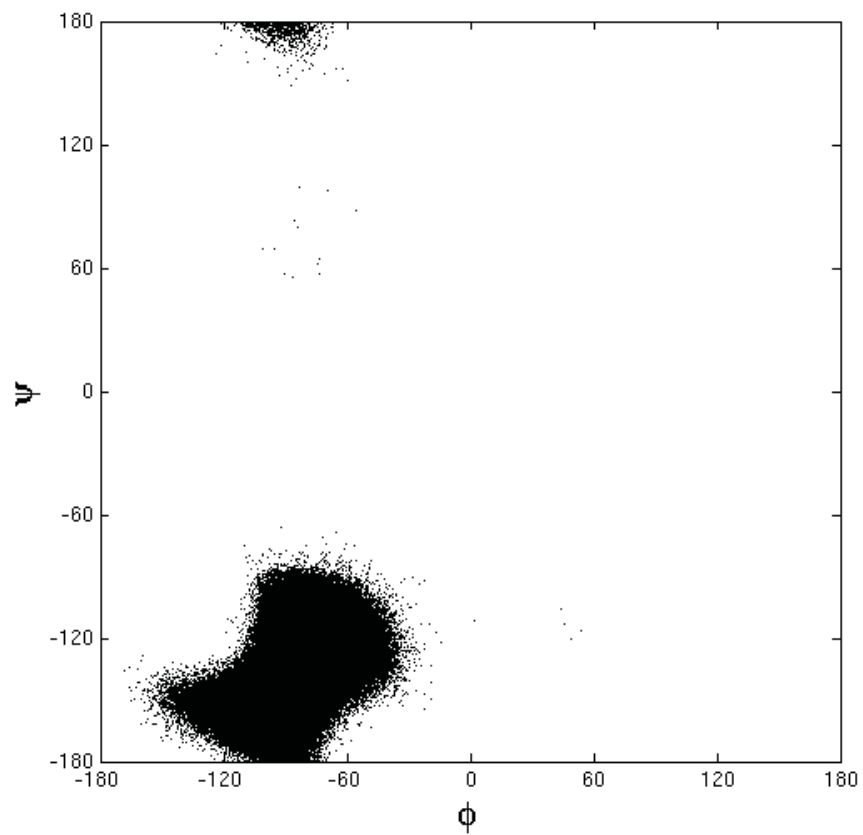
Figure 3.16. Scatterplot of the φ-ψ dihedral angle distribution of the glucan glycosidic linkages at 300 K and 1 atm in water for the dissociated chains.

(A)

$$\vec{v}_{O4-O4'} = \vec{r}_{O4'} - \vec{r}_{O4}$$

$$\vec{v}_{C2-O2} = \vec{r}_{O2} - \vec{r}_{C2}$$

$$\vec{v}_{C3-O3} = \vec{r}_{O3} - \vec{r}_{C3}$$

$$\vec{x}' = \frac{\dfrac{\vec{v}_{O2-C2}}{\left|\vec{v}_{O2-C2}\right|} + \dfrac{\vec{v}_{O3-C3}}{\left|\vec{v}_{O3-C3}\right|}}{\left\|\dfrac{\vec{v}_{O2-C2}}{\left|\vec{v}_{O2-C2}\right|} + \dfrac{\vec{v}_{O3-C3}}{\left|\vec{v}_{O3-C3}\right|}\right\|} \qquad \begin{aligned} \vec{z} &= \frac{\vec{v}_{O4-O4'}}{\left|\vec{v}_{O4-O4'}\right|} \\ \vec{y} &= \vec{z} \times \vec{x}' \\ \vec{x} &= \vec{y} \times \vec{z} \end{aligned}$$

(B)

☆ - Glucan Ring Center-of-Mass

☆ - Water Center-of-Mass

$$\vec{z} = \vec{v}_{COM}$$
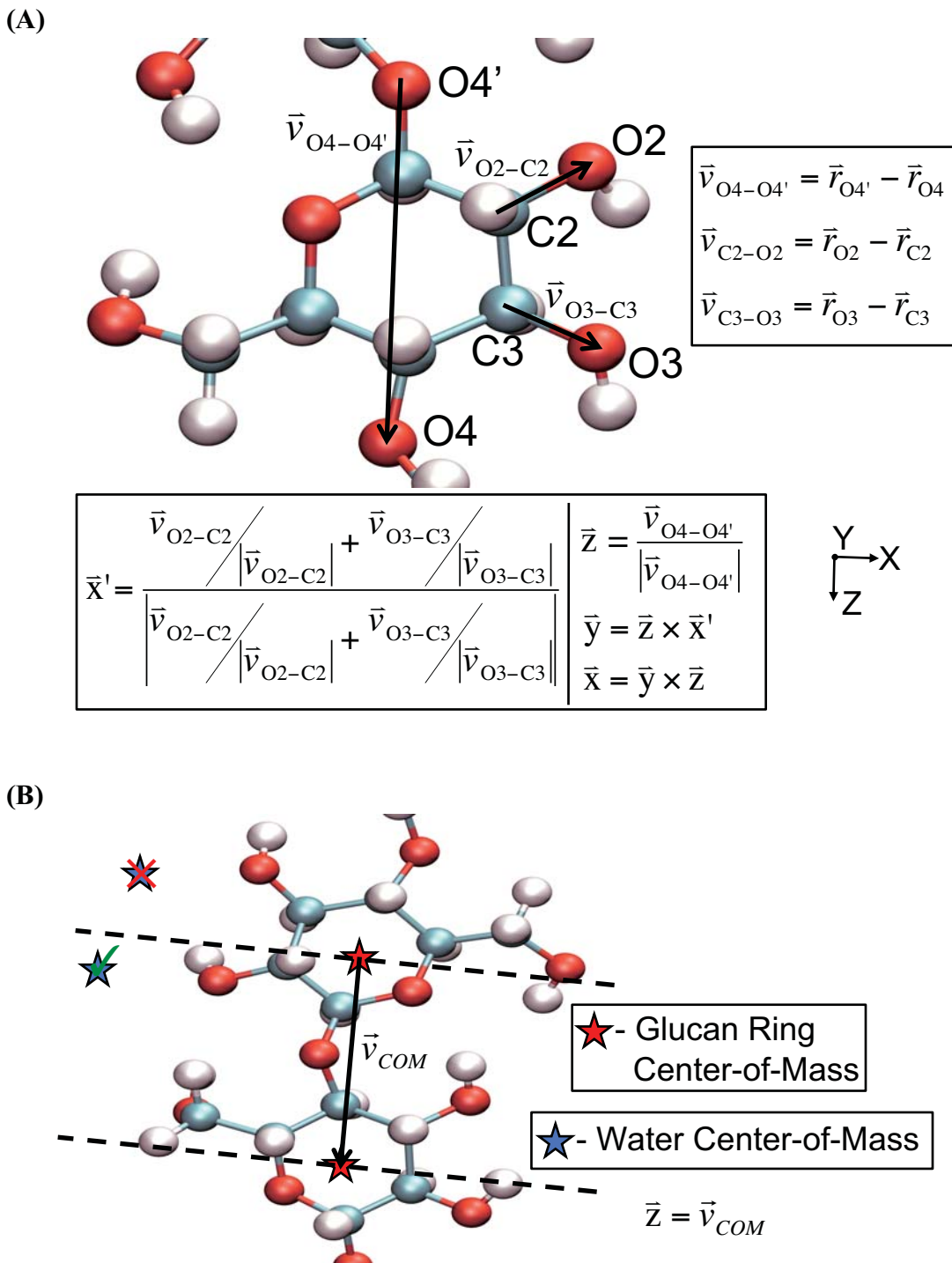
Figure 3.17. Schematic for the construction of the Cartesian coordinate system used for calculating the local number density distribution of solvent around glucans in the dissociated chains (A). The formulas for the three inter-atomic vectors used in calculating the unit axes are shown on the right, while the formulas for the x-, y-, and z-axes, as well as the x'-axis, are shown on the bottom of the figure. Diagram of the cylindrical coordinate system used for calculation of the cylindrical glucan-solvent RDFs about the

dissociated chains (B). Two hypothetical water positions are shown. Only waters that lie between the two planes that define the ends of the cylinder (shown as the two dotted lines) are considered. The z-axis of the cylinder is the vector connecting the centers-of-mass of two consecutive glucan rings.
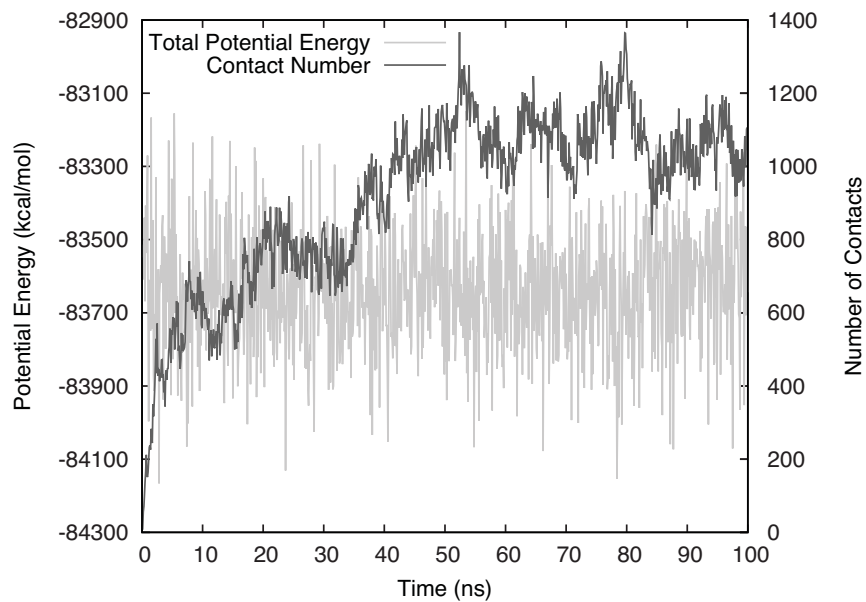
Figure 3.18. Behavior of the 18 chain system during collapse. Time zero is the moment the constraints on the sugar atoms were removed. Number of contacts is defined as the number of cellulose atom pairs where the atoms are within 3 Å of each other and on different chains.

(A)
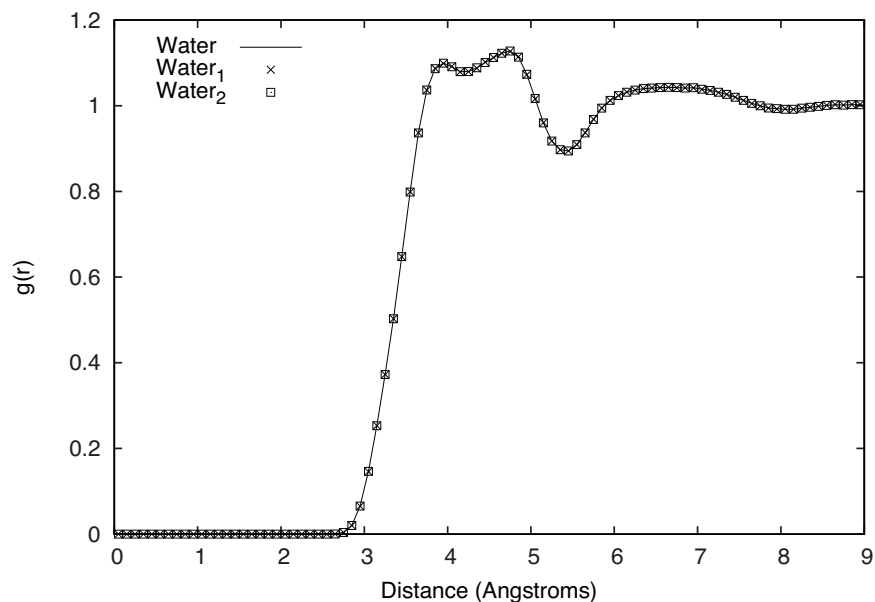
(B)
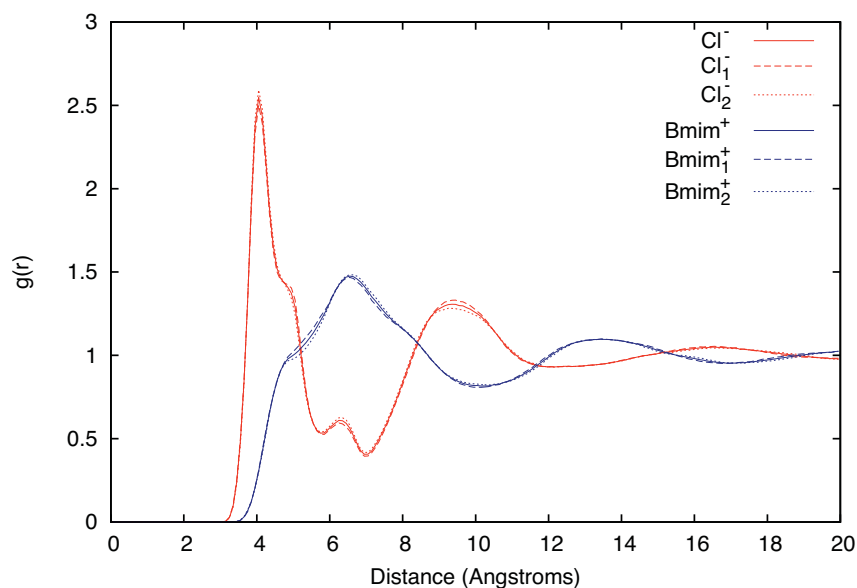
Figure 3.19. Convergence of water structure about the dissociated glucan chains at 300 K (A). 'Water' refers to the g(r) calculated from the entire 15 ns simulation, while 'Water$_1$' and 'Water$_2$' refer to a g(r) constructed from data of only the first 7.5 ns or second 7.5 ns of the simulation, respectively. Convergence of IL structure about the dissociated glucan chains at 450 K (B). 'Cl' curves show the g(r) of Cl$^-$ about the chains, while 'Bmim' curves the same for the whole Bmim$^+$ cation. The subscripts 1 and 2 refer to curves constructed from either the first 15 ns or last 15 ns of the simulation. A lack of subscript indicates a curve calculated with the entire 30 ns trajectory.

(A)
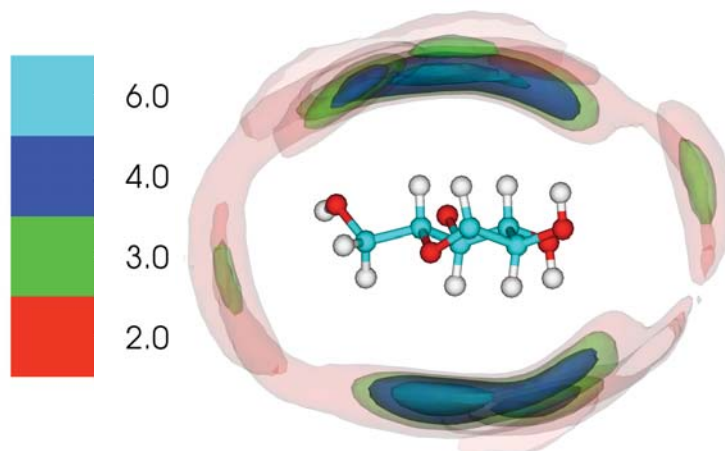
(B)
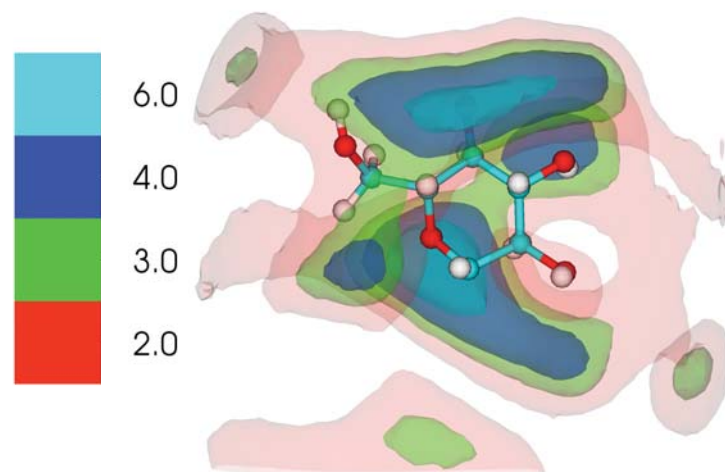
Figure 3.20. The calculated profile of the temporally and spatially averaged number densities of the centers of masses of the Bmim[+] tail sites around a glucan chain normalized by bulk values at 450 K and 1 atm, viewed from along the chain axis (A) and from the top of the glucan ring (B).

(A)
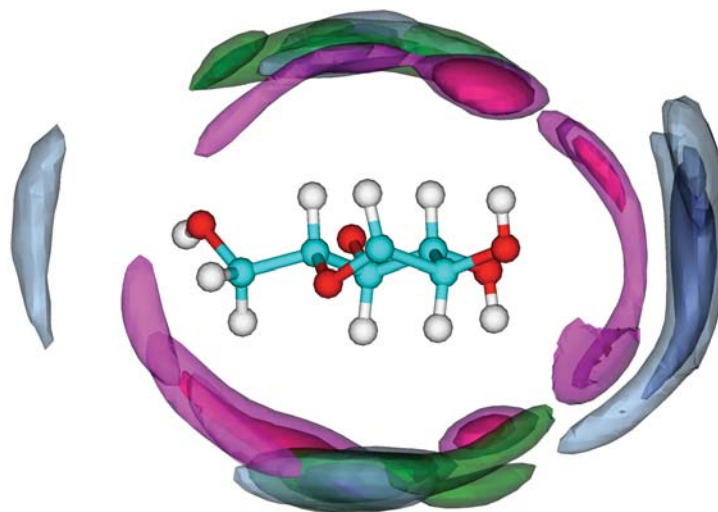
(B)
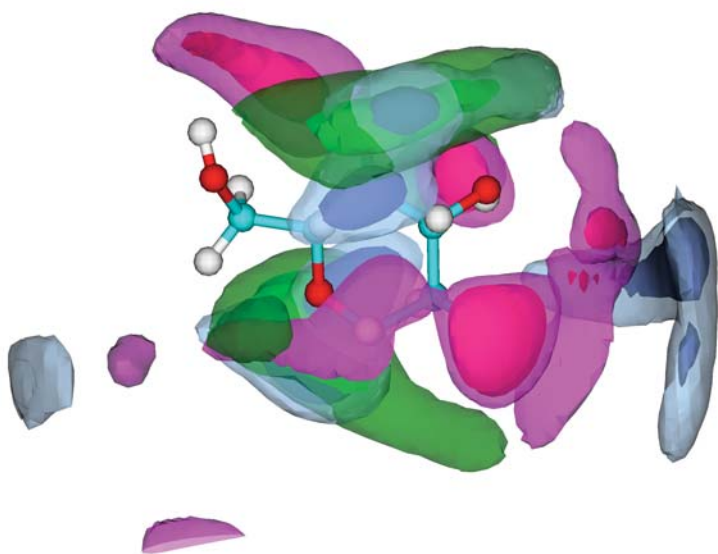
Figure 3.21. Combined three-dimensional number density profiles of the Cl⁻ anions (red/pink), Bmim⁺ rings (blue), and Bmim⁺ tails (green) at 450 K and 1 atm. Cl⁻ densities are drawn at 10 and 15 times pure solvent density at the same conditions, Bmim⁺ ring densities at 4 and 3 times pure solvent densities, and Bmim⁺ tail densities at 6 and 4 times pure solvent densities.
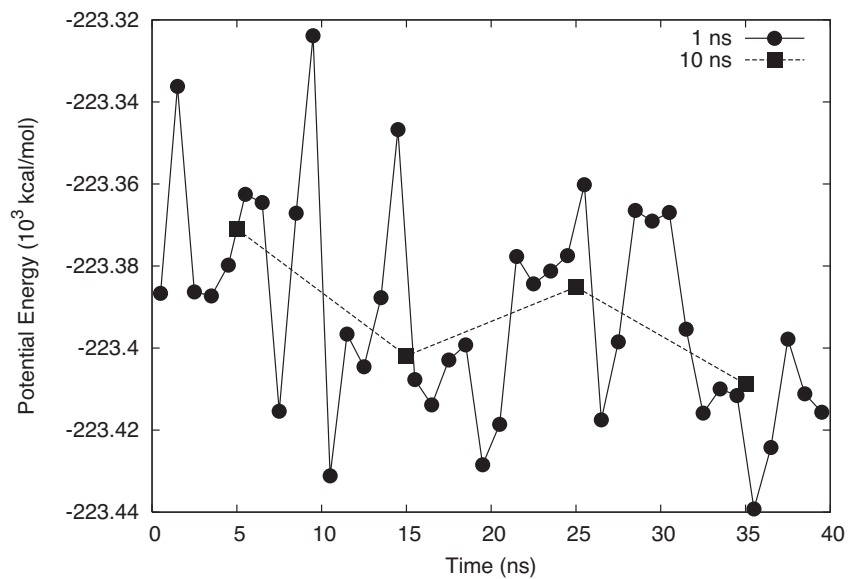
Figure 3.22. Running average of the total potential energy of the unconstrained dissociated cellulose/water system during glucan chain collapse using both 1 ns and 10 ns window averages.

References

(1)   Schubert, C. *Nat. Biotechnol.* **2006**, *24*, 777-784.

(2)   Huber, G. W.; Iborra, S.; Corma, A. *Chem. Rev.* **2006**, *106*, 4044-4098.

(3)   Farrell, A. E.; Plevin, R. J.; Turner, B. T.; Jones, A. D.; O'Hare, M.; Kammen, D. M. *Science* **2006**, *311*, 506-508.

(4)   Wyman, C. E. *Appl. Biochem. Biotech.* **2001**, *91-3*, 5-21.

(5)   Ragauskas, A. J.; Williams, C. K.; Davison, B. H.; Britovsek, G.; Cairney, J.; Eckert, C. A.; Frederick, W. J.; Hallett, J. P.; Leak, D. J.; Liotta, C. L.; Mielenz, J. R.; Murphy, R.; Templer, R.; Tschaplinski, T. *Science* **2006**, *311*, 484-489.

(6)   Demirbas, A. *Energ. Convers. Manage.* **2008**, *49*, 2106-2116.

(7)   Perlack, R. D. *Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasability of a Billion-Ton Annual Supply*; DOE/GO-102995-2135; Oak Ridge National Laboratory: Oak Ridge, TN, 2005.

(8)   Lynd, L. R.; Wyman, C. E.; Gerngross, T. U. *Biotechnol. Progr.* **1999**, *15*, 777-793.

(9)   Perez, S.; Mazeau, K. In *Polysaccharides: Structural Diversity and Functional Versatility*, 2nd ed.; Dumitriu, S., Ed.; Marcel Dekker: New York, 2005; pp 41-68.

(10)   Yang, B.; Wyman, C. E. *Biofuel. Bioprod. Bior.* **2008**, *2*, 26-40.

(11)   Dale, B. *J. Agr. Food. Chem.* **2008**, *56*, 3885-3891.

(12)   Simmons, B. A.; Loque, D.; Blanch, H. W. *Genome. Biol.* **2008**, *9*, 242.

(13)   Wyman, C. E. *Trends Biotechnol.* **2007**, *25*, 153-157.

(14)   Swatloski, R. P.; Spear, S. K.; Holbrey, J. D.; Rogers, R. D. *J. Am. Chem. Soc.* **2002**, *124*, 4974-4975.

(15)   Moulthrop, J. S.; Swatloski, R. P.; Moyna, G.; Rogers, R. D. *Chem. Commun.* **2005**, 1557-1559.

(16)   Singh, S.; Simmons, B. A.; Vogel, K. P. *Biotechnol. Bioeng.* **2009**, *104*, 68-75.

(17)   Himmel, M. E.; Ding, S. Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W.; Foust, T. D. *Science* **2007**, *315*, 804-807.

(18)   Gross, A. S.; Chu, J. W. *J. Phys. Chem. B* **2010**, *114*, 13333-13341.

(19)   Kondo, T. In *Polysaccharides: Structural Diversity and Functional Versatility*, 2nd ed.; Dumitriu, S., Ed.; Marcel Dekker: New York, 2005; pp 69-98.

(20)   Marechal, Y.; Chanzy, H. *J Mol Struct* **2000**, *523*, 183-196.

(21)   Watanabe, A.; Morita, S.; Ozaki, Y. *Biomacromolecules* **2006**, *7*, 3164-3170.

(22)   Watanabe, A.; Morita, S.; Ozaki, Y. *Biomacromolecules* **2007**, *8*, 2969-2975.

(23)   Kondo, T. *J Polym Sci Pol Phys* **1997**, *35*, 717-723.

(24)   Heiner, A. P.; Sugiyama, J.; Teleman, O. *Carbohyd Res* **1995**, *273*, 207-223.

(25)   Tashiro, K.; Kobayashi, M. *Polymer* **1991**, *32*, 1516-1530.

(26)   Bergenstrahle, M.; Berglund, L. A.; Mazeau, K. *J Phys Chem B* **2007**, *111*, 9138-9145.

(27)   Shen, T. Y.; Gnanakaran, S. *Biophys J* **2009**, *96*, 3032-3040.

(28)   Mazeau, K.; Heux, L. *J Phys Chem B* **2003**, *107*, 2394-2403.

(29)   Paavilainen, S.; Rog, T.; Vattulainen, I. *J Phys Chem B* **2011**, *115*, 3747-3755.

(30)   Cho, H. M.; Gross, A. S.; Chu, J.-W. *J Am Chem Soc* **2011**, *133*, 14033-14041.

(31)   Cousins, S. K.; Brown, R. M. *Polymer* **1995**, *36*, 3885-3888.

(32)   Bergenstrahle, M.; Wohlert, J.; Himmel, M. E.; Brady, J. W. *Carbohyd. Res.* **2010**, *345*, 2060-2066.

(33)   Payne, C. M.; Himmel, M. E.; Crowley, M. F.; Beckham, G. T. *J Phys Chem Lett* **2011**, *2*, 1546-1550.

(34)   El Seoud, O. A.; Koschella, A.; Fidale, L. C.; Dorn, S.; Heinze, T. *Biomacromolecules* **2007**, *8*, 2629-2647.

(35)   Pinkert, A.; Marsh, K. N.; Pang, S. S.; Staiger, M. P. *Chem. Rev.* **2009**, *109*, 6712-6728.

(36)   Remsing, R. C.; Swatloski, R. P.; Rogers, R. D.; Moyna, G. *Chem. Commun.* **2006**, 1271-1273.

(37)   Remsing, R. C.; Hernandez, G.; Swatloski, R. P.; Massefski, W. W.; Rogers, R. D.; Moyna, G. *J. Phys. Chem. B* **2008**, *112*, 11071-11078.

(38)   Youngs, T. G. A.; Holbrey, J. D.; Deetlefs, M.; Nieuwenhuyzen, M.; Gomes, M. F. C.; Hardacre, C. *Chemphyschem* **2006**, *7*, 2279-2281.

(39)   Youngs, T. G. A.; Hardacre, C.; Holbrey, J. D. *J. Phys. Chem. B* **2007**, *111*, 13765-13774.

(40)   Liu, H. B.; Singh, S.; Sale, K. L.; Simmons, B. A. *J Phys Chem B* **2011**, *115*, 10251-10258.

(41)   Liu, H. B.; Sale, K. L.; Holmes, B. M.; Simmons, B. A.; Singh, S. *J. Phys. Chem. B* **2010**, *114*, 4293-4301.

(42)   Mackerell, A. D.; Wiorkiewiczkuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946-11975.

(43)   MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586-3616.

(44)   Foloppe, N.; MacKerell, A. D. *J. Comput. Chem.* **2000**, *21*, 86-104.

(45)   Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187-217.

(46)   Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; Mackerell, A. D. *J. Comput. Chem.* **2008**, *29*, 2543-2564.

(47)   Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. *J. Chem. Theory Comput.* **2009**, *5*, 2353-2370.

(48)   Pomelli, C. S.; Chiappe, C.; Vidis, A.; Laurenczy, G.; Dyson, P. J. *J. Phys. Chem. B* **2007**, *111*, 13014-13019.

(49)   Holbrey, J. D.; Reichert, W. M.; Reddy, R. G.; Rogers, R. D. *Acs Sym. Ser.* **2003**, *856*, 121-133.

(50)   Cadena, C.; Zhao, Q.; Snurr, R. Q.; Maginn, E. J. *J. Phys. Chem. B* **2006**, *110*, 2821-2832.

(51)   Yan, T. Y.; Burnham, C. J.; Del Popolo, M. G.; Voth, G. A. *J. Phys. Chem. B* **2004**, *108*, 11877-11881.

(52)   Habasaki, J.; Ngai, K. L. *J. Chem. Phys.* **2008**, *129*, 194501.

(53)   Youngs, T. G. A.; Holbrey, J. D.; Mullan, C. L.; Norman, S. E.; Lagunas, M. C.; D'Agostino, C.; Mantle, M. D.; Gladden, L. F.; Bowron, D. T.; Hardacre, C. **2011**, *2*, 1594-1605.

(54)   Bowron, D. T.; D'Agostino, C.; Gladden, L. F.; Hardacre, C.; Holbrey, J. D.; Lagunas, M. C.; McGregor, J.; Mantle, M. D.; Mullan, C. L.; Youngs, T. G. A. *J Phys Chem B* **2010**, *114*, 7760-7768.

(55)   Nishiyama, Y.; Langan, P.; Chanzy, H. *J. Am. Chem. Soc.* **2002**, *124*, 9074-9082.

(56)   Ding, S. Y.; Himmel, M. E. *J. Agr. Food Chem.* **2006**, *54*, 597-606.

(57)   Horii, F.; Yamamoto, H.; Kitamaru, R.; Tanahashi, M.; Higuchi, T. *Macromolecules* **1987**, *20*, 2946-2949.

(58)   Beckham, G. T.; Matthews, J. F.; Peters, B.; Bomble, Y. J.; Himmel, M. E.; Crowley, M. F. *J Phys Chem B* **2011**, *115*, 4118-4127.

(59)   Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-935.

(60)   Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781-1802.

(61)   Martyna, G. J.; Tobias, D. J.; Klein, M. L. *J. Chem. Phys.* **1994**, *101*, 4177-4189.

(62)   Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613-4621.

(63)   Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press ; Oxford University Press: Oxford England; New York, 1987.

(64)   Wilkes, J. S.; Levisky, J. A.; Wilson, R. A.; Hussey, C. L. *Inorg. Chem.* **1982**, *21*, 1263-1264.

(65)   Shen, T. Y.; Langan, P.; French, A. D.; Johnson, G. P.; Gnanakaran, S. *J. Am. Chem. Soc.* **2009**, *131*, 14786-14794.

(66)   Umemura, M.; Yuguchi, Y.; Hirotsu, T. *J. Phys. Chem. A* **2004**, *108*, 7063-7070.

(67)   De Podesta, M. *Understanding the Properties of Matter*, 2nd ed.; Taylor & Francis: London ; New York, 2001.

(68)   Schwarz, F. P. *J. Solution. Chem.* **1996**, *25*, 471-484.

(69)   Jasra, R. V.; Ahluwalia, J. C. *J. Solution. Chem.* **1982**, *11*, 325-338.

(70)   Vietor, R. J.; Newman, R. H.; Ha, M. A.; Apperley, D. C.; Jarvis, M. C. *Plant. J.* **2002**, *30*, 721-731.

(71)   Chandler, D. *Nature* **2005**, *437*, 640-647.

(72)   Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; DiStasio, R. A.; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Van Voorhis, T.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.;

Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachsel, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C. P.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simmonett, A. C.; Subotnik, J. E.; Woodcock, H. L.; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W. J.; Schaefer, H. F.; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172-3191.

(73)   Matthews, J. F.; Bergenstrahle, M.; Beckham, G. T.; Himmel, M. E.; Nimlos, M. R.; Brady, J. W.; Crowley, M. F. *J Phys Chem B* **2011**, *115*, 2155-2166.

(74)   Saalwachter, K.; Burchard, W.; Klufers, P.; Kettenbach, G.; Mayer, P.; Klemm, D.; Dugarmaa, S. *Macromolecules* **2000**, *33*, 4094-4107.

Chapter 4

Entropy of Cellulose Dissolution in Water and in the Ionic Liquid 1-Butyl-3-Methylimidazolim Chloride

Abstract

The entropic driving forces of cellulose dissolution in water and in the ionic liquid 1-butyl-3-methylimidazolium chloride (BmimCl) are investigated via molecular dynamics simulations and the two-phase thermodynamic model. An atomistic model of cellulose was simulated at a dissociated state and a microfibril state to represent dissolution. The calculated values of entropy and internal energy changes between the two states inform the interplay of entropic and energetic driving forces in cellulose dissolution. In both water and BmimCl, we found that the entropy associated with the solvent degrees of freedom (DOF) decreases upon cellulose dissolution. However, solvent entropy reduction in BmimCl is much smaller than that in water and counteracts the entropy gain from the solute DOF to a much lesser extent. Solvent entropy reduction in water also plays a major role in making the free energy change of cellulose dissolution unfavorable at room temperature. In BmimCl, the interaction energies between solvent molecules and glucan chains and the total entropy change both contribute favorably to the dissolution free energy of cellulose. Calculations at different temperatures in the two solvents indicate that changes in total internal energy are a good indicator of the sign of the free energy change of cellulose dissolution.

4.1 Introduction

Of all the structural components in the plant cell wall, cellulose is the most abundant (40-50 wt %) and robust. Others include amorphous hemicellulose, which cross-links with cellulose, and the random polymer lignin, which encases the cellulose-hemicellulose composite in a supportive sheath.[1-3] Although biologically synthesized lignocellulosic materials potentially can be used as a feedstock for producing liquid transportation fuels and other chemicals,[1,4-7] the chemistry and organization of its aforementioned building blocks render the material highly recalcitrant towards degradation. Therefore, without pretreatment of the feedstock to separate and disaggregate the different components of lignocellulose, downstream synthesis of fuel molecules is slow. Consequently, an understanding of the factors controlling cellulose disaggregation should prove useful in guiding the development of viable technology for the conversion of biomass to fuels.[8-10]

Cellulose is composed of linear polymers of β-1,4-linked glucose monomers. Earlier efforts on resolving the molecular causes of cellulose recalcitrance revealed much information on its structural origins. It is well established that microfibrils are the main form of glucan chains in biomass. The crystalline structures of cellulose microfibrils are collectively known as cellulose I.[3,11,12] Since the surfaces of a cellulose microfibril expose different glucan residue moieties, distinct patterns in the organization of nearby solvent molecules are observed near different crystal faces.[13,14] The potentials of mean

force (PMFs) of deconstructing a glucan chain from a cellulose microfibril[15-17] and the solvent-glucan interaction energies at different stages of the deconstruction process[18] have also been resolved via molecular simulation. This work aims to supplement the current knowledge of cellulose recalcitrance by characterizing the entropic driving forces in the dissolution of a cellulose microfibril.

X-ray and neutron scattering of crystalline cellulose establishes that the interaction network in microfibrils consists of lateral OH—O hydrogen bonds (HBs) that organize glucan chains into flat sheets and axial intersheet CH—O contacts (pseudo HBs) and van der Waals interactions that stack sheets together.[19,20] Intersheet contacts have been identified to be most responsible for cellulose recalcitrance and stability.[13,15,16,18,21] Although the interaction network in cellulose is exceedingly robust, it can be completely disrupted in certain classes of ionic liquids (ILs),[22-24] specifically those with a heterocyclic ring functionalized with alkyl tail groups as the cation and a small conjugate base as the anion.[25,26] A canonical example is 1-butyl-3-methylimidazolium chloride (BmimCl), in which cellulose solubility can reach as high as 25 wt %.[22] Resolving the molecular driving forces provided by ILs such as BmimCl to dissolve cellulose can reveal the fundamental principles for overcoming biomass recalcitrance.

Both NMR measurements, quantum chemistry calculations, and atomistic molecular dynamics (MD) simulations have shown that IL anions strongly interact with the hydroxyl groups of glucan chains via OH—anion HBs,[14,18,27-34] while quantum mechanical calculations have shown that the acidic hydrogens of the cation ring only weakly interact with the glucan OH groups.[32,34] Empirically, cellulose solubility has been found to be sensitive to the molecular architecture of the cations.[25] Analysis of the three-dimensional density profiles of BmimCl around dissolved glucan chains[14] and dissection of the force interactions between BmimCl and glucose residues during cellulose deconstruction reveal the role of cations as interacting with the ether linkages along the glucan backbone and the hydroxymethyl side chain off the sugar rings.[18] The PMF of transferring a glucan chain from the microfibril surface to the BmimCl phase was calculated to be around -2.0 kcal/mol/glucan-residue.[18] In contrast, the potential of mean force (PMF) of transferring the glucan chain to water is unfavorable, around 2.0 kcal/mol/glucan.[16,18] Pairwise PMFs of solvent-glucan interactions indicate that water lacks driving forces to interact with sugar rings and linker oxygen. Conversely, in BmimCl, Bmim$^+$ cations strongly couple to side chains and linker oxygens and Cl$^-$ anions bind to OH groups to facilitate cellulose dissolution.[18]

Current understanding of IL-induced cellulose dissolution is based primarily on the structural details and PMFs calculated with molecular simulations. An essential piece of information that is still missing is the interplay between the energetic and entropic driving forces in cellulose dissolution. This information is particularly important for guiding the molecular design and engineering of pretreatment solvents for mitigating cellulose recalcitrance. To shed light on the role of entropic driving forces, in this work we compare the behaviors of cellulose in a crystalline microfibril state and in a dissolved state via atomistic MD simulations.

4.2 Methods

Following reference 14, dissolution of a microfibril is modeled by arranging 36 glucan chains, each 16 residues long, in either a mutually dissociated state (D) or in a crystalline microfibril state (M) (Figure 4.1). The two states were solvated in either TIP3P[35] water (28443 waters) or BmimCl (12180 IL pairs). The BmimCl force field used was developed and validated previously.[14] For the dissociated state, at least 4 solvation shells of solvent were added to separate individual chains, and the same number of solvent molecules was also used to solvate the microfibril state. All MD runs were performed with the NAMD software[36] at the designated temperature (425, 450, and 500 K in BmimCl and 300, 325, and 350 K in water) and pressure (1 atm) using a Langevin thermostat and Nose-Hoover Langevin barostat[37,38] with periodic boundary conditions. From previously conducted atomistic MD simulations,[14] 15 100 ps simulations were started and run from restarts at regularly spaced intervals of 1 ns for cellulose/water systems and 2 ns for cellulose/BmimCl systems. In both the microfibril and dissociated states, harmonic restraints were put on the C1 carbons of all terminal glucans in the MD simulations (force constant 5 kcal/mol/Å^2 for water systems and 10 kcal/mol/Å^2 for IL systems). This allowed for a representation of cellulose with glucan chains much shorter than the commonly observed degrees of polymerization found in nature, and the use of an impractically large simulation cell at the atomistic scale was avoided.[14] More details of the simulation setup and MD procedures can be found in reference 14. The entropies of the simulated cellulose/solvent systems were calculated with the two-phase thermodynamic (2PT) model.[39]

In the 2PT method, the total entropy of a molecular system is calculated as the sum of the entropies of the translational, rotational, and intramolecular degrees of freedom (DOF) of each atom.[39,40] The key step is fitting the velocity density of states of each atom to a linear combination of two reference systems, an ideal solid and a model hard sphere gas, for which the absolute entropies are known as a function of density of states.

$$\rho(\omega) = \rho_g(\omega) + \rho_{HO}(\omega)$$

The velocity density of states is the temporal Fourier transform of the mass-weighted velocity autocorrelation function.

$$\rho(\omega) = \frac{2}{kT} \lim_{\tau' \to \infty} \int_{-\tau'}^{\tau'} \left[\lim_{\tau \to \infty} \frac{m}{2\tau} \int_{-\tau}^{\tau} v(t'+t)v(t')dt \right] e^{-i\omega t} dt$$

With the 2PT model, the total entropy of a molecular system can be estimated from an MD trajectory by applying this two-phase representation to all atoms and summing their entropy values together. Error bars in entropy values were calculated via bootstrapping methods with the 15 100 ps trajectories of each solvent/cellulose conformation combination. More details of our 2PT calculations can be found in the supporting information (SI).

The 2PT approximation has been applied to calculate the absolute entropies of Lennard-Jones fluids, water, and molecular solvents. Quantitative agreement of the 2PT entropy values with experimental data and with the results of other simulation methods have been demonstrated repeatedly.[39-43] A novelty of this work is applying the 2PT

method to compute the entropies of the dissociated and microfibril states of cellulose in water and BmimCl to reveal the entropic driving forces of cellulose dissolution.

4.3 Results and Discussion

Contributions from the solvent degrees of freedom to the entropy of cellulose dissolution ($\Delta S_{DM}^{solvent} = S_D^{solvent} - S_M^{solvent}$) are shown in Figure 4.2A. In both water and BmimCl, separating glucan chains to the dissociated state leads to an entropy reduction of the solvent molecules, the translational DOF constituting the highest percentage of the change. In water, the translational DOF contributes ~90% of the solvent entropy reduction. There is no contribution from the intramolecular DOF to the entropy reduction of water due to the use of the rigid TIP3P[35] model. The entropy reduction of BmimCl upon cellulose dissolution in the investigated temperature range is only 30-60% of that of water. Regarding the total reduction in solvent entropy, the translational DOF of BmimCl contribute ~70%, the rotational DOF ~25%, and the intramolecular DOF the remaining 5%.

The apparent solvent entropy reduction in both water and BmimCl can be rationalized by the fact that the number of solvent molecules bound to the glucans and therefore constricted due to solvent-glucan interactions[18] is higher in the dissociated state than in the microfibril state. From the three-dimensional density profiles, solvent-glucan interactions lead to localization of water and IL molecules into specific regions around the sugars and hence limitation of their positional flexibility.[13,14,29-31] Coupled with the extended structure of the glucan chains, this effect creates a long-range network of glucan-associated solvent molecules. As an example, water entropy in either the dissociated or microfibril simulation is less than that of neat water (See the SI). Despite this common trend, solvent entropy reductions in water and BmimCl have very different behaviors. Unlike the pronounced temperature dependence of water entropy reduction, reduction in BmimCl entropy upon cellulose dissolution in the investigated temperature range is nearly constant. The small size of water molecules leads to solvent structures and hydrophobicity that are very sensitive to temperature.[44] On the other hand, IL-IL interactions are ionic in nature, and a significant temperature dependence in the entropy reduction of BmimCl was not observed. As illustrated later, solvent entropy reduction plays a key role in directing the sign of the free energy change of cellulose dissolution.

The calculated entropy changes of glucan chains upon cellulose dissolution, ΔS_{DM}^{cell}, are shown in Figure 4.2B. Although the glucan chains are restrained around their initial positions by the harmonic constraints, they still possess translational and rotational entropy due to vibrational motion in either of these two modes. In both the dissociated and microfibril states, each glucan chain is treated with the 2PT method as a separate entity with its own translational, rotational, and intramolecular DOF. For the microfibril state, calculations of cellulose entropy by treating the entire 36 chains as one single body gives the same values of total entropy as those obtained by the chain-by-chain approach (See the SI for more details). The latter convention is used for consistent comparison with the dissociated state.

In water, dissolution leads to an entropy increase of glucan chains as expected. From our MD trajectories with the end restraints on the chains,[14] the dominant contribution to this increase in 2PT entropy comes from the intramolecular DOF of the

glucan chains. Glucan chains becoming more flexible after dissociation in water is thus the main cause of the entropy increase observed in Figure 4.2B, and is a driving force for dissolution. An earlier study on glucan chains of the same size showed that end restraints only prevent the net translation and rotation of glucans but do not lead to noticeable differences in their internal flexibility.[14] Therefore, underestimation of the entropy increase of the intramolecular DOF due to the end restraints is expected to be minimal. Entropy contributions from the net translation and rotation of individual chains will be incorporated later via analytical models.

In BmimCl, the net changes in cellulose entropy as a result of dissolution have much smaller magnitudes compared to those in water (Figure 4.2B). At 450 K and 500 K, only slight increases are observed. At 425 K, glucan chains actually have a lower value of entropy in the dissociated state. The apparent differences in the changes of cellulose entropy for dissolution in BmimCl (positive or negative, small) and in water (positive, large) are due to the distinct contributions coming from the intramolecular DOF. In the dissociated state of cellulose in BmimCl, the intramolecular DOF of glucan chains have lower entropy than those in the microfibril state; in water, the opposite trend is observed. Since BmimCl molecules couple to glucan chains much more strongly in the dissociated state based on the deconstruction PMFs and the pair interaction forces calculated from atomistic MD simulations,[18] our calculations indeed show that the intramolecular DOF of glucan chains are more restricted upon dissociation in the IL. Conversely, in water the deconstruction PMF is unfavorable and water molecules lack the driving forces to couple to the sugar rings and linker oxygens of glucan residues. Dissociation of cellulose in water thus leads to significant entropy increases in the intramolecular DOF of glucan chains as shown in Figure 4.2B. Without accounting for the effects of the end restraints on the net translation and rotation of glucan chains, the calculated increases in 2PT entropies for the translational and rotational DOF of cellulose upon dissolution are also much smaller (20-50%) in BmimCl than those in water, similar to the trend of the entropy changes of the intramolecular DOF.

Adding harmonic restraints on the chain ends enables finite-size simulations of a cellulose microfibril and maintains a consistent dissociated state in water and in BmimCl for comparison. However, without accounting for the net translation and rotation of glucan chains, the total entropies of the dissociation process are underestimated. As an estimation of the entropic contributions of these DOF to cellulose dissociation, the entropies of rigid-body translation and rotation[40,45] of individual glucan chains and the microfibril are calculated using the averaged system volumes and moments of inertia as input parameters. The inclusion of these increases the dissolution entropies by ~6 cal/K/mol-glucan. The total dissolution entropy of cellulose calculated by the 2PT method, $\Delta S_{\mathrm{DM}}^{\mathrm{2PT}}$, and the total dissolution entropy with the estimated contributions from the net translation and rotation of the dissociated chains and the microfibril added to $\Delta S_{\mathrm{DM}}^{\mathrm{2PT}}$, ΔS_{DM}, are both shown in Figure 4.3. In water and in BmimCl, the values of $\Delta S_{\mathrm{DM}}^{\mathrm{2PT}}$ are negative at all simulated temperatures due to the solvent DOF. Taking the net translation and rotation of glucan chains into account, ΔS_{DM} is positive for all systems except for cellulose dissolution in water at 300 K.

The results shown in Figures 4.2-3 indicate that for cellulose dissolution in water and in BmimCl, the entropy gain from glucan chains becoming free to translate and rotate

is counteracted to different extents by solvent entropy reduction. In BmimCl, ΔS_{DM} remains positive and the solvent entropy reduction is not sensitive to temperature. Conversely, in water, a high temperature dependence of the solvent entropy reduction is observed, and ΔS_{DM} is in fact negative at a lower temperature of 300 K. Overall, the calculated values of ΔS_{DM} are mostly positive, indicating that entropy is likely a driving force for cellulose dissolution.

To estimate the Helmholtz free energy change of cellulose dissolution, ΔA_{DM}, the values of ΔS_{DM} are combined with the reported values of internal energy change of cellulose dissolution, ΔE_{DM}.[14] The ΔE_{DM} value in water is ~0 at room temperature and becomes increasingly positive at higher temperatures. In BmimCl, the value of ΔE_{DM} is negative at all investigated temperatures. The calculated ΔE_{DM} values also agree well with the heats of solution of monosaccharides in water.[46,47] The calculated values of ΔA_{DM} are shown in Figure 4.4. ΔA_{DM} is positive in water and negative in BmimCl, consistent with the experimental observation that cellulose can be dissolved in BmimCl but not in water. These values, 0.5 kcal/mol-glucan in water and -2.0 kcal/mol-glucan in BmimCl, are also of the same sign and similar magnitudes as those calculated with a thermodynamic integration approach on a distinct but related process of cellulose deconstruction.[18] Also, as Figure 4.4 shows, the magnitudes of the energetic and entropic terms are of the same order, and neither overwhelms the other in determining the free energy change of dissolution.

4.4 Conclusions

With the 2PT method, we show that the energetic and entropic driving forces are not favorable for cellulose dissolution in water at 300 K. At 325 K and 350 K in water, although the total entropy change shifts sign to positive, the internal energy change shifts in the opposite direction, and the resulting free energy change is unfavorable. In BmimCl, both energetic and entropic driving forces favor cellulose dissolution at all investigated temperatures. The stronger interactions between the IL molecules and glucan chains at the dissolved state lead to negative entropy changes from the intramolecular DOF of the solute and the translational DOF of the solvent, counteracting some of the entropy gains from the net translation and rotation of the dissolved glucan chains.

Furthermore, we show that the magnitudes of energetic and entropic driving forces are both in the range of 1-3 k_BT. The favorable free energy change of cellulose dissolution is thus a delicate balance of the energetic and entropic driving forces emergent from complex molecular structures and interactions. A repeating pattern is that the energy of dissolution is a predictor of the sign of the dissolution free energy. Out of the six calculated free energy values, only two have $|\Delta E_{DM}| < |T\Delta S_{DM}|$, and in these two cases, the sign of ΔA_{DM} is consistent with that of ΔE_{DM} (see the SI for further details). The calculations in this work thus suggest that internal energy change could be used as an effective indicator for *in silico* screening of cellulose pretreatment solvents. Testing of this proposition is ongoing and will appear in future work.

4.5 Acknowledgements

4.6 Supporting Information

4.6.1 Application of the Two-Phase Thermodynamic Model to Determine Cellulose Dissolution Entropy

The two-phase thermodynamic (2PT) method[39,40] was used to calculate the entropy of water, BmimCl, and cellulose in order to determine the entropy of solution of cellulose in those two solvents. From previously conducted simulations,[14] 15 100 ps simulations were started and run from restarts at regularly spaced intervals of 1 ns for cellulose/water systems and 2 ns for cellulose/BmimCl systems. Position and velocity data was collected every 4 fs for analysis. The 2PT entropy of each component, solvent or cellulose, was then calculated in the following manner:

1. Calculation of the velocity time autocorrelation functions

The velocity of each atom at each time step was split up into translational, rotational, and intramolecular components

$$\vec{v}_i(t) = \vec{v}_i^{\,trans}(t) + \vec{v}_i^{\,rot}(t) + \vec{v}_i^{\,\text{intra}}(t)$$

where i refers to atom and t to time. The translational velocity is calculated as

$$\vec{v}_i^{\,trans}(t) = \frac{\sum_{j \in M} m_j \vec{v}_j(t)}{\sum_{j \in M} m_j}$$

where m_j refers to the mass of atom j, and the sum is over all atoms within a single molecule or body (each atom can only belong to one molecule). The translational velocity is the velocity of the center of mass of the body to which the selected atom belongs.

The rotational velocity of an atom is calculated with the following equations

$$\vec{L}_M(t) = \sum_{j \in M} m_j \left(\vec{r}_{jM}(t) \times \vec{v}_j(t) \right) = \underline{\underline{I_M(t)}} \vec{\omega}_M(t)$$

$$\vec{v}_i^{\,rot}(t) = \vec{\omega}_M(t) \times \vec{r}_{iM}(t)$$

L_M is the angular momentum of molecule M, to which atom i belongs, r_{jM} is the position vector between atom j and the center of mass of body M, I_M is the moment of inertia tensor of M, and ω_M the angular velocity of M. As in the case for the translational velocity, the summation is over all atoms that are part of the body M. In practice, the rotational velocity of atom i is found by inverting the moment of inertia tensor, and multiplying it with the angular momentum vector to calculate the angular velocity vector. The rotational velocity is the cross product of this with the position vector.

The intramolecular velocity is then the balance

$$\vec{v}_i^{\,\text{intra}}(t) = \vec{v}_i(t) - \vec{v}_i^{\,trans}(t) - \vec{v}_i^{\,rot}(t)$$

With the three velocities types calculated for all atoms at all time points, the velocity time autocorrelation functions of each atom type are calculated

$$C_j^x(t) = \left\langle v_j^x(\tau + t) \cdot v_j^x(\tau) \right\rangle$$

C is the autocorrelation function, j refers to a specific atom type, and x refers to type of motion or degree of freedom (translational, rotational, or intramolecular).

2. Calculation of the density of states functions

The density of states function of each molecular species for each degree of freedom type (translational, rotational, or intramolecular) is the mass weighted Fourier transform of the velocity time autocorrelation function associated with that type of motion.[48]

$$\rho^x(\omega) = FT\left[C^x(t)\right] = FT\left[N_M \sum_{j \in M} m_j \left\langle v_j^x(\tau + t) \cdot v_j^x(\tau) \right\rangle\right]$$

N_M is the number of molecules of type M, x again refers to degree of freedom type, j refers to atom type, m_j is the mass of atom type j, ω is frequency, and the summation is over all atoms types that exist in molecule type M. Each molecule type will have its own density of states functions associated with it. FT stands for Fourier Transform.

After the density of state functions are calculated, they are split up according to the 2PT methodology. This is done via a fluidity factor, which ranges from zero to one. Values closer to zero indicate a more solid-like system, while values near one a more fluid-like one.[40] The factor is a function of the thermodynamic state point of the system, as well as $\rho(0)$, the zero frequency density of state.[39,40] Once it is calculated, the density of states is split into two parts, one that represents a hard sphere gas, and one that represents an ideal, harmonic solid. This division of the density of states function is the key step in the 2PT methodology.

$$\rho_g^x(\omega) = \frac{\rho^x(0)}{1 + \left[\dfrac{\pi \rho^x(0)\omega}{6f^x N}\right]^2}$$

$$\rho_s^x(\omega) = \rho^x(\omega) - \rho_g^x(\omega)$$

The subscript g and s denote either the gas or solid system, respectively, N is the number of that type of molecule, and f is the fluidity factor. Note the following:

$$\rho_s^x(0) = 0$$

$$\rho_g^{intra}(\omega) = 0 \text{ for all } \omega$$

Also, integration of the density of states function gives the number of degrees of freedom associated with that type of motion.

3. Calculation of the Entropy

      With the density of states functions known, the entropy of each molecule type can be calculated

$$S^x = k_b \left[\int \rho_s^x(\omega) S_{HO}(\omega) d\omega + \int \rho_g^x(\omega) S_{HS}^x \, d\omega \right]$$

$$S_{HO}(\omega) = \frac{\beta h \omega}{\exp(\beta h \omega) - 1} - \ln[1 - \exp(-\beta h \omega)]$$

$$S_{HS}^{trans} = \frac{5}{2} + \ln\left[\left(\frac{2\pi m k_b T}{h^2} \right)^{3/2} \frac{V}{f^{trans}} z(y) \right] + \frac{y(3y-4)}{(1-y)^2}$$

$$S_{HS}^{rot} = \ln\left[\left(\frac{\pi^{1/2} e^{3/2}}{\sigma} \right) \left(\frac{T^3}{\Theta_A \Theta_B \Theta_C} \right)^{1/2} \right]$$

S^x is the translational, rotational, or intramolecular entropy, and is the sum of that of the ideal solid and hard sphere gas entropies. S_{HO} is the entropy of an ideal harmonic oscillator as a function of frequency. S_{HS} is the entropy of the hard sphere gas, the exact formula of which depends on whether the translational or rotational entropy is being calculated. In the translational hard sphere entropy equation, f is the translational fluidity factor, y is a function of the state point of the system as well as the zero frequency density of state, and z(y) is the Carnahan-Starling equation of state of a hard sphere gas.[40] In the rotational hard sphere entropy equation, the Θs are the three rotational temperatures of the molecule type, and σ is the symmetry factor.[40,45] Note that the intramolecular entropy only has contributions from the ideal solid term, as its gas density of states function is always zero.

      After the translational, rotational, and intramolecular are calculated for each molecule type in the system, they are summed together to give the total system entropy. The entropy of dissolution of cellulose in either water or BmimCl is then calculated as

the differences in entropy values between the dissociated and microfibril simulations (dissociated – microfibril). The values presented here are the results of 15 simulations for each solvent/cellulose conformation combination and temperature choice.

4.6.2 Microfibril Degree of Freedom Partition and Entropy

The translational, rotational, and intramolecular time autocorrelation functions must be calculated as part of the 2PT analysis. Embedded in this process is the choice of what collection of sites constitutes an independent body. For instance, the translational velocities of each body are calculated as the mass-weighted sum of the velocities of each atom within that body. The rotational velocities are calculated according to the formulas for rigid-body rotation.[40,45] For a small molecule like water or $Bmim^+$, the choice of what sites should define each body is obvious, a set of atoms connected by covalent bonds to each other, but not connected to any other atoms by such bonds, i.e. all the atoms in a single molecule. The oxygen and two hydrogens of each water would constitute one such body. This definition is readily extended to the dissociated chain conformation of cellulose. However, for the microfibril, the case is not so simple. It can be viewed either as a collection of 36 individual chains, each with its own independent translational, rotational, and intramolecular DOF, or, as one collective body, wherein movement of one chain relative to another is considered internal motion. The decision of how to partition the microfibril's DOF will affect the absolute entropy values calculated for these DOF of the microfibril. In Table 4.1A-B, we compare the microfibril entropies calculated using both approaches. As expected, considering each chain as its own, independent entity (as done in the main text) yields a higher absolute entropy value for the microfibril's translational and rotational entropies compared to treating the microfibril as one collective body, as in the former method, cellulose has more DOF associated with these types of motion. The same reasoning explains why this first approach gives a lower intramolecular entropy compared to the second. Once relative motion between chains is considered intramolecular motion, and not translational or rotational motion, more DOF become associated with the intramolecular term. However, when considering the net entropy, both approaches are equivalent, and give the same total entropy for the microfibril. The choice of using the chain-by-chain approach was made for consistency with the dissociated state, which used that basis for calculating the glucan entropy.

4.6.3 Comparison of Water Entropy from the Cellulose Simulations with Neat Water Entropy

In Table 4.2A, we show the entropy of water from the dissociated and microfibril simulations, while in Table 4.2B, we show the entropy of neat water at 300. In either simulation, the effect of cellulose is to decrease the entropy of water relative to the neat case. Essentially, the entirety of the change occurs with translational entropy, which decreases by 0.4 cal/K/mol, while the rotational entropy is basically unchanged. The combination of the glucan chains' many OH groups along with its extended structure serve to bind nearby waters in a collective manner and limit their freedom more severely compared to the more localized, short range structure of pure water.

4.6.4 Pressure-Volume Contribution to the Free Energy Change of Cellulose Dissolution

In Table 4.3, we show the pV-work term associated with cellulose dissolution in water and BmimCl from our MD simulations. The pΔV values are negligible, at least 5 orders of magnitude smaller than the calculated Helmholtz free energy terms shown in Figure 4.4. Thus the Helmholtz and Gibbs free energy changes for this process are equivalent.

4.6.5 Internal Energy Change as a Predictor of the Sign of Free Energy Change for Cellulose Dissolution

In Table 4.4A-B we show the values of ΔE, ΔS, and ΔA of cellulose dissolution in either water or BmimCl calculated from our MD simulations. In all cases except for cellulose in water at 300 K and in BmimCl at 500 K, |ΔE|>TΔS|, and energy change determines the sign of the free energy change. For those two simulations where |ΔE|<TΔS|, ΔE still has the same sign as ΔA. Thus, energy of solution could be a good predictor for the sign of the free energy change. This is beneficial, as energy of solution is both quicker and simpler to calculate than entropy of solution, and could be used as an *in silico* test to screen potential cellulose solvents. In future work, this will be investigated to test whether this result extends to other solvents, IL or otherwise.

Table 4.1A. Comparison of 2PT entropies calculated for the cellulose microfibril in water using either the whole microfibril or each individual chain as the basis for the DOF decomposition of the glucans.

Temp (K)	300		325		350	
Basis	Whole Microfibril	Chain-by-Chain	Whole Microfibril	Chain-by-Chain	Whole Microfibril	Chain-by-Chain
Trans Entropy[a]	0.04	0.93	0.04	0.96	0.04	1.00
Rot Entropy[a]	0.04	0.81	0.04	0.84	0.04	0.88
Intra Entropy[a]	41.60	39.94	45.47	43.75	49.37	47.57
Total Entropy[a]	41.68	41.68	45.54	45.55	49.45	49.45

[a] cal/K/mol-gluc

Table 4.1B. Comparison of 2PT entropies calculated for the cellulose microfibril in BmimCl using either the whole microfibril or each individual chain as the basis for the DOF decomposition of the glucans.

Temp (K)	425		450		500	
Basis	Whole Microfibril	Chain-by-Chain	Whole Microfibril	Chain-by-Chain	Whole Microfibril	Chain-by-Chain
Trans Entropy[a]	0.04	1.06	0.04	1.08	0.04	1.13
Rot Entropy[a]	0.04	0.93	0.04	0.95	0.04	1.01
Intra Entropy[a]	59.50	57.59	63.00	61.05	70.04	67.98
Total Entropy[a]	59.58	59.58	63.07	63.08	70.12	70.13

[a] cal/K/mol-gluc

Table 4.2A. Entropy of water from the dissociated chain and microfibril simulations

Temp (K)	300		325		350	
Simulation	Dissociated	Microfibril	Dissociated	Microfibril	Dissociated	Microfibril
Trans Entropy[a]	13.419 ± 0.003	13.585 ± 0.002	14.245 ± 0.002	14.380 ± 0.001	14.991 ± 0.003	15.091 ± 0.003
Rot Entropy[a]	3.163 ± 0001	3.181 ± 0.001	3.667 ± 0.001	3.685 ± 0.000	4.166 ± 0.001	40181 ± 0.001
Total Entropy[a]	16.582 ± 0.003	16.766 ± 0.002	17.912 ± 0.003	18.065 ± 0.002	19.157 ± 0.004	19.272 ± 0.004

[a] cal/K/mol

Table 4.2B. Entropy of neat water at 300 K

Trans Entropy[a]	13.834
Rot Entropy[a]	3.160
Total Entropy[a]	16.994

[a] cal/K/mol!

105

Table 4.3. Pressure-volume contribution to the free energy change of cellulose dissolution.

Solvent	Temp (K)	ΔV (Å^3)	$p\Delta V$ (kcal/mol-gluc)
Water	300	-544.1	-1.38E-05
Water	325	-79.9	-2.02E-06
Water	350	215.8	5.46E-06
BmimCl	425	227.3	5.76E-06
BmimCl	450	869.0	2.20E-05
BmimCl	500	1533.8	3.88E-05

Table 4.4A. Energy, entropy and Helmholtz free energy change of dissolution of cellulose in water.

Temp (K)	300			325			350		
ΔE (kcal/mol-glucan)	0.057	±	0.037	0.769	±	0.029	1.183	±	0.031
$T\Delta S$ (kcal/mol-glucan)	-0.383	±	0.059	0.146	±	0.057	0.899	±	0.095
ΔA (kcal/mol-glucan)	0.440	±	0.070	0.623	±	0.064	0.283	±	0.100

Table 4.4B. Energy, entropy and Helmholtz free energy change of dissolution of cellulose in BmimCl.

Temp (K)	425			450			500		
ΔE (kcal/mol-glucan)	-1.726	±	0.111	-1.195	±	0.102	-0.669	±	0.069
$T\Delta S$ (kcal/mol-glucan)	0.947	±	0.252	0.967	±	0.286	1.390	±	0.284
ΔA (kcal/mol-glucan)	-2.673	±	0.276	-2.163	±	0.304	-2.059	±	0.292
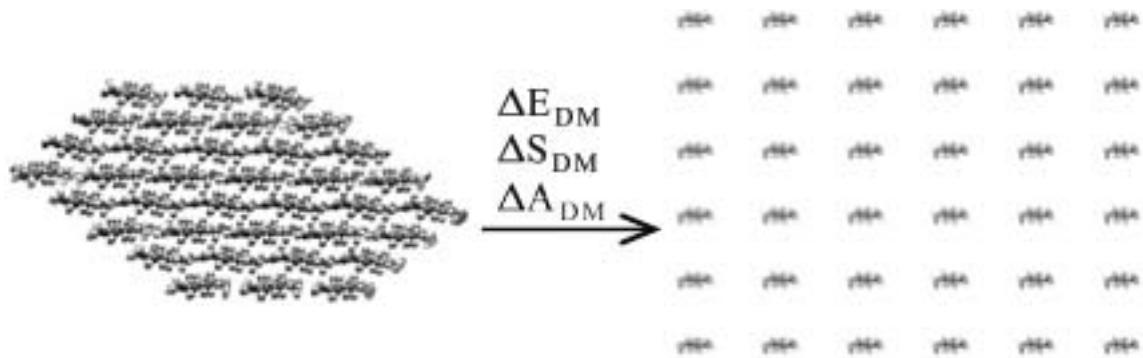
Figure 4.1. Cross view of the two different conformational states of cellulose used in this work to represent cellulose dissolution. Left: the microfibril state. Right: the dissociated state. Each state contains 36 glucan chains, with 16 glucans per chain.

(A)
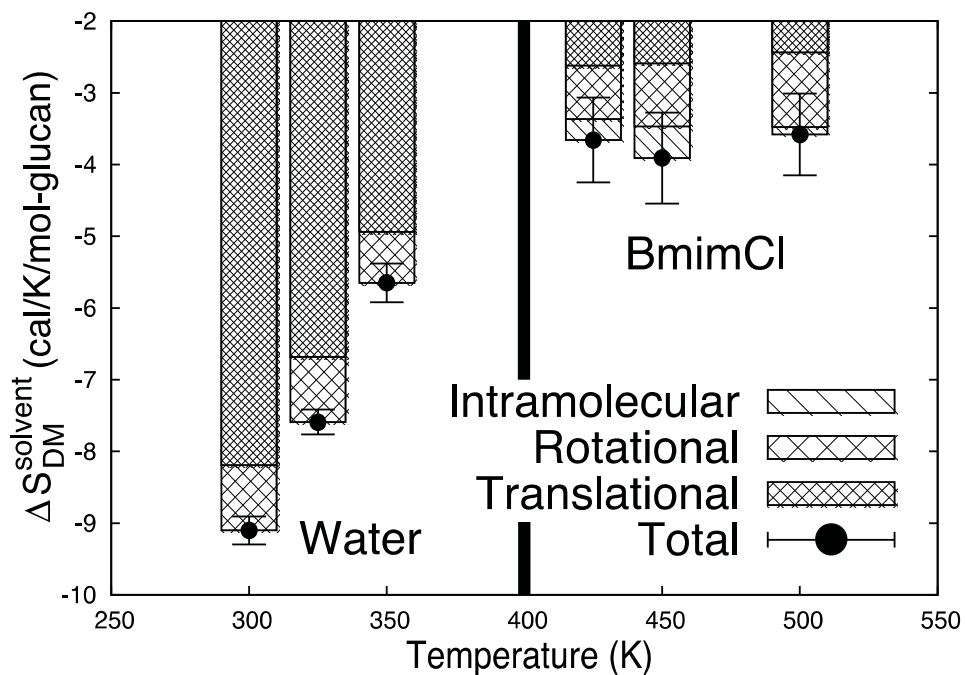
(B)
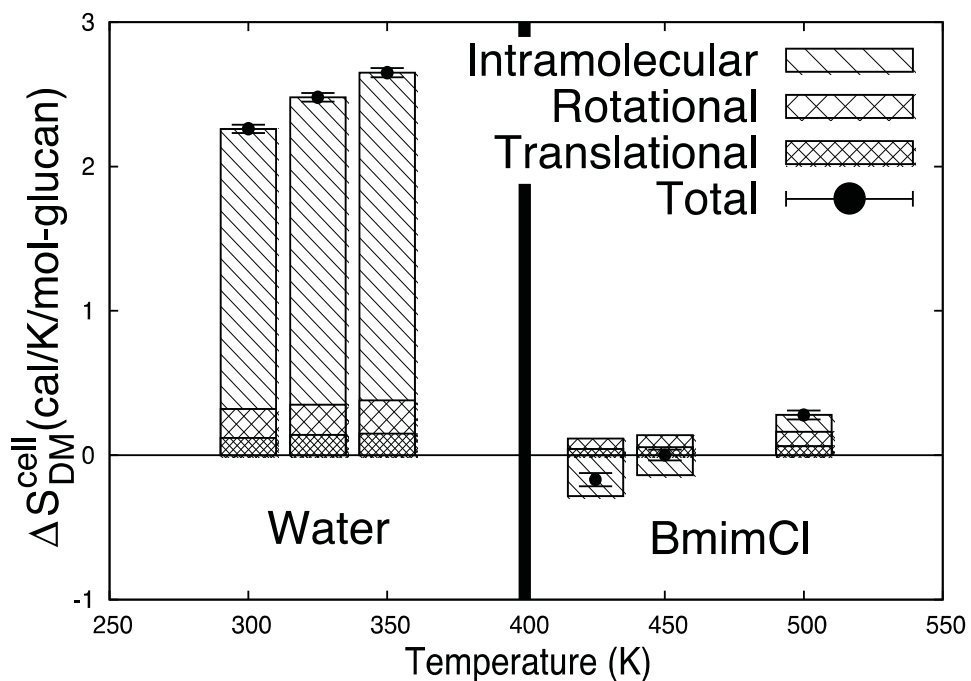
Figure 4.2. (A) The change in entropy of the solvent DOF upon cellulose dissolution in water and BmimCl. (B) The change in entropy of the cellulose DOF upon dissolution in water and BmimCl.
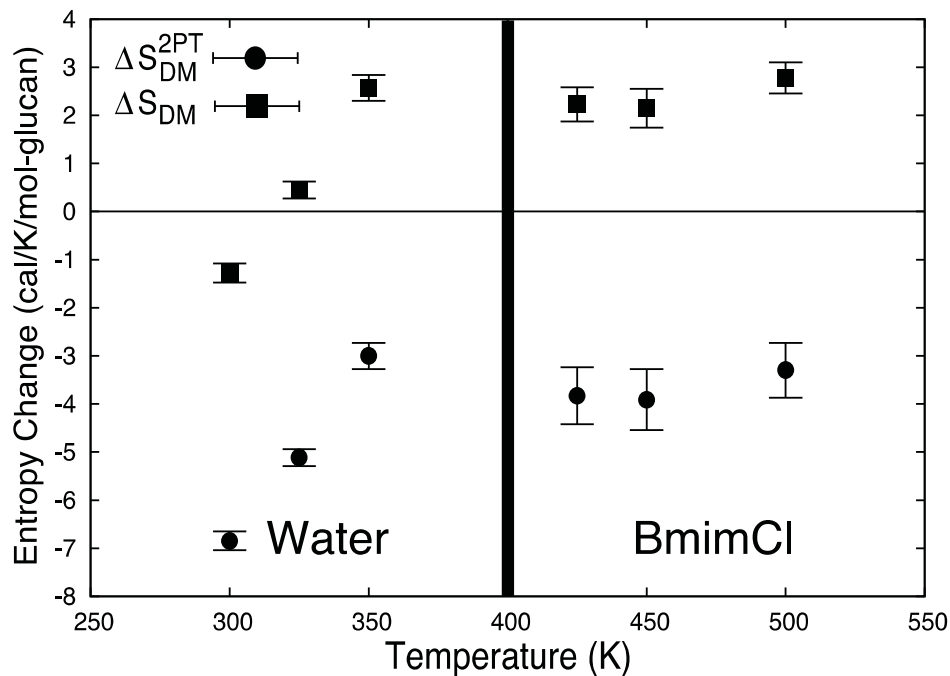
Figure 4.3. The change in total entropy of the simulation systems upon cellulose dissolution in water and BmimCl calculated with the 2PT method, ΔS_{DM}^{2PT}. Adding the estimated entropy changes for the net translation and rotation of glucan chains (see text for details) to ΔS_{DM}^{2PT} gives the total dissolution entropy, ΔS_{DM}.
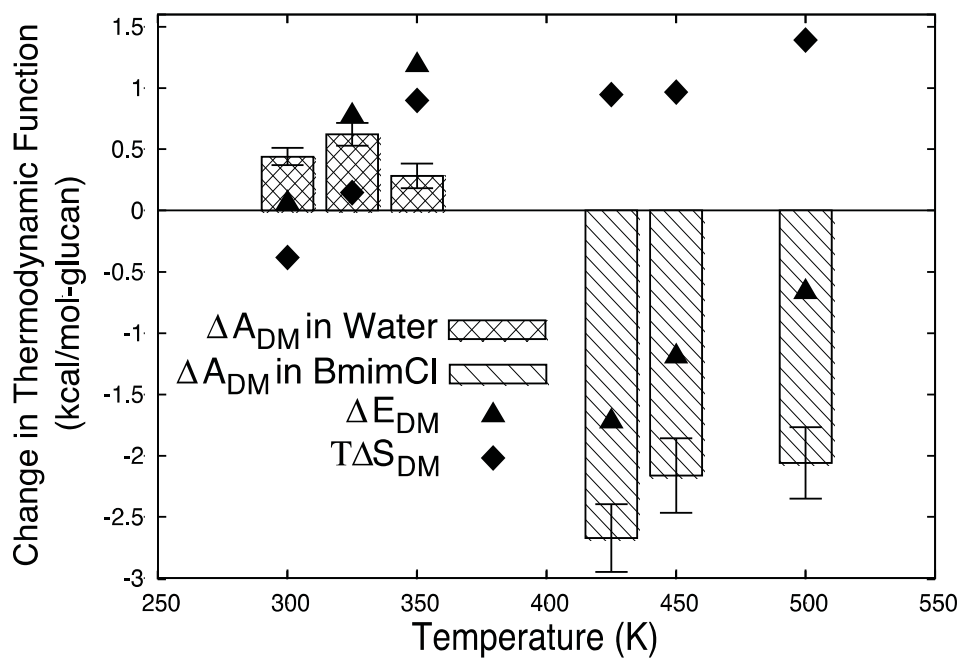
Figure 4.4. Energy, entropy, and Helmholtz free energy change of dissolution for cellulose in either water or BmimCl.

References

  (1) Ragauskas, A. J.; Williams, C. K.; Davison, B. H.; Britovsek, G.; Cairney, J.; Eckert, C. A.; Frederick, W. J.; Hallett, J. P.; Leak, D. J.; Liotta, C. L.; Mielenz, J. R.; Murphy, R.; Templer, R.; Tschaplinski, T. *Science* **2006**, *311*, 484-489.

  (2) Somerville, C.; Bauer, S.; Brininstool, G.; Facette, M.; Hamann, T.; Milne, J.; Osborne, E.; Paredez, A.; Persson, S.; Raab, T.; Vorwerk, S.; Youngs, H. *Science* **2004**, *306*, 2206-2211.

  (3) Wertz, J. L.; Bedue, O.; Mercier, J. P. *Cellulose Science and Technology*; EPFL Press: Lausanne, 2010.

  (4) Perlack, R. D. *Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasability of a Billion-Ton Annual Supply*; DOE/GO-102995-2135; Oak Ridge National Laboratory: Oak Ridge, TN, 2005.

  (5) Dale, B. *J. Agr. Food. Chem.* **2008**, *56*, 3885-3891.

  (6) Wyman, C. E. *Mrs Bull* **2008**, *33*, 381-383.

  (7) Fairley, P. *Nature* **2011**, *474*, S2-S5.

  (8) Himmel, M. E.; Ding, S. Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W.; Foust, T. D. *Science* **2007**, *315*, 804-807.

  (9) Yang, B.; Wyman, C. E. *Biofuel. Bioprod. Bior.* **2008**, *2*, 26-40.

  (10) Sanderson, K. *Nature* **2011**, *474*, S12-S14.

  (11) Kajiwara, K.; Miyamoto, T. In *Polysaccharides: Structural Diversity and Functional Versatility*, 2nd ed.; Dumitriu, S., Ed.; Marcel Dekker: New York, 2005; pp 1-40.

  (12) Perez, S.; Mazeau, K. In *Polysaccharides: Structural Diversity and Functional Versatility*, 2nd ed.; Dumitriu, S., Ed.; Marcel Dekker: New York, 2005; pp 41-68.

  (13) Gross, A. S.; Chu, J. W. *J. Phys. Chem. B* **2010**, *114*, 13333-13341.

  (14) Gross, A. S.; Bell, A. T.; Chu, J. W. *J Phys Chem B* **2011**, *115*, 13433-13440.

  (15) Bergenstrahle, M.; Wohlert, J.; Himmel, M. E.; Brady, J. W. *Carbohyd. Res.* **2010**, *345*, 2060-2066.

  (16) Payne, C. M.; Himmel, M. E.; Crowley, M. F.; Beckham, G. T. *J Phys Chem Lett* **2011**, *2*, 1546-1550.

  (17) Peri, S.; Karim, M. N.; Khare, R. *Carbohyd Res* **2011**, *346*, 867-871.

  (18) Cho, H. M.; Gross, A. S.; Chu, J.-W. *J Am Chem Soc* **2011**, *133*, 14033-14041.

  (19) Nishiyama, Y.; Langan, P.; Chanzy, H. *J. Am. Chem. Soc.* **2002**, *124*, 9074-9082.

  (20) Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P. *J Am Chem Soc* **2003**, *125*, 14300-14306.

  (21) Cousins, S. K.; Brown, R. M. *Polymer* **1995**, *36*, 3885-3888.

  (22) Swatloski, R. P.; Spear, S. K.; Holbrey, J. D.; Rogers, R. D. *J. Am. Chem. Soc.* **2002**, *124*, 4974-4975.

  (23) Moulthrop, J. S.; Swatloski, R. P.; Moyna, G.; Rogers, R. D. *Chem. Commun.* **2005**, 1557-1559.

(24)   Singh, S.; Simmons, B. A.; Vogel, K. P. *Biotechnol. Bioeng.* **2009**, *104*, 68-75.

(25)   Pinkert, A.; Marsh, K. N.; Pang, S. S.; Staiger, M. P. *Chem. Rev.* **2009**, *109*, 6712-6728.

(26)   El Seoud, O. A.; Koschella, A.; Fidale, L. C.; Dorn, S.; Heinze, T. *Biomacromolecules* **2007**, *8*, 2629-2647.

(27)   Remsing, R. C.; Swatloski, R. P.; Rogers, R. D.; Moyna, G. *Chem. Commun.* **2006**, 1271-1273.

(28)   Remsing, R. C.; Hernandez, G.; Swatloski, R. P.; Massefski, W. W.; Rogers, R. D.; Moyna, G. *J. Phys. Chem. B* **2008**, *112*, 11071-11078.

(29)   Youngs, T. G. A.; Holbrey, J. D.; Deetlefs, M.; Nieuwenhuyzen, M.; Gomes, M. F. C.; Hardacre, C. *Chemphyschem* **2006**, *7*, 2279-2281.

(30)   Youngs, T. G. A.; Hardacre, C.; Holbrey, J. D. *J. Phys. Chem. B* **2007**, *111*, 13765-13774.

(31)   Liu, H. B.; Sale, K. L.; Holmes, B. M.; Simmons, B. A.; Singh, S. *J. Phys. Chem. B* **2010**, *114*, 4293-4301.

(32)   Janesko, B. G. *Phys Chem Chem Phys* **2011**, *13*, 11393-11401.

(33)   Guo, J. X.; Zhang, D. J.; Duan, C. G.; Liu, C. B. *Carbohyd Res* **2010**, *345*, 2201-2205.

(34)   Guo, J. X.; Zhang, D. J.; Liu, C. B. *J Theor Comput Chem* **2010**, *9*, 611-624.

(35)   Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-935.

(36)   Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781-1802.

(37)   Martyna, G. J.; Tobias, D. J.; Klein, M. L. *J. Chem. Phys.* **1994**, *101*, 4177-4189.

(38)   Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613-4621.

(39)   Lin, S. T.; Blanco, M.; Goddard, W. A. *J Chem Phys* **2003**, *119*, 11792-11805.

(40)   Lin, S. T.; Maiti, P. K.; Goddard, W. A. *J Phys Chem B* **2010**, *114*, 8191-8198.

(41)   Pascal, T. A.; Abrol, R.; Mittal, R.; Wang, Y.; Prasadarao, N. V.; Goddard, W. A. *J Biol Chem* **2010**, *285*, 37753-37761.

(42)   Pascal, T. A.; Lin, S. T.; Goddard, W. A. *Phys Chem Chem Phys* **2011**, *13*, 169-181.

(43)   Huang, S. N.; Pascal, T. A.; Goddard, W. A.; Maiti, P. K.; Lin, S. T. *J Chem Theory Comput* **2011**, *7*, 1893-1901.

(44)   Chandler, D. *Nature* **2005**, *437*, 640-647.

(45)   McQuarrie, D. A.; Simon, J. D. *Physical chemistry : a molecular approach*; University Science Books: Sausalito, Calif., 1997.

(46)   Jasra, R. V.; Ahluwalia, J. C. *J. Solution. Chem.* **1982**, *11*, 325-338.

(47)   Schwarz, F. P. *J. Solution. Chem.* **1996**, *25*, 471-484.

(48)   Berens, P. H.; Mackay, D. H. J.; White, G. M.; Wilson, K. R. *J Chem Phys* **1983**, *79*, 2375-2389.

Chapter 5

Preferential Interactions between Lithium Chloride and Glucan Chains in N,N-Dimethylacetamide Drive Cellulose Dissolution

Abstract

Cellulose, the most abundant component of lignocellulose, is composed of linear glucan chains held together by a strong, robust interaction network that renders the material insoluble in most solvents. One of the few cellulose solvents to have been discovered is lithium chloride (LiCl) dissolved in N,N-dimethylacetamide (DMA). By the integrated application of all-atom molecular dynamics (MD) simulations, reaction path optimization, free-energy calculations, and a coarse-graining (CG) force-matching (FM) analysis, we establish that the DMA mediated preferential interactions between LiCl and the glucan chains of cellulose causes cellulose dissolution in the LiCl/DMA solvent system. The key property of DMA that causes this behavior is that it is a poor solvent in regards to both dissolving LiCl and cellulose. Because of their small size, Li^+ cations can exploit these solvent-mediated preferential interactions to strongly couple to multiple interaction sites simultaneously on glucan chains, including the spatially restricted regions around the ether linkages connecting neighboring glucose rings. Cations in the LiCl/DMA system were thus identified as the main component responsible for dissolving cellulose. The mechanism of cellulose dissolution in LiCl/DMA deduced from the atomistic-scale simulations conducted in this work is also consistent with most of the empirical observations of cellulose solubility in salt/amide solvent systems.

5.1 Introduction

The abundance of lignocellulosic biomass makes it an attractive feedstock for the production of renewable fuels and chemicals.[1,2] Cellulose, the main component of lignocellulosic biomass, consists of linear polymers of glucose arranged into slender aggregates called microfibrils.[3,4] Found in the plant cell wall, microfibrils contain a crystalline interaction network consisting of intrachain and interchain OH—O hydrogen bonds (HBs), as well as unconventional CH—O HBs and van der Waals (vdW) interactions between glucan chains of different sheets (intersheet interactions).[5,6] This network renders cellulose insoluble in most solvents, and subsequently harsh conditions are required to deconstruct the material.[7-11]

Nevertheless, a handful of solvent systems that are able to dissolve cellulose have been reported, such as ionic liquids (ILs), mixtures of N-methyl-N-morpholine-N-oxide (NMMO) and water, and aqueous transition metal complexes.[12,13] Elucidating how molecular coupling in these solvent systems disrupts the robust interaction network of cellulose can reveal the structure-function relationships of solvent-cellulose interactions and contribute to developing pretreatment technologies for converting biomass to fuels. A well-known cellulose solvent system is lithium chloride (LiCl) in N,N-dimethylacetamide (DMA), or its cyclic form, 1-methyl-2-pyrrolidinone (NMP).[14,15] This non-derivatizing solvent[16] has been shown to dissolve cellulose up to 10-15 wt %, with solubility

proportional to LiCl concentration (1-10 wt%).[14,17] Substitution of DMA or LiCl to other analogs may also result in a cellulose solvent, but the resulting solubility tends to be significantly lower.[18-22]

In the LiCl/DMA solvent system without any solute, cations appear to tightly associate with the carbonyl oxygen of DMA according to measurements using ^{13}C NMR,[16] infrared spectroscopy,[23] X-ray crystallography,[24] and thermochemistry.[23-25] Quantum mechanical (QM) calculations of LiCl-DMA clusters also reveal strong DMA-Li^+ interactions.[26] Conversely, Cl^- anions do not appear to strongly couple to DMA molecules. QM calculations indicate that the anion primarily interacts with the methyl hydrogens of DMA.[26]

In solutions of cellulose dissolved in LiCl/DMA, anions were found to replace the OH—O HBs between glucan chains with OH—Cl^- HBs.[14,19] The role of Cl^- in disrupting chain-chain interactions is similar to that of IL anions in cellulose dissolving ILs.[27-31] ^{7}Li NMR showed that the chemical shift of cations is independent of LiCl concentration in DMA over a wide range, most likely because the DMA-Li^+ complex persists in these solutions. Conversely, ^{7}Li NMR chemical shift was observed to be a strong function of cellulose concentration.[32] Therefore, both the DMA-Li^+ complex and Cl^- anions appear to favor coupling to cellulose. The emergent picture of how the LiCl/DMA system dissolves cellulose is thus direct interactions between the dissociated salt and the polymer solute.[19,20] The specific roles and relative strengths of Li^+ and Cl^- in coupling to glucose residues, though, have yet to be resolved. Since cellulose dissolution does not occur after adding LiCl to water or most other solvents, DMA-mediated preferential interactions are likely the driving force. That is, the effective LiCl-cellulose coupling is stronger in DMA due to the specific balance of DMA-LiCl and DMA-cellulose interactions. The preferential interaction theory is widely accepted to rationalize how excipients such as urea and polyols destabilize or stabilize protein folding, respectively,[33,34] but has not yet been established for cellulose dissolution in salt-organic solvent mixtures. Solvent-mediated ion-solute interactions may also explain the observed capability of the LiCl/DMA system in dissolving polyamides.[35]

In this work, we explicitly show the preferential interactions between LiCl and cellulose in DMA through the use of molecular simulations. All-atom molecular dynamics (MD) and free energy simulations of cellulose-dissolving processes are conducted in four solvent systems: pure DMA, LiCl/DMA, pure water, and LiCl/water. Besides computing ion densities around cellulose in different states of dissolution, the results of atomistic simulations are subject to a coarse-graining (CG) force-matching (FM) analysis to quantify the strengths of the preferential interactions. These methods also allow other unanswered questions related to cellulose dissolution in the LiCl/DMA system, such as how the specific coupling between glucose and solvent moieties leads to disruption of the cellulose interaction network, to be addressed.

5.2 Methods

5.2.1 Reaction-Path Optimization and Potential of Mean Force of Cellulose Deconstruction Calculations

The potential of mean force (PMF) of peeling a single glucan chain from the surface of a 12-chain, 10 glucose residue per chain, partial microfibril (Figure 5.1A-B) into solution was calculated in DMA, LiCl/DMA, water, and LiCl/water. The arrangement of glucan chains in the model represents the top three layers of a typically sized cellulose microfibril.[36] In the initial state of the deconstruction pathway, the partial microfibril is intact in the crystalline I_α configuration.[37] In the final state, six glucoses of the top left corner chain are peeled off the microfibril surface from the reducing end (Figure 5.1A-B). To replicate the rigidity of an intact, complete, multi-layered microfibril, harmonic restraints with 5 kcal/mol/Å^2 force constants were placed on the crystalline positions of all heavy ring atoms of sugar residues in the bottommost layer. To effectively represent the long lengths of cellulose microfibrils,[4] additional restraints of the same strength were put on the terminal carbons (C1 or C4 depending on the end) of the glucan chains, except the free end of the peeled chain. The energy-minimized deconstruction pathway was used as the initial configurations of cellulose in the PMF calculations, and was setup and optimized using the same protocols as reported in previous works.[38,39] The final pathway used for the PMF calculations contained 32 replicas.

A constant amount of solvent molecules is used in the simulation of all replicas along the path in the PMF calculations. The DMA system contains 1750 solvent molecules. The LiCl/DMA system contains 1750 DMA molecules and 170 LiCl molecules, corresponding to 4.5 wt % of salt, which is within the experimentally observed range of ion concentrations sufficient for dissolving cellulose.[17] The LiCl/water system has 8766 water molecules and 170 LiCl molecules, corresponding to the same molarity of salt as in the LiCl/DMA system. The pure water system contained 6004 water molecules. All simulations were performed with the NAMD[40] software. Periodic boundary conditions were employed, and the system temperature and pressure were held constant at 350 K and 1 atm using a Langevin thermostat and Nose-Hoover Langevin barostat.[41,42] Cellulose, water, LiCl, and DMA were modeled using the CHARMM[43] carbohydrate force field,[44,45] TIP3P[46] force field, CHARMM ion force field,[47,48] and a newly developed force field based on the standardized CHARMM protocols for force field generation (see the supporting information), respectively. For simulations of the LiCl/DMA solvent system, ion-ion interactions were modified to better reproduce experimentally measured thermodynamic behavior of LiCl in DMA (see the SI). The PMFs were calculated along the deconstruction pathway via integration of the mean force given by harmonic restraint potentials (force constant 5 kcal/mol/Å^2) placed on 32 equally spaced values of a collective variable.[38] The collective variable was the sum of two contact numbers. The first, C—O contacts between the peeled chain and the two chains below it in the microfibril, was included because intersheet interactions are the main cause of cellulose insolubility.[38] To prevent the off-path event of the peeled glucan chain collapsing back onto the far right side of the microfibril surface, a second contact number, C—O contacts between the peeled chain and the top right chain furthest from it

in the top layer, was added to the collective variable. In conducting the PMF calculations in each solvent system, all 32 replicas were energy minimized, gradually heated to 350 K, and equilibrated for at least 5 ns before data collection in the production stage. The equilibration period is to ensure that the PMF profile has stabilized. Production runs of at least 15 ns were then used to calculate the PMFs of cellulose deconstruction.

5.2.2 Simulation Model of Cellulose Dissolution

As a complement to the partial microfibril deconstruction free energy calculations described above, we performed MD simulations of four glucan chains, each 10 glucose residues long, solvated in both a fibril state and in a dissociated state (Figure 5.1C). In the fibril state, the four chains are closely packed together in the crystalline cellulose I_α conformation,[37] while in the dissociated state they are separated by ~20 Å. Taken together, they represent the two end states of a dissolution process. To maintain the four chain model in the fibril or dissociated state in the MD simulations, harmonic restraint potentials with a force constant of 5 kcal/mol/Å^2 were placed on the positions of the terminal carbons (C1 or C4) of each chain.[31] Both states were solvated by 1574 DMA molecules, 1574 DMA molecules and 152 LiCl molecules (4.5 wt % salt), or 8852 water molecules and 152 LiCl molecules. The molarity of salt is the same in the LiCl/DMA and LiCl/water systems. Other simulation details are the same as described in the deconstruction section. During the equilibration stage, system volume, number of ions in contact with the cellulose, and average cluster size of the ions were monitored until stable values reached. Afterwards, data from production runs of at least 30 ns was collected.

To compute the effective interactions between solvent, ion, and glucose moieties before and after dissolution, the coarse-graining (CG) scheme shown in Figure 5.2 is adopted. The Li^+ and Cl^- ions are explicit CG sites in this analysis. The interaction potentials between CG sites were computed from the production-stage trajectories of the all-atom MD simulations via the force-matching (FM) method.[49] We also calculated the three-dimensional ion densities around DMA or the dissociated cellulose chains using a computational procedure described in previous work.[31] Other simulation and analysis details can be found in the SI.

5.3 Results and Discussion

The calculated PMF profiles of cellulose deconstruction are shown in Figure 5.3. In LiCl/DMA, the free energy reduction of detaching the glucan chain from the microfibril surface into the solution phase is -1.3 ± 0.04 kcal/mol-glucose, indicating the capability of dissolving cellulose. In pure DMA, pure water, and LiCl/water, monotonic increases in the PMF with progress of deconstruction indicate that these solvents cannot dissolve cellulose. The calculated free energy cost of microfibril deconstruction in DMA is 3.0 ± 0.11 kcal/mol-glucose, significantly higher than that in water, 1.8 ± 0.03 kcal/mol-glucose. The calculated free energy cost in LiCl/water, 1.6 ± 0.05 kcal/mol-glucose, is very close to that in pure water, indicating that the addition of LiCl does not change the insolubility of cellulose in water. Therefore, addition of LiCl qualitatively shifts the behavior of cellulose solubility in DMA but not in water.

Cellulose dissolution was also modeled by simulating a four glucan chain system in both a fibril and a dissociated state (Figure 5.1C). The smaller system size allows for more extensive sampling and reduced statistical noise of the calculated molecular structures and interactions. To inspect the emergent behaviors of preferential interactions, the average three-dimensional ion densities relative to the bulk value around the glucose residues of the dissociated chains were calculated. These results are plotted from two orthogonal perspectives in Figure 5.4 for cellulose in the LiCl/water system. In water, LiCl localization around glucose residues relative to the bulk is not significant. Cl$^-$ anions are closer to sugar residues than the Li$^+$ cations. Local densities of Cl$^-$ are high near the equatorial OH groups of the sugar ring, signifying OH-anion interactions. Noticeable Cl$^-$ densities are also observed around the axial CH groups due to CH—Cl$^-$ HBs. However, cellulose-ion interactions in water are not sufficiently strong to cause cellulose dissolution (Figure 5.3)

The molar solubility of LiCl in DMA is an order of magnitude lower than that in water.[50,51] The distribution of LiCl around DMA observed in all-atom MD simulations of the LiCl/DMA solvent system without cellulose also reveals clear molecular signatures of DMA being a poor solvent for the salt (details reported in the SI). The CG interaction potentials calculated from the results of atomistic simulations indicate that ion-DMA interactions are highly frustrated, with the AMD-Li$^+$ and ME-Cl$^-$ interactions strongly attractive and the AMD-Cl$^-$ and ME-Li$^+$ interactions strongly repulsive. Since the free energy cost of cellulose deconstruction in pure DMA is higher than that in pure water (Figure 5.3), DMA is also a poorer solvent than water in regards to dissolving cellulose. Conversely, water is a good solvent for dissolving LiCl but a poor one for dissolving cellulose, and adding LiCl to it does not affect its inability to dissolve cellulose. Therefore, solvent properties appear to play an essential role in mediating interactions between the added salt and cellulose in achieving dissolution.

In LiCl/DMA, the three-dimensional ion density profiles normalized by bulk values around the dissociated glucan chains are shown in Figure 5.5A-B. The local enhancement of ion concentrations relative to the bulk is much higher in DMA than that in water. In DMA, both cations and anions bind tightly to sugar moieties from both the equatorial and axial directions of the glucose ring. The two ions also strongly couple to each other near glucose residues to ensure local charge neutrality. In water, a good LiCl solvent, enhancement of ion concentration around cellulose is low and strong anion-cation coupling is not observed around the polymer solute (Figure 5.4).

In DMA, ion localization is especially high near the oxygen containing moieties of the glucan chains. Both Li$^+$ and Cl$^-$ densities are high around OH groups, and Li$^+$ density is also high near the O5 oxygen of the glucose ring. In contrast to the ion distributions in water (Figure 5.4), in DMA, the high-density regions of Li$^+$ are closer to the glucose residues than those of the anions. Therefore, cations play a key role in driving cellulose dissolution in the LiCl/DMA system. Although OH-anion interactions have been identified as a molecular interaction that facilitates cellulose dissolution in LiCl/DMA,[19,20] strong cellulose-Li$^+$ interactions have not been recognized. An interesting feature is the accessibility of Li$^+$ cations to the interaction sites of a glucan chain that bridge neighboring glucose residues. An example is shown in Figure 5.5C, wherein a single Li$^+$ interacts with the O6 and O5 of one glucose, and the O4 and O3 of its neighbor. Upon full solvent exposure of the glucan chains in LiCl/DMA, the average

numbers of ions within 3 Å of a glucose residue are 8 (cation) and 16 (anion) more in the dissociated state than in the fibril state (Figure 5.6). In water, the corresponding increases in the number of sugar-contacting ions due to chain dissociation are only 0.25 (cation) and 1 (anion). Thus, upon chain separation, the solvation environment created by DMA causes many more Li^+ and Cl^- ions to associate with the glucan chains than that created by water.

The observed preferential interactions in DMA between LiCl and the glucan chains that drive cellulose deconstruction can be revealed quantitatively via the interaction potentials between CG moieties (Figure 5.2) calculated from the all-atom MD trajectories of the solvated fibril and dissociated states of cellulose (Figure 5.1C). The interaction potentials of Li^+ and Cl^- with the four CG sites of a glucose residue of cellulose are plotted in Figure 5.7A and 5.7B, respectively. The rich features in these interaction potentials are due to the complex liquid-state structures that exist around the glucan chains. Nevertheless, the occurrence of attractive interactions and well-defined first nearest-neighbor minima are clear.

For the Li^+ cation, SC-Li^+ and LO-Li^+ potentials in DMA are strongly attractive, with strength greater than 6 kcal/mol around the minima at a small separation of ~2 Å (Figure 5.7A). The profiles of the fibril and dissociated states are very similar. Therefore, the greater number of ion-sugar contacts in the dissociated state (Figure 5.6) drives cellulose dissolution. The attractive interactions between the glucose OH sites and Li^+ are also significant, but they are slightly weaker than the SC-Li^+ and LO-Li^+ interactions. Also, the RNG-Li^+ potential has a deep local minimum at ~2.5 Å that is more attractive in the dissociated state than that in the fibril state. Comparatively, in water, cellulose-Li^+ interactions are far less attractive (Figure 5.22).

For the interactions between the larger Cl^- anions and glucose CG sites in DMA (Figure 5.7B), the attractive interactions are significantly weaker than those of the glucose-Li^+ interactions. Interactions of the RNG-Cl^- and OH-Cl^- pairs are attractive at short distances in the dissociated state but are repulsive in the fibril state. These results indicate that anions do also contribute to driving cellulose dissolution, but these contributions are less than those from the glucose-Li^+ interactions. The potentials between Cl^- and the other two glucose sites, SC-Cl^- and LO-Cl^-, do not exhibit significant attractive interactions at close distances. Furthermore, the strengths of glucose-Cl^- potentials in DMA are in fact similar to those in water (Figure 5.23) despite the additional features in the PMF profiles in DMA due to the more complex liquid structures.

The results presented above indicate that interactions between Li^+ and glucose residues in DMA indeed appear as the main driver of cellulose dissolution. Although Cl^- anions do interact with the hydroxyl groups of glucose, the DMA-mediated preferential interactions between Li^+ and glucan chains are much stronger. The small size of the Li^+ cation allows for multiple simultaneous attractive interactions with a glucan chain, particular around the linker oxygen and hydroxymethyl side group of the glucose residues.

In regards to solvent-ion coupling, the calculated interaction potentials clearly show that the amide group of DMA strongly interacts with Li^+, but this attractive interaction is counteracted by the highly repulsive ME-Li^+ interactions (Figure 5.16). The result of this is that Li^+ anions are highly frustrated in the bulk and favor interactions with glucose residues instead. In DMA, Cl^- anions interact attractively with solvent methyl

groups but repulsively with the solvent amide group and so also exhibit strong preferential interactions with cellulose moieties. However, Cl$^-$ anions cannot access as many interaction sites on the glucan chains as the Li$^+$ cations because of their larger size. Our calculations using all-atom models thus identify that Li$^+$ cations, via solvent-mediated preferential interactions, are most responsible for causing cellulose dissolution in DMA. This theory is consistent with the empirical observation that NaCl or salts with larger cations in DMA are not able to dissolve cellulose,[14] but LiBr/DMA still can, although with a lower solubility.[22]

5.4 Conclusions

By a combined use of all-atom MD simulations, reaction path optimization, free energy calculations, and a coarse-graining force-matching analysis, we found that preferential interactions between LiCl and glucan chains cause cellulose dissolution in DMA. DMA-mediated LiCl-glucose interactions lead to a free energy reduction upon peeling of a glucan chain from a microfibril surface into solution, a molecular process that involves disrupting the intrinsic interaction network of cellulose. Since DMA is a poor solvent in regards to both dissolving LiCl and dissolving cellulose, the two species strongly couple to each other in the solvent system. Such solvent-mediated interactions do not occur in water, a good solvent for dissolving LiCl but a poor one for dissolving cellulose. The small size of the Li$^+$ ions gives it extensive access to many interaction sites along glucan chains, including the more restricted regions near the oxygen linkers connecting neighboring glucose residues. The calculated strengths of the cellulose-Li$^+$ interactions are much stronger than those between cellulose and the larger Cl$^-$ anions.

A number of experimental observations are consistent with the aforementioned mechanism deduced from the atomistic simulations described in this work. First, cellulose solubility is an increasing function of LiCl concentration in DMA,[17] since more ions are available to interact with the glucan chains to disrupt the cellulose interaction network at higher salt loadings. Second, substitution of DMA to other more polar amides that better solvate LiCl, such as dimethylformamide, results in a weaker ability or total inability to dissolve cellulose.[14,18,21] Based on the mechanism discussed earlier, the key property of DMA in mediating the preferential interactions that can dissolve glucan chains in solution is that it is a poor solvent for both LiCl and cellulose. That is, although DMA only dissolves a limited amount of LiCl since it is a weak solvent for the salt, the combined effects of weak ion-solvent and cellulose-solvent coupling create sufficiently strong preferential interactions between the salt and glucose residues to achieve dissolution. Furthermore, replacing Li$^+$ with larger cations, including sodium, potassium, calcium, and barium all result in ineffective cellulose solvents.[14] As discussed earlier in our theory, the small size of Li$^+$ is important for accessing the spatially restricted interaction sites around glucan chains and forming O-Li$^+$ interactions. Replacing Cl$^-$ with Br$^-$ in DMA while keeping Li$^+$ as the cation gives a weaker, but still effective cellulose solvent.[22] This is consistent with our findings that the cations in the LiCl/DMA system are the dominant contributors to cellulose dissolution. However, since replacing Cl$^-$ in LiCl/DMA with a larger anion does lower cellulose solubility,[14,19,22] anions do contribute to the disruption of the cellulose interaction network. The theory developed in this work suggests that LiF/DMA could be a more effect cellulose solvent than LiCl/DMA given

equivalent amounts of dissolved salt because F$^-$ is a smaller and stronger HB accepter than Cl$^-$.[52]

Integration of the simulation results obtained in this work with the previous experimental observations indicates that the DMA-mediated preferential interactions between salt and sugar require a delicate balance between the interactions of the different components of the ion/sugar/solvent system to give rise to the capability of dissolving cellulose. DMA being a poor solvent for both the salt and the substrate and a small cation size are identified as two key molecular features for this mechanism.

5.5 Acknowledgements

5.6 Supporting Information

5.6.1 Modification of the Li-Cl Interaction Potential in DMA

The ion-ion interaction potentials in DMA were modified slightly due to inaccurate phase behavior in our all-atom molecular dynamics simulations using the unmodified ion force field. The CHARMM small ion force field was developed to accurately model the behavior of ions in water, and was validated using such metrics as free energy of solvation.[47,48] However, because the properties of DMA differ from those of water, particularly that the dielectric constant of DMA is approximately half that of water, the ion-ion interactions are too strong for their simulation in DMA. Therefore, we modified the ion-ion potentials slightly. In Figure 5.8, we show both the original LiCl ion-ion interactions, and the new modified ones we used in our all-atom simulations. The potentials were modified so that from the distance of the interaction minimum of the Li$^+$-Cl$^-$ interaction inwards, the strength of the interaction was as if the ions had |0.9e| charges instead of the original |1.0e| charges (the signs of the charges were not modified). This is equivalent to scaling the dielectric constant of the ion-ion Coulombic interactions by (1/0.9^2) for the same distances. From the distance of the Li-Cl interaction minimum out to 12 Å, this perturbation was continuously switched off, and after 12 Å, the ion-ion potentials were unchanged. Practically, this was accomplished through modification of the ion-ion van der Waals interactions, while the Coulombic interactions were not changed. Also, the ion-sugar and ion-solvent interactions were not adjusted.

We determined the 0.9 charge scaling for the ion-ion interactions by examination of the ion-ion clustering in DMA. In Figure 5.9 we show the trajectories of the average cluster size (in total number of ions) for the unmodified force field and for different modifications of the ion-ion interactions. We also show these trajectories for simple scaling of the ion charges as well. The difference between the simple scaling and the modifications described above is that for the former, all interactions that involve an ion are affected, while for the latter, only ion-ion interactions are affected. All simulations

120

were run using the NAMD software[40] with periodic boundary conditions at 350 K and 1 atm using a Langevin thermostat and Nose-Hoover Langevin barostat.[41,42] The system composition was 921 DMA molecules and 79 LiCl molecules (4.0 wt % LiCl in DMA).

Over the 10 ns trajectory, the amount of clustering for the unmodified force field did not deviate from its initial value. The same is true of the modified potential with |0.95e| effective charges. For both systems, ion clustering is irreversible. That is, ion clusters will never break apart over time, only increase in size, indicating the insolubility of the LiCl in DMA with these force fields. Once the modification is increased to |0.9e| effective charges, the average cluster size decreases over time. At higher levels of modification, |0.85e| and lower, the amount of clustering is much lower, near an average cluster size of one (that is, no clustering at all). This amount is low, lower than the amount in water at the same temperature and ion molarity (Figure 5.10), even though the dielectric constant of water is twice that of DMA. Also, visual inspection of the trajectories showed that the low cluster size was in part due to unphysical formation of loose aggregates of like ions (data not shown). This did not occur with the |0.9e| and |0.95e| modified force fields. Thus, the |0.9e| effective charge force field gave the best results, lying in between the irreversible clustering of the lesser perturbations and unphysical aggregation of the higher perturbations. For the cluster size trajectories from the simulations ran with the simple scaling of ion charges, only the |0.6e| trajectory showed reversible clustering. Lesser scalings, |0.9e| and |0.8e|, did not. However, use of the |0.6e| simple scaling in a cellulose deconstruction PMF calculation in LiCl/DMA did not give an effective cellulose solvent, and so, does not accurately model LiCl/DMA in this respect. The cellulose deconstruction PMF with this force field was positive (data not shown).

For comparison, we performed the same simulations of LiCl in water at the same conditions, 350 K and 1 atm, the results of which we show in Figure 5.10. These simulations contained 4875 TIP3P[46] water molecules and 79 LiCl molecules. The molarity of ions in the system is the same as that in the LiCl/DMA simulations. For the unmodified force field the average cluster size was slightly under 2, while for all the modified ones, it was one. Like in the case of DMA, this was because of the formation of unphysical loose aggregates of like ions. Thus, in water, the solvent for which the LiCl force field was created, no modification of the force field should be used, as expected.

As another check, we examined the solid density, melting point, and heat of fusion of pure LiCl for the original force field, as well as for the modified ones. Simulations containing 4000 LiCl molecules were run at constant temperature and pressure (1 atm) in the same manner as those above. At a given temperature, simulations were started from both a crystalline, solid state, and an amorphous, liquid state. The simulated melting point was taken as when the simulations that began in the liquid state underwent a discontinuous density increase with decreasing temperature. Below this point, the density from the simulations that began in the solid state were used for the equilibrium density, while above it, the densities from the liquid initialized simulations were used. The heat of fusion at each temperature was calculated as the difference in enthalpy between the liquid and solid initialized simulations. We show these results in Figures 5.11-12. All force fields accurately predicted the melting point of LiCl (878 K, shown by the vertical line in Figure 5.11), but the modified |0.8e| simulations deviated significantly from the experimental liquid state density while the others did not. Also, for

those force fields that the enthalpy of fusion was calculated, the |0.9e| effective charge one gave results closest to the experimental value of 4.8 kcal/mol. Thus overall, from examining the behavior of LiCl in DMA and also in its neat state, the |0.9e| ion-ion effective charge modification gave the most accurate results when compared to experimentally known behavior.

5.6.2 Construction of the DMA Force Field

As no DMA CHARMM force field existed, we had to create one. The force field was based off the existing force fields for acetamide, N-methylacetamide, and the methyl-substituted ammonias, from which essentially all the atom types, bonds, angles, and dihedrals in DMA already existed. The force field was validated by calculating the liquid phase density and heat of vaporization of DMA between 275 K and 400 K and 1 atm. Simulations were run at constant pressure and temperature in the same manner as described previously. The simulated system contained 1000 DMA molecules. The results of our density calculations are shown in Figure 5.13. The force field accurately reproduces the experimental DMA density across a wide range of temperatures. The deviations of the calculated values from the experimental ones are all less than 1 %. In Figure 5.14, we show the calculated heats of vaporization of DMA. Like in the case of density, the calculated values are close to the experimental ones, and the deviation between the two ranges between 2 and 6.5 %. The parameters for the force field are given in Tables 5.1 and 5.2.

5.6.3 Calculation of the Three-Dimensional Density Profiles

The three-dimensional density profiles were calculated in the same manner as those from previous work.[31] For the central molecule around which the density is to be calculated, a standardized local coordinate system and origin is determined. For DMA, the origin is the midpoint of the C-N amide bond. The x-axis is along this bond, the y-axis is perpendicular to the amide plane, and the z-axis is the cross product of the two. For glucose, the origin is the center-of-mass of the heavy atoms of the glucose ring. The x-axis is in the direction of the C2-O2 and C3-O3 bonds, the z-axis is the direction of polymerization of the sugar chain, and the y-axis is the cross product of the two. After creating the local coordinate system, the space around the central molecule is divided up into three-dimensional bins, and the occupancy of each bin by the selected solvent molecules calculated. These occupancies are then averaged over all central molecules and time, and then normalized by the bulk density.

5.6.4 DMA is a Poor LiCl Solvent

The driving force for accumulation of LiCl at the cellulose surface and subsequent dissolution of the sugars is the preferential interactions caused by DMA being a poor solvent for LiCl and cellulose. In Table 5.3A-B we list the maximum solubility of LiCl in water and DMA.[50,51] The molar solubility of LiCl per volume is an order of magnitude lower in DMA than in water. Energetically, this is due to water being more polar than DMA. Water is capable of forming electrostatically driven interactions with Li^+ through

O-Li$^+$ contacts and with Cl$^-$ through OH—Cl$^-$ hydrogen bonds. The only polar group that DMA posses is the central amide functionality. While this provides the possibility of energetically favorable ion-solvent interactions, that this functionality is surrounded by apolar methyl groups limits this effect. To understand the predominant LiCl/DMA interactions, in Figure 5.15A-B, we show the average density of LiCl normalized by the bulk value around DMA in neat LiCl/DMA at 350 K and 1 atm. Li$^+$ density is strongly localized near the amide oxygen, with values greater than 100, indicating the formation of O-Li$^+$ interactions. Cl$^-$ is located by the methyl hydrogen groups. The methyl hydrogen to Cl$^-$ interaction is energetically weak compared to a Li$^+$-Cl$^-$ interaction, or to a O-Li$^+$ interaction (either in DMA or in water). This is evident in the lower values of the Cl$^-$ localization, as well as its more diffuse volume in space seen in Figure 5.15B. Also, although it has been suggested in some mechanisms,[20] we see no evidence of C-Cl$^-$ interactions with the amide carbon atom, or of N-Li$^+$ interactions with the amide nitrogen atom.

Besides the two caused by direct ion-DMA interactions just discussed, other regions of ion density localization are evident in Figure 5.15A-B. However, these result not from ion-DMA interactions, but from ion-ion interactions in the vicinity of a DMA. For instance, the volumes of high Cl$^-$ density by the amide oxygen are due to Cl$^-$ coordinating to a Li$^+$ that itself has coordinated to an amide oxygen. The second region of Li$^+$ density in this area is then due to coordination of Li$^+$ ions to those Cl$^-$ ions. Likewise, the high density of Li$^+$ near the methyl groups is due to coordination of Li$^+$ ions to Cl$^-$ ions which themselves are near a methyl group.

The fact that each ion has a distinct and separate part of DMA to which it preferentially interacts with and highly localizes to signals that DMA is a poor solvent of LiCl. The Li$^+$ prefers the amide group, while the Cl$^-$ prefers the methyl group. That there is no cross over between the two interaction motifs implies that Li$^+$ interacting with a methyl group or Cl$^-$ interacting with an amide group would be extremely unfavorable thermodynamically. This is seen in the LiCl/DMA force matching results, which we show in Figure 5.16. While the AMD-Li$^+$ and ME-Cl$^-$ interactions are very favorable, the ME-Li$^+$ and AMD-Cl$^-$ curves are just as unfavorable. Since the total interaction any one ion has with a DMA molecule would be the sum of its interactions with each of the four groups, the energetic gains made by the favorable interaction(s) are counterbalanced by the repulsion(s) of the unfavorable ones. Also, the distinct partitioning of the favorable interactions of the ions with the two DMA functionalities and the high amount of localization of each ion type around these functionalities represent a large entropic penalty compared to a solvent where the ions would not be so highly localized in space. This is in addition to any weaker energies of interaction that LiCl has with DMA compared to water. The combination of the two make DMA a poor solvent for LiCl compared to water.

Table 5.1. CHARMM Topology file for N,N-dimethylacetamide.

```
MASS   101 HA      1.00800   H
MASS   102 CT3    12.01100   C
MASS   103 C      12.01100   C
MASS   104 O      15.99940   O
MASS   105 NH0    14.00700   N


RESI NDMA        0.00 ! N,N-dimethylacetamide
GROUP
ATOM CL   CT3   -0.27
ATOM HL1  HA     0.09
ATOM HL2  HA     0.09
ATOM HL3  HA     0.09
ATOM C    C      0.48
ATOM O    O     -0.48
ATOM N    NH0   -0.32
ATOM CR1  CT3  -0.11
ATOM H11  HA     0.09
ATOM H12  HA     0.09
ATOM H13  HA     0.09
ATOM CR2  CT3  -0.11
ATOM H21  HA     0.09
ATOM H22  HA     0.09
ATOM H23  HA     0.09


BOND CL  HL1  CL  HL2  CL  HL3  CL  C
BOND C  O   C  N   N  CR1  N  CR2
BOND CR1 H11  CR1 H12  CR1 H13  CR2 H21
BOND CR2 H22  CR2 H23
IMPR N  C  CR1 CR2
IMPR C  CL N   O
PATCH FIRS NONE LAST NONE
```

Table 5.2. CHARMM parameter file for N,N-dimethylacetamide.

Bonds
CT3 C   250.000   1.4900
HA   CT3 322.000   1.1110
O   C   620.000   1.2300
NH0 CT3 320.000   1.4300
NH0 C   370.000   1.3450
NH3 HC   403.000   1.0400

Angles
HA   CT3 HA   35.500   108.40   5.40 1.80200
HA   CT3 C   33.000   109.50 30.00 2.16300
O   C   CT3   80.000 121.0000
NH0 C   CT3   80.000 116.5000
O   C   NH0   80.000 122.5000
CT3 NH0 C   50.000 120.0000
NH0 CT3 HA   51.500 109.5000
CT3 NH0 CT3   53.000 120.0000

Dihedrals
X   CT3 C   X   0.0500 6 180.00
CT3 C   NH0 CT3   2.5000 2 180.00
O   C   NH0 CT3   2.5000 2 180.00
X   CT3 NH0 X   0.1100 6   0.00

Improper Dihedrals
O   X   X   C   120.0000   0   0.0000
NH0 X   X   CT3   20.0000   0   0.0000

Nonbonded
CT3   0.000000 -0.080000   2.060000 0.000000 -0.010000   1.900000
C   0.000000 -0.110000   2.000000
O   0.000000 -0.120000   1.700000 0.000000 -0.120000   1.400000
HA   0.000000 -0.022000   1.320000
NH0   0.000000 -0.200000   1.850000 0.000000 -0.200000   1.550000

Table 5.3A. Maximum solubility of LiCl in water.

Temp (K)	10	20	25	30	40
Mass Fraction[a]	0.425	0.453	0.458	0.463	0.473
Molarity (M)	12.790	13.813	13.999	14.161	14.538

[a] Solubility and density data from Ref 50

Table 5.3B. Maximum solubility of LiCl in DMA.

Temp (K)	25
Mass Fraction[a]	0.085
Molarity (M)	2.006

[a] Solubility and density data from Ref 51

(A)

(B)

(C)

Fibril   Dissociated
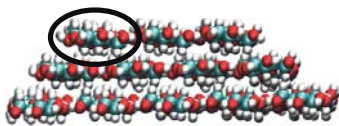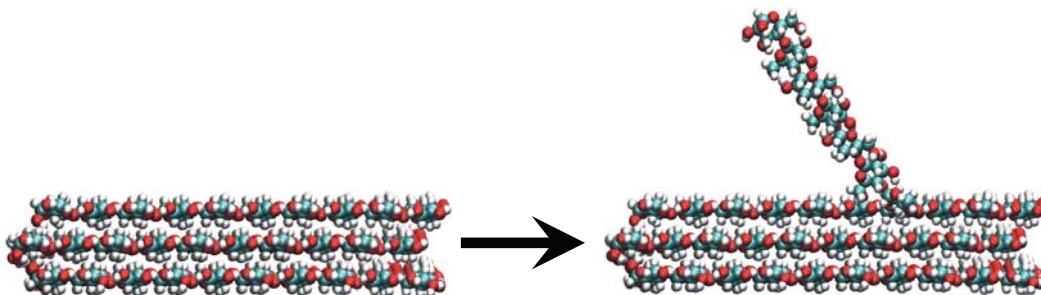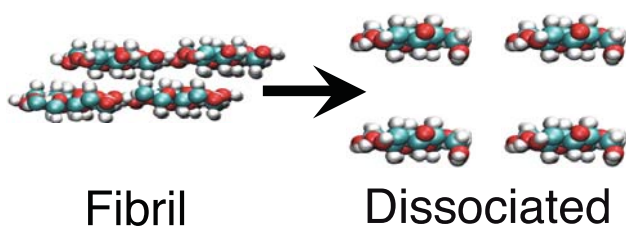
Figure 5.1. Molecular models and processes employed for investigating cellulose dissolution in this work. (A) Cross-section of the partial microfibril subject to deconstruction. The circled chain is peeled off from the microfibril surface in the deconstruction process. (B) Side view of the initial and final states of the partial microfibril in the deconstruction process. (C) Cross section of the fibril and dissociated states of the four chain clusters. The fibril state is in the crystalline I_α conformation. In the dissociated state, the chains are separated by ~20 Å.
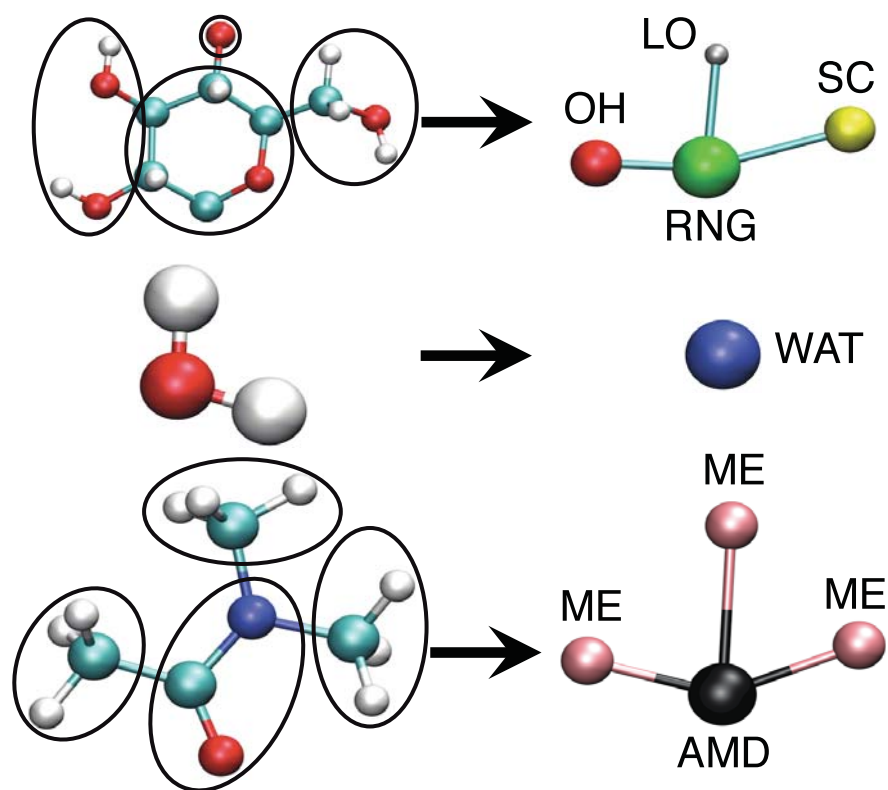
Figure 5.2. The coarse-graining scheme employed for glucose, water, and DMA used in computing the force-matched interaction potentials.
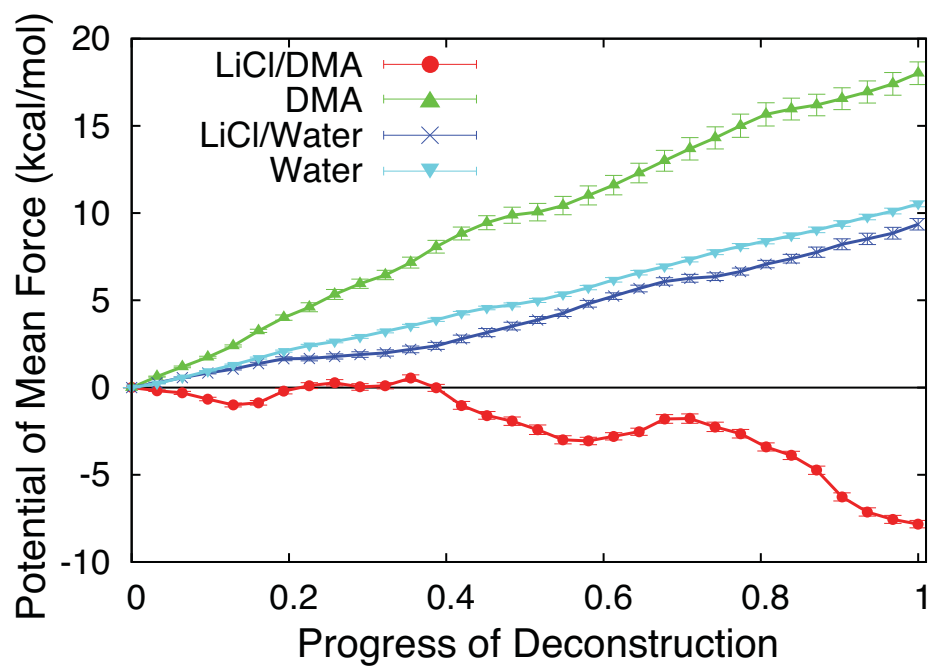
Figure 5.3. The PMFs of cellulose deconstruction (Figure 5.1A-B) in LiCl/DMA, DMA, LiCl/water, and water.
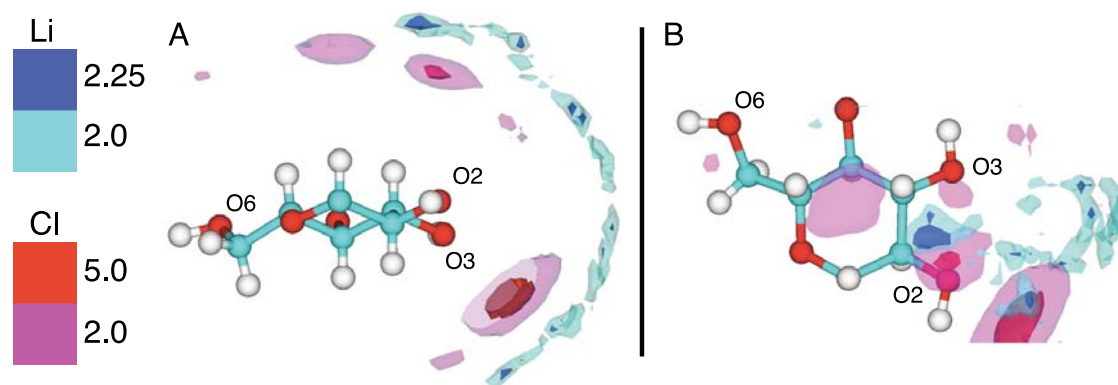
Figure 5.4. The average densities of Li$^+$ and Cl$^-$ around the dissociated glucan chains (Figure 5.1C) in water normalized by the bulk values. (A) Front view and (B) top view.
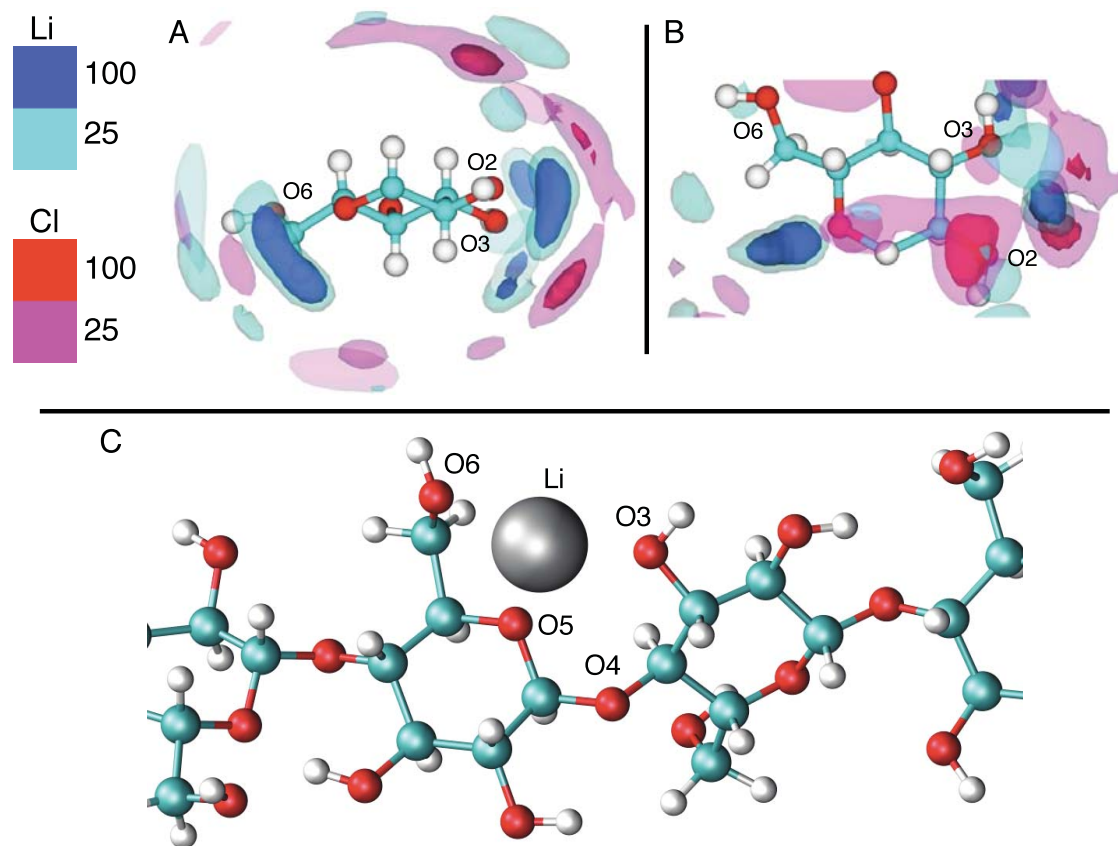
Figure 5.5. The average densities of Li$^+$ and Cl$^-$ around the dissociated glucan chains (Figure 5.1C) in DMA normalized by the bulk values. (A) Front view and (B) top view. (C) A configuration of Li$^+$ binding to the restricted region between neighboring glucose resides in a dissociated chain. This configuration is from a snapshot of the all-atom MD simulation.
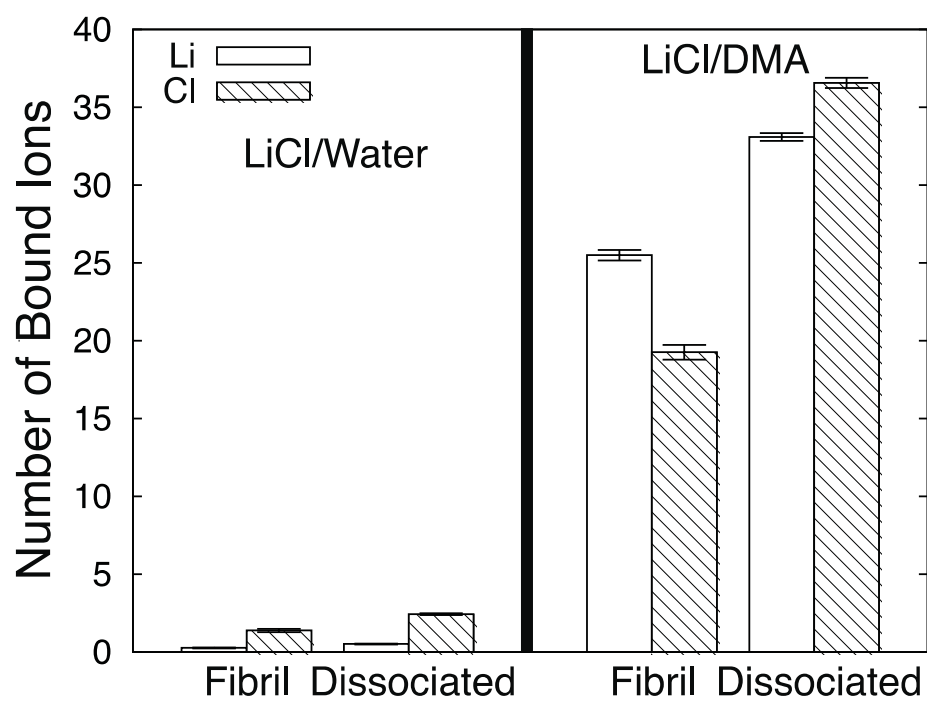
Figure 5.6. The average number of Li$^+$ and Cl$^-$ ions within 3 Å of cellulose atoms calculated from the MD simulations of the fibril and dissociated states of cellulose (Figure 5.1C) in LiCl/water and LiCl/DMA.
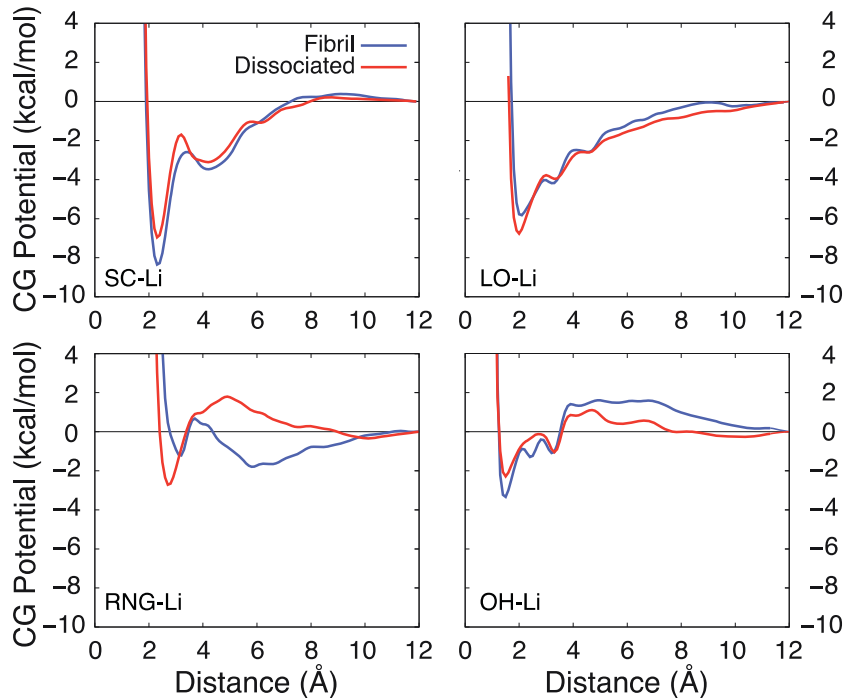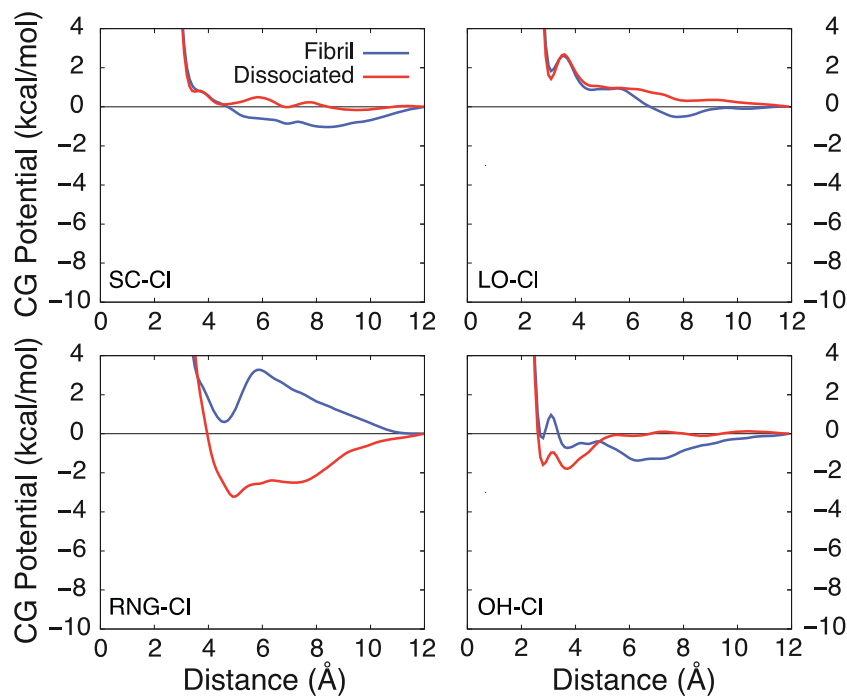
Figure 5.7. Interaction potentials between LiCl and the CG sites of cellulose in LiCl/DMA. (A) Li$^+$ and (B) Cl$^-$. CG sites of cellulose are shown in Figure 5.2.
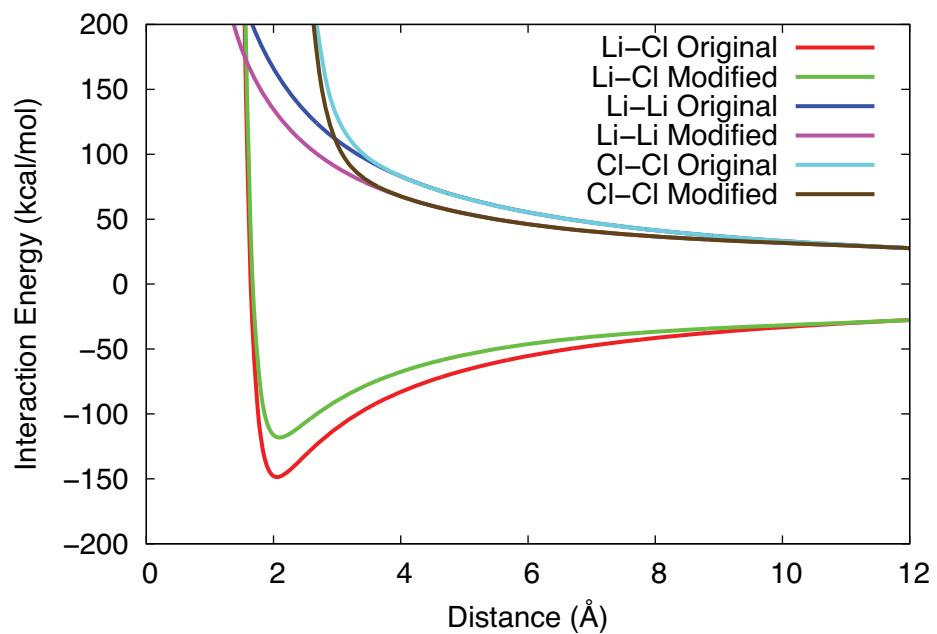
Figure 5.8. Modified ion-ion interaction potentials in DMA used in this work. Original unmodified potentials shown for comparison.
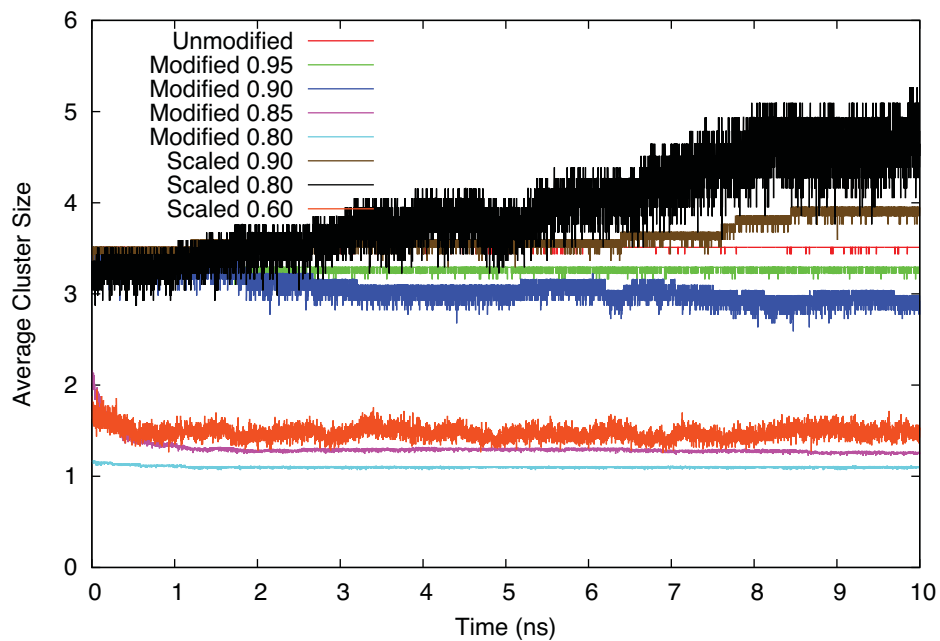
Figure 5.9. Trajectories of the average cluster size of LiCl in DMA for the original unmodified force field, modified ion force fields, and scaled ion force fields.
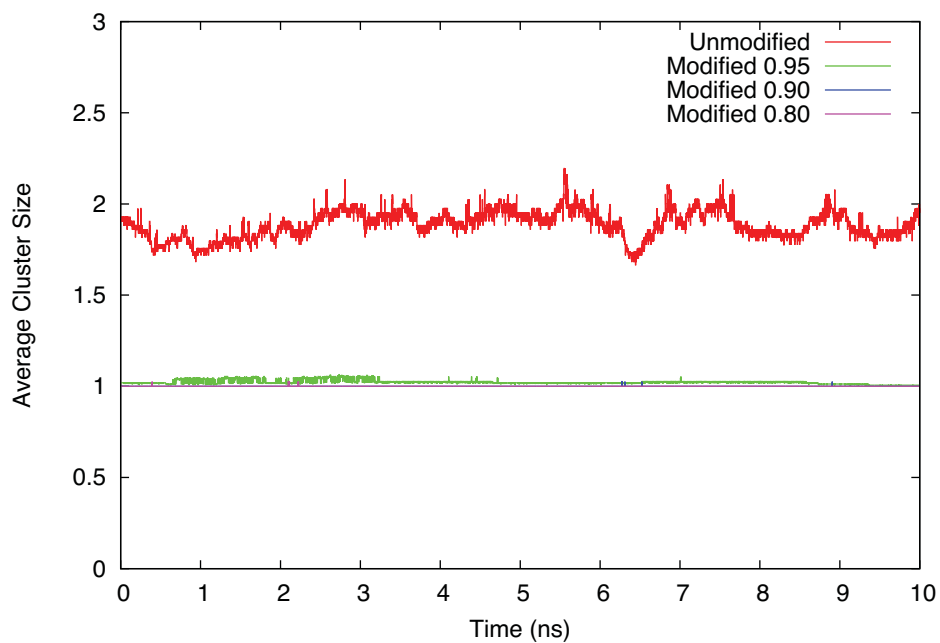
Figure 5.10. Trajectories of the average cluster size of LiCl in water for the original unmodified force field and modified ion force fields.
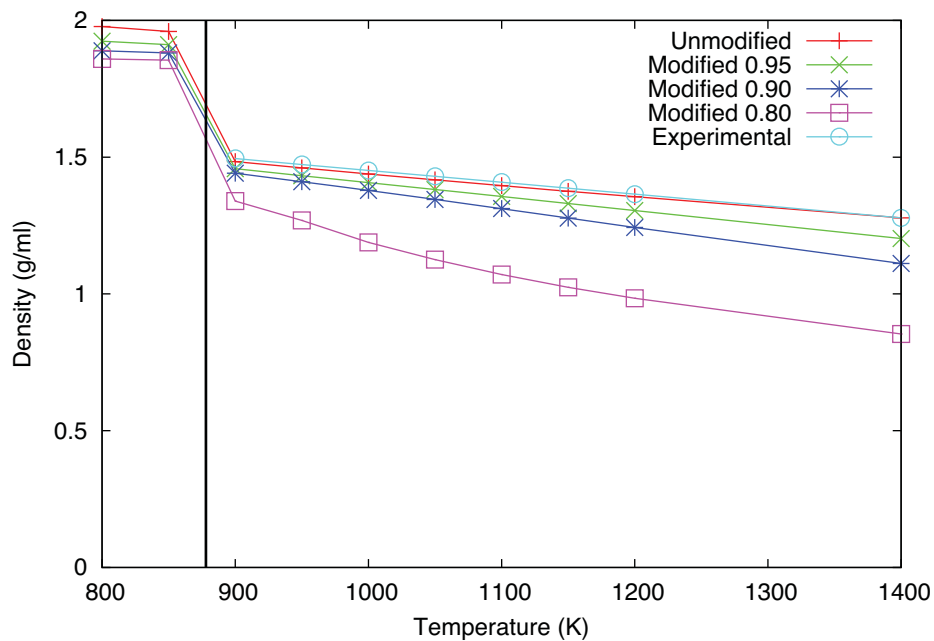
Figures 5.11. Density of neat LiCl at various temperatures for the original unmodified force field, and several modified force fields, and the experimental density for comparison. The vertical line shows the experimental melting point of LiCl at 1 atm (878 K).
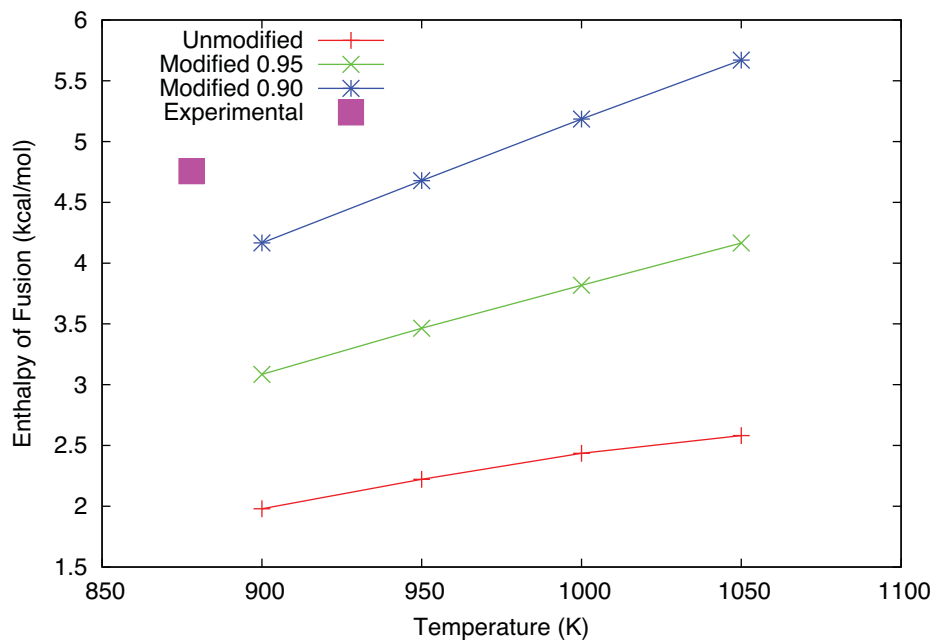
Figure 5.12. Enthalpy of fusion of LiCl for the original unmodified force field and several modified force fields. The experimental enthalpy of fusion at the melting point is also shown.
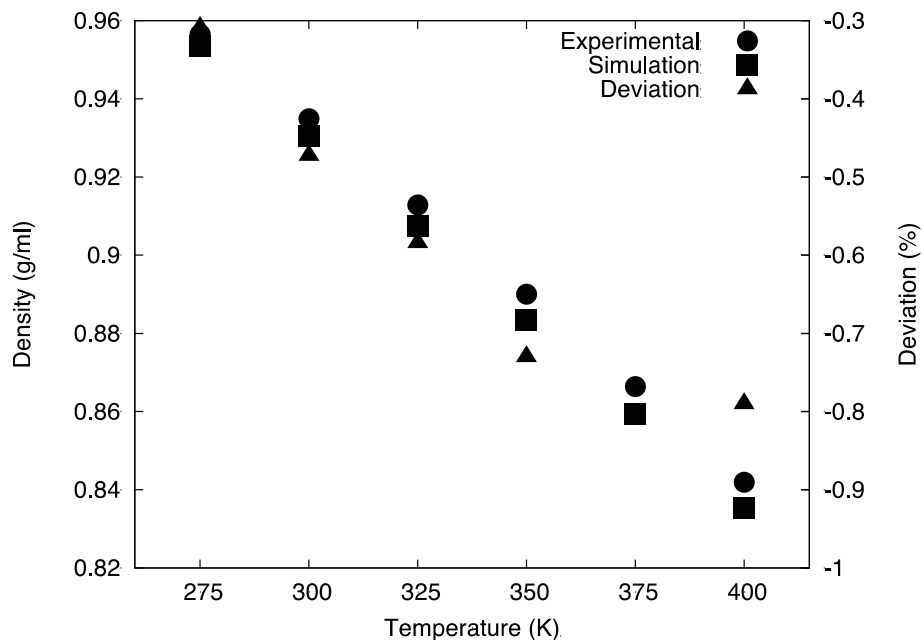
Figure 5.13. Experimental and simulated density of DMA at various temperatures. Percent deviation between the two is also shown.
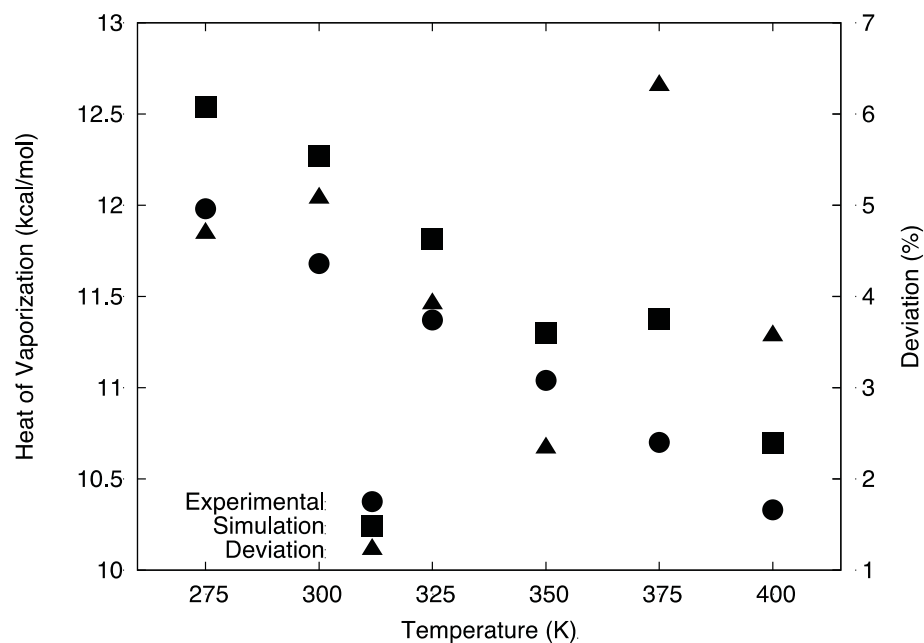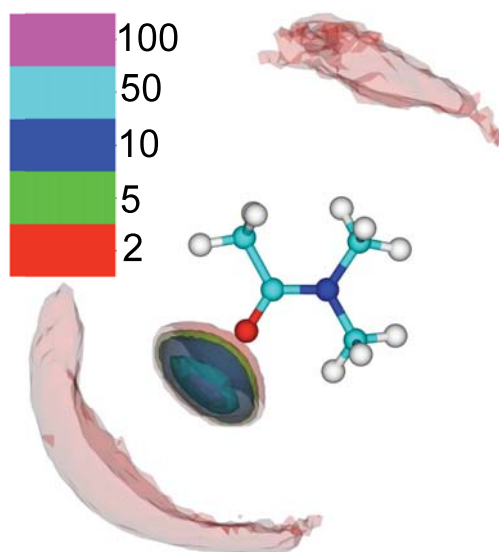

Figure 5.14. Experimental and simulated heat of vaporization of DMA at various temperatures. Percent deviation between the two is also shown.
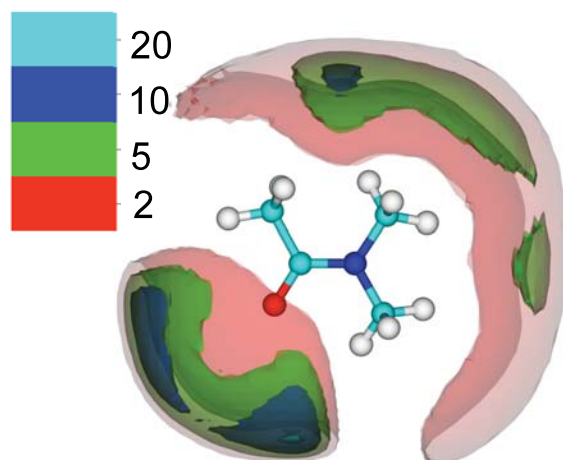
(A)

(B)

Figure 5.15. Three-dimensional density relative to the bulk of Li^+ (A) and Cl^- (B) around DMA in LiCl/DMA at 350 K and 1 atm.
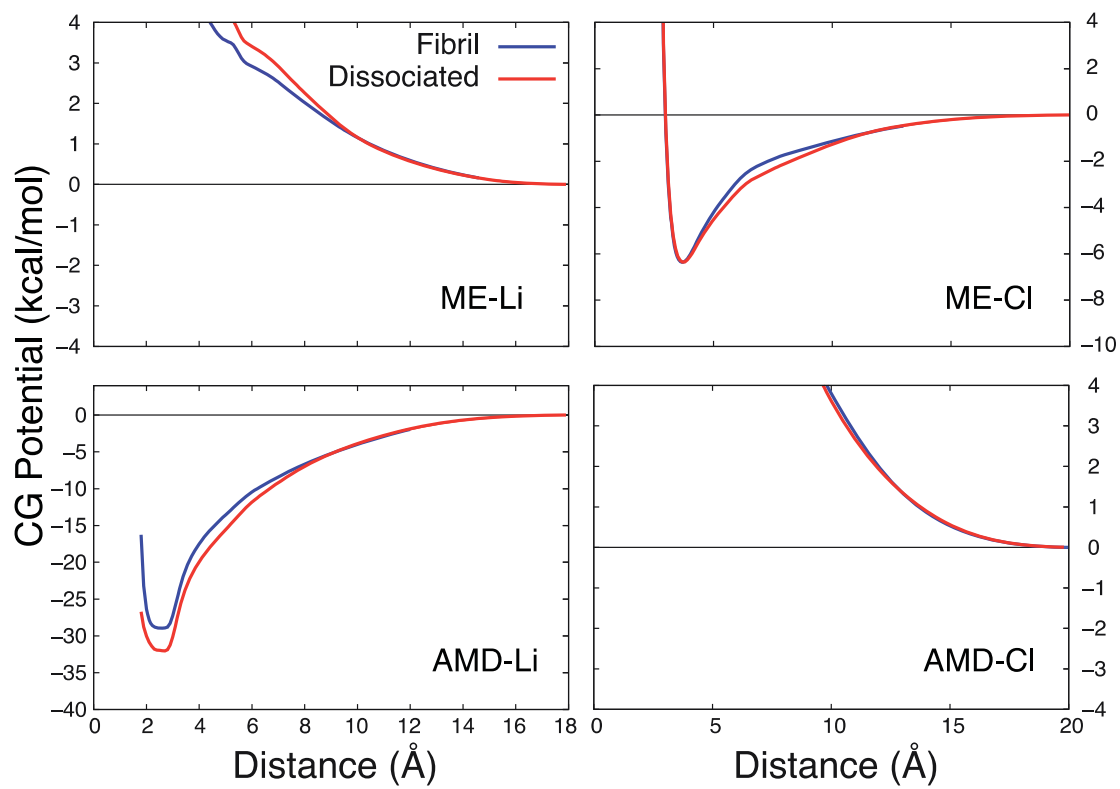
Figure 5.16. CG interaction potentials between LiCl and the coarse-grained functionalities of DMA (defined in Figure 5.2) in LiCl/DMA.
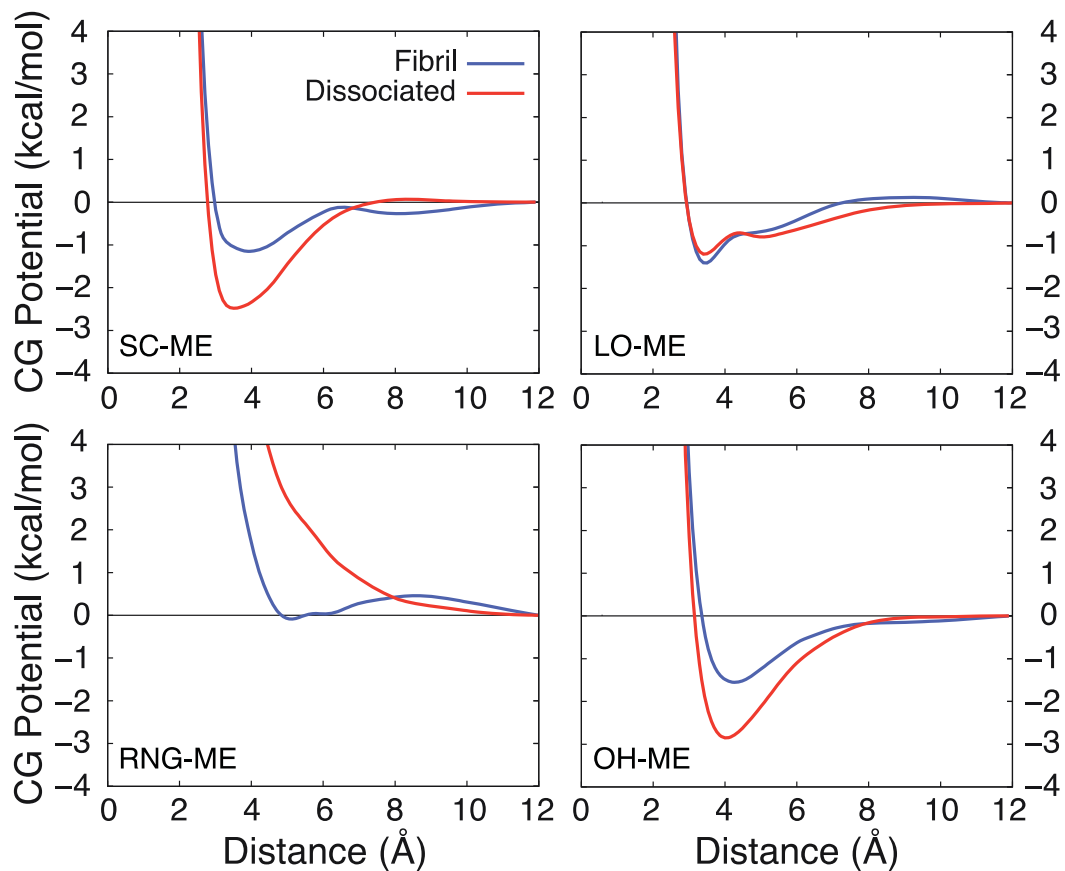
Figure 5.17. CG interaction potentials between the ME coarse-grained group of DMA and the coarse-grained cellulose functionalities in LiCl/DMA. Coarse-grained groups are defined in Figure 5.2.
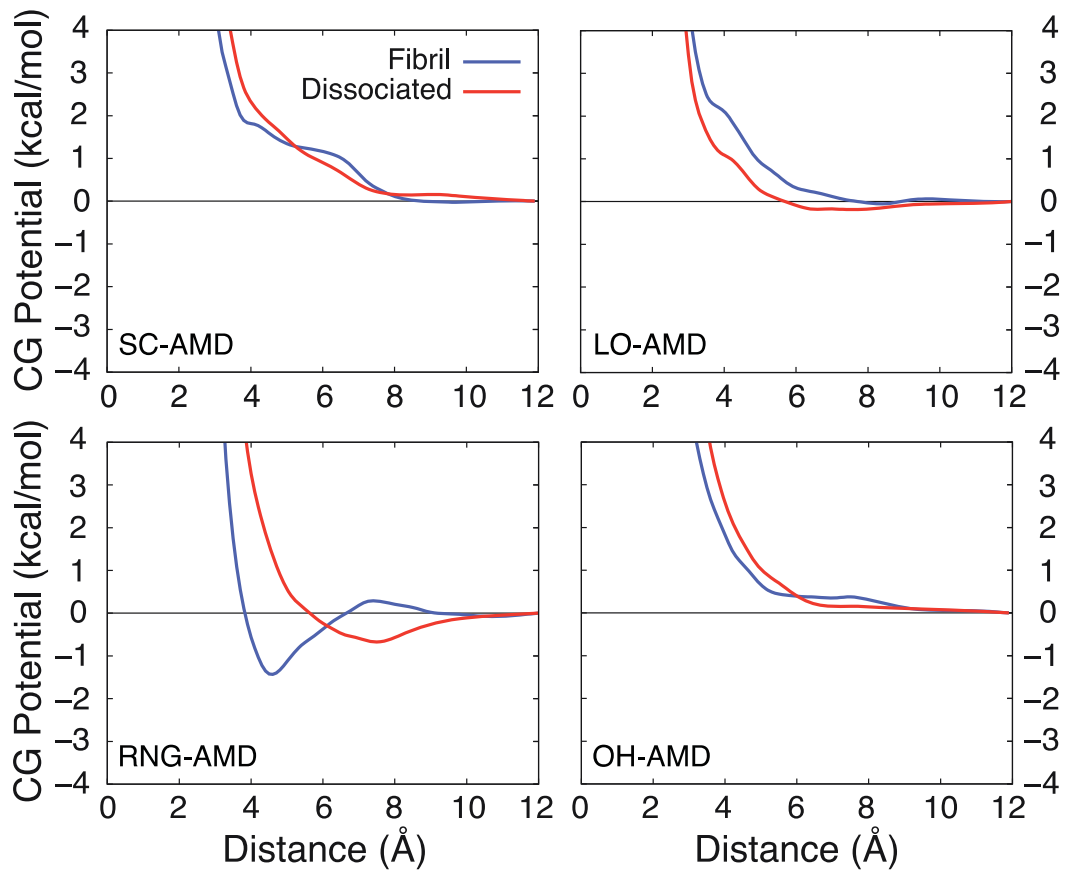
Figure 5.18. CG interaction potentials between the AMD coarse-grained group of DMA and the coarse-grained cellulose functionalities in LiCl/DMA. Coarse-grained groups are defined in Figure 5.2.

Figure 5.19. CG interaction potentials between the ME coarse-grained group of DMA and the coarse-grained cellulose functionalities in pure DMA. Coarse-grained groups are defined in Figure 5.2.

Figure 5.20. CG interaction potentials between the AMD coarse-grained group of DMA and the coarse-grained cellulose functionalities in pure DMA. Coarse-grained groups are defined in Figure 5.2.
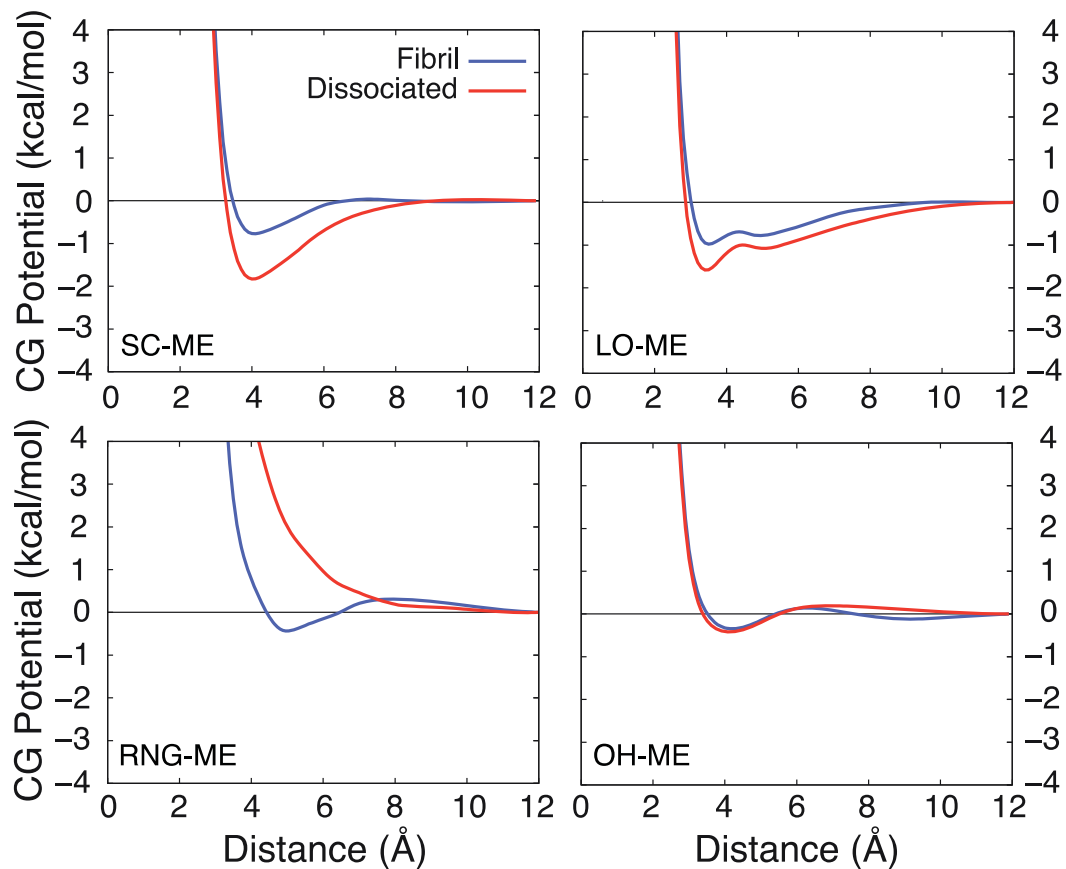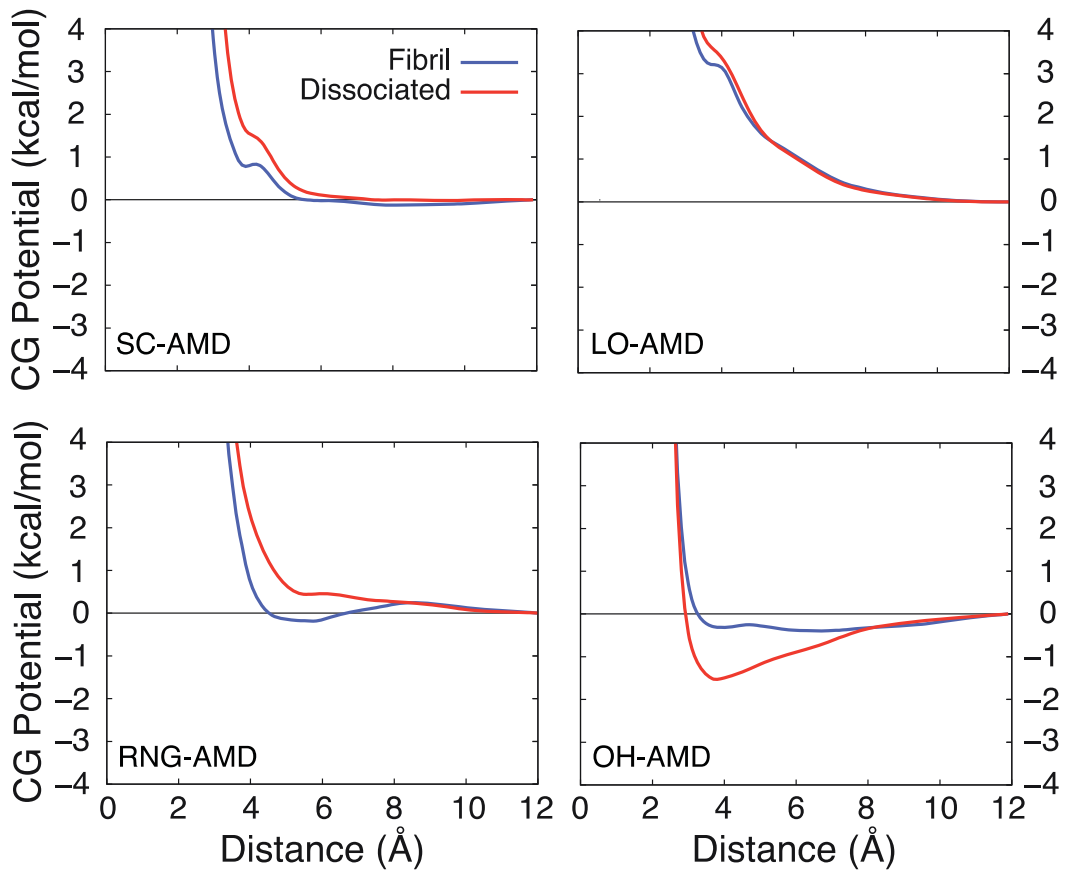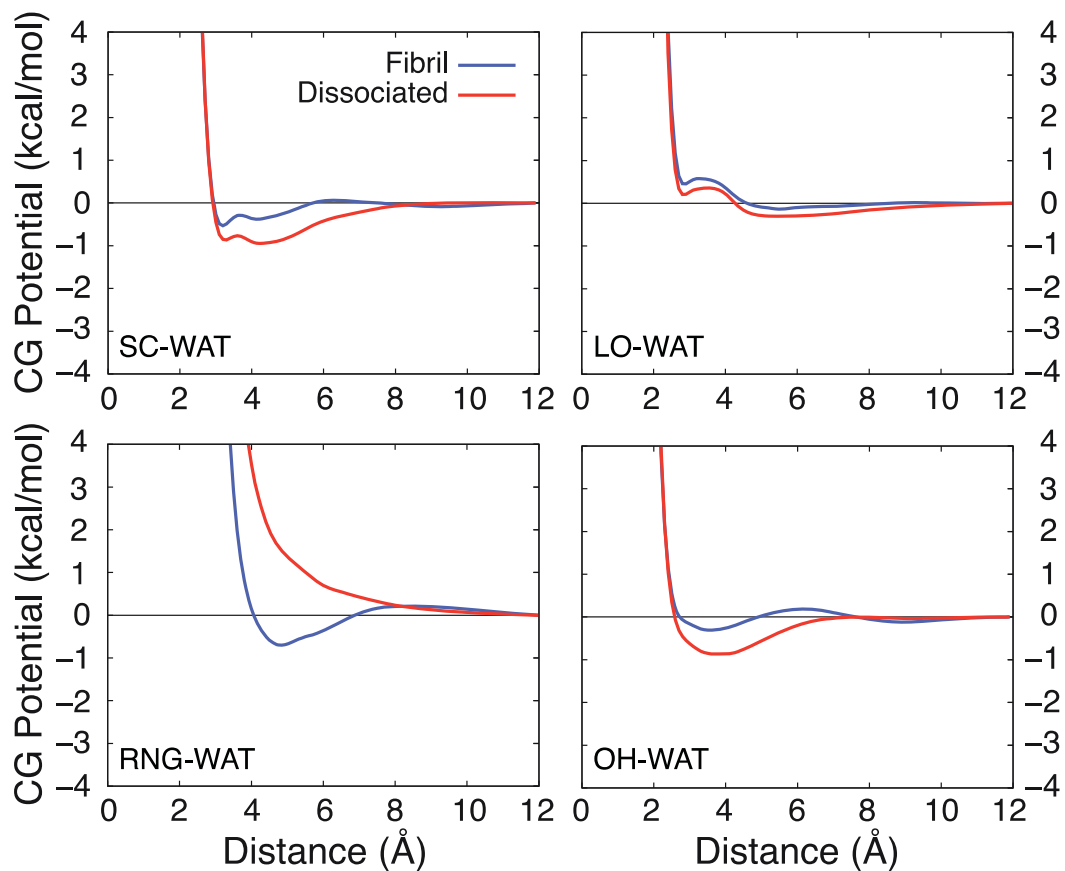
Figure 5.21. CG interaction potentials between the WAT coarse-grained group of water and the coarse-grained cellulose functionalities in LiCl/water. Coarse-grained groups are defined in Figure 5.2.

Figure 5.22. CG interaction potentials between the Li$^+$ cation and the coarse-grained cellulose functionalities in LiCl/water. Coarse-grained groups are defined in Figure 5.2.

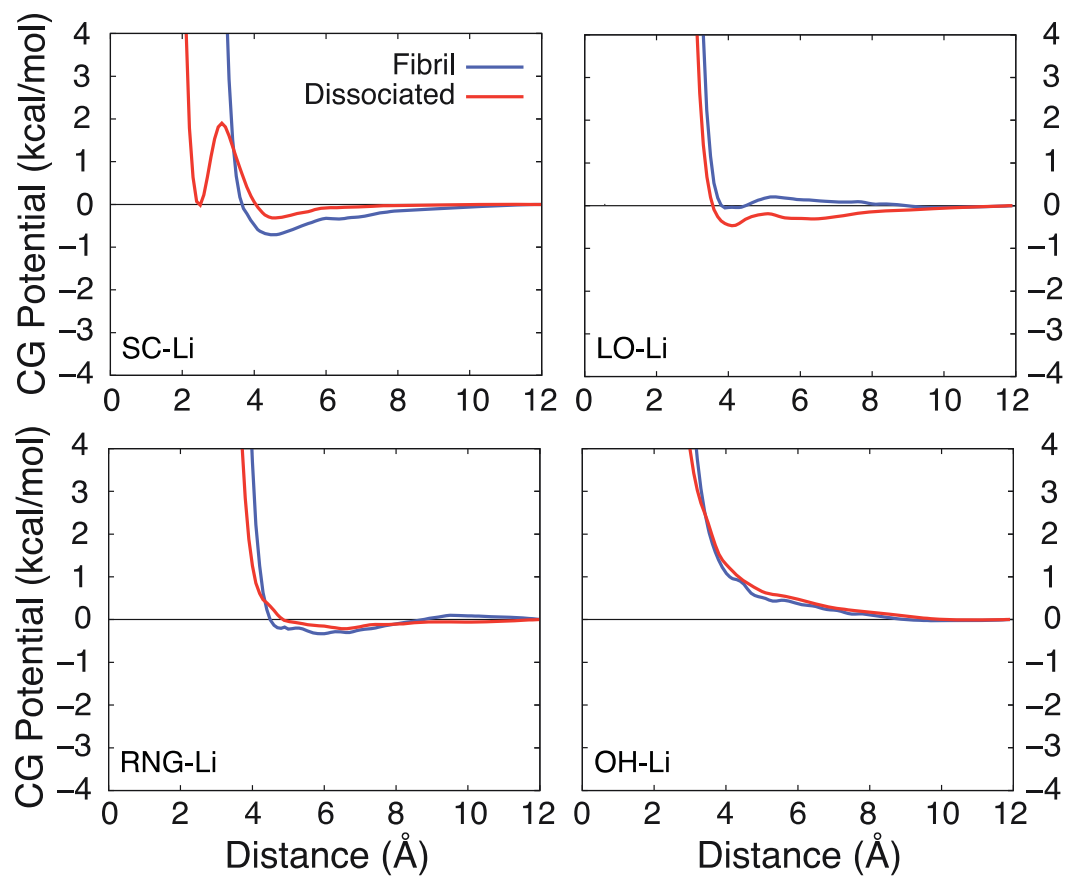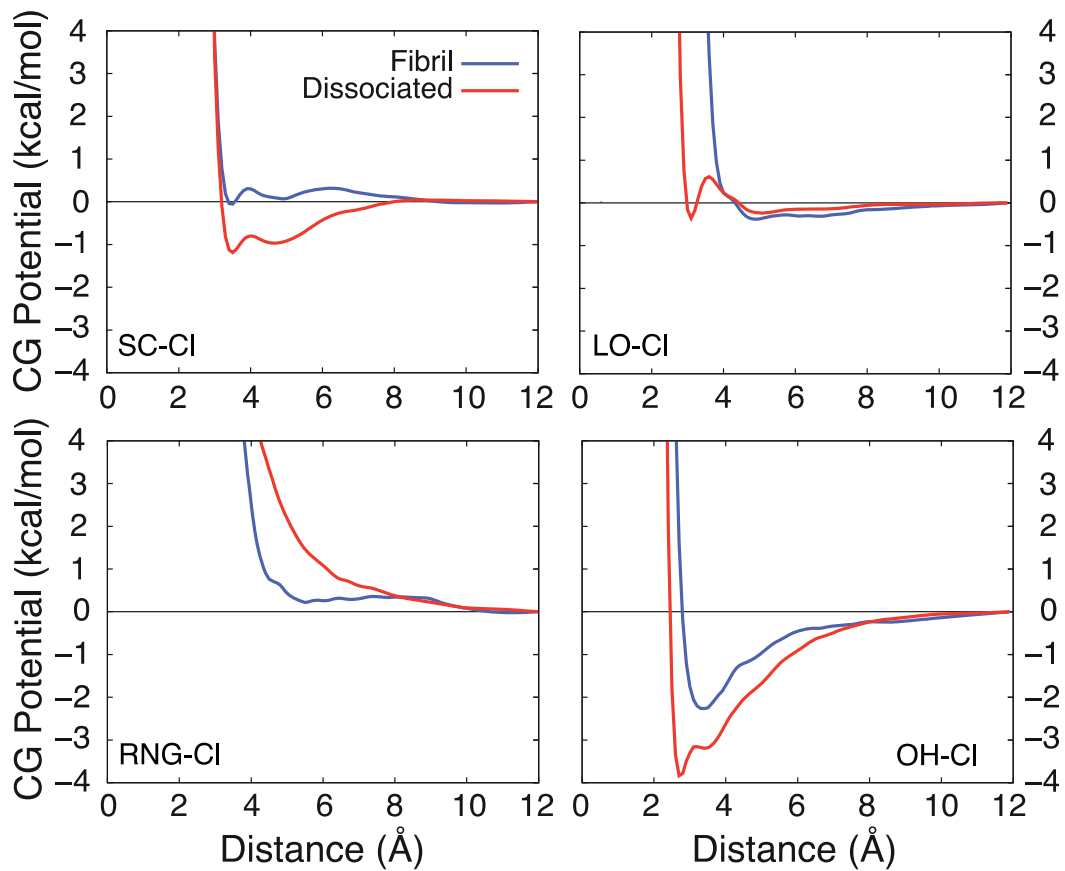**Figure 5.23.** CG interaction potentials between the Cl⁻ anion and the coarse-grained cellulose functionalities in LiCl/water. Coarse-grained groups are defined in Figure 5.2.

References

(1)   Huber, G. W.; Iborra, S.; Corma, A. *Chem. Rev.* **2006**, *106*, 4044-4098.

(2)   Perlack, R. D. *Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasability of a Billion-Ton Annual Supply*; DOE/GO-102995-2135; Oak Ridge National Laboratory: Oak Ridge, TN, 2005.

(3)   Ragauskas, A. J.; Williams, C. K.; Davison, B. H.; Britovsek, G.; Cairney, J.; Eckert, C. A.; Frederick, W. J.; Hallett, J. P.; Leak, D. J.; Liotta, C. L.; Mielenz, J. R.; Murphy, R.; Templer, R.; Tschaplinski, T. *Science* **2006**, *311*, 484-489.

(4)   Perez, S.; Mazeau, K. In *Polysaccharides: Structural Diversity and Functional Versatility*, 2nd ed.; Dumitriu, S., Ed.; Marcel Dekker: New York, 2005; pp 41-68.

(5)   Himmel, M. E.; Ding, S. Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W.; Foust, T. D. *Science* **2007**, *315*, 804-807.

(6)   Gross, A. S.; Chu, J.-W. *J. Phys. Chem. B* **2010**, *114*, 13333-13341.

(7)   Wyman, C. E. *Trends Biotechnol.* **2007**, *25*, 153-157.

(8)   Yang, B.; Wyman, C. E. *Biofuel. Bioprod. Bior.* **2008**, *2*, 26-40.

(9)   Wyman, C. E.; Dale, B. E.; Elander, R. T.; Holtzapple, M.; Ladisch, M. R.; Lee, Y. Y. *Bioresource Technol* **2005**, *96*, 1959-1966.

(10)   Lynd, L. R.; Wyman, C. E.; Gerngross, T. U. *Biotechnol. Progr.* **1999**, *15*, 777-793.

(11)   Chang, R.; Gross, A. S.; Chu, J. W. *J Phys Chem B* **2012**, *116*, 8074-8083.

(12)   Wertz, J. L.; Bedue, O.; Mercier, J. P. *Cellulose Science and Technology*; EPFL Press: Lausanne, 2010.

(13)   Klemm, D. *Comprehensive cellulose chemistry*; Wiley-VCH: Weinheim ; New York, 1998.

(14)   Mccormick, C. L.; Callais, P. A.; Hutchinson, B. H. *Macromolecules* **1985**, *18*, 2394-2401.

(15)   Mccormick, C. L.; Lichatowich, D. K. *J Polym Sci Pol Lett* **1979**, *17*, 479-484.

(16)   El-Kafrawy, A. *J Appl Polym Sci* **1982**, *27*, 2435-2443.

(17)   Chrapava, S.; Touraud, D.; Rosenau, T.; Potthast, A.; Kunz, W. *Phys Chem Chem Phys* **2003**, *5*, 1842-1847.

(18)   Morgenstern, B.; Berger, W. *Acta Polym* **1993**, *44*, 100-102.

(19)   Dawsey, T. R.; Mccormick, C. L. *J Macromol Sci R M C* **1990**, *C30*, 405-440.

(20)   Morgenstern, B.; Kammer, H. W. *Trends Polym Sci* **1996**, *4*, 87-92.

(21)   Petrus, L.; Gray, D. G.; Bemiller, J. N. *Carbohyd Res* **1995**, *268*, 319-323.

(22)   Furuhata, K. I.; Koganei, K.; Chang, H. S.; Aoki, N.; Sakamoto, M. *Carbohyd Res* **1992**, *230*, 165-177.

(23)   Balasubramanian, D.; Shaikh, R. *Biopolymers* **1973**, *12*, 1639-1650.

(24)   Bello, J.; Haas, D.; Bello, H. R. *Biochemistry-Us* **1966**, *5*, 2539-&.

(25)   Nakamura, T. *B Chem Soc Jpn* **1975**, *48*, 1447-1451.

(26)   Zelenkovskii, V. M.; Fen'ko, L. A.; Bil'dyukevich, A. V. *Polym Sci Ser B+* **2006**, *48*, 28-31.

(27)   Remsing, R. C.; Hernandez, G.; Swatloski, R. P.; Massefski, W. W.; Rogers, R. D.; Moyna, G. *J. Phys. Chem. B* **2008**, *112*, 11071-11078.

(28)   Remsing, R. C.; Swatloski, R. P.; Rogers, R. D.; Moyna, G. *Chem. Commun.* **2006**, 1271-1273.

(29)   Youngs, T. G. A.; Holbrey, J. D.; Deetlefs, M.; Nieuwenhuyzen, M.; Gomes, M. F. C.; Hardacre, C. *Chemphyschem* **2006**, *7*, 2279-2281.

(30)   Youngs, T. G. A.; Hardacre, C.; Holbrey, J. D. *J. Phys. Chem. B* **2007**, *111*, 13765-13774.

(31)   Gross, A. S.; Bell, A. T.; Chu, J.-W. *J Phys Chem B* **2011**, *115*, 13433-13440.

(32)   Morgenstern, B.; Kammer, H. W.; Berger, W.; Skrabal, P. *Acta Polym* **1992**, *43*, 356-357.

(33)   Hua, L.; Zhou, R. H.; Thirumalai, D.; Berne, B. J. *P Natl Acad Sci USA* **2008**, *105*, 16928-16933.

(34)   Street, T. O.; Bolen, D. W.; Rose, G. D. *P Natl Acad Sci USA* **2006**, *103*, 13997-14002.

(35)   Panar, M.; Beste, L. F. *Macromolecules* **1977**, *10*, 1401-1406.

(36)   Ding, S. Y.; Himmel, M. E. *J. Agr. Food Chem.* **2006**, *54*, 597-606.

(37)   Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P. *J Am Chem Soc* **2003**, *125*, 14300-14306.

(38)   Cho, H. M.; Gross, A. S.; Chu, J.-W. *J Am Chem Soc* **2011**, *133*, 14033-14041.

(39)   Brokaw, J. B.; Haas, K. R.; Chu, J. W. *J Chem Theory Comput* **2009**, *5*, 2050-2061.

(40)   Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781-1802.

(41)   Martyna, G. J.; Tobias, D. J.; Klein, M. L. *J. Chem. Phys.* **1994**, *101*, 4177-4189.

(42)   Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613-4621.

(43)   Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187-217.

(44)   Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; Mackerell, A. D. *J. Comput. Chem.* **2008**, *29*, 2543-2564.

(45)   Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. *J. Chem. Theory Comput.* **2009**, *5*, 2353-2370.

(46)   Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-935.

(47)   Beglov, D.; Roux, B. *J Chem Phys* **1994**, *100*, 9050-9063.

(48)   Lamoureux, G.; Roux, B. *J Phys Chem B* **2006**, *110*, 3308-3322.

(49)   Noid, W. G.; Chu, J. W.; Ayton, G. S.; Voth, G. A. *J Phys Chem B* **2007**, *111*, 4116-4127.

(50)   Cohen-Adas, R.; Lorimer, J. W. *Solubility Data Series*; Pergamon Press: Elmsford, New York, 1991; Vol. 47.

(51)   Potthast, A.; Rosenau, T.; Buchner, R.; Roder, T.; Ebner, G.; Bruglachner, H.; Sixta, H.; Kosma, P. *Cellulose* **2002**, *9*, 41-53.

(52)   Kovacs, A.; Varga, Z. *Coordin Chem Rev* **2006**, *250*, 710-727.