# The Dissolution of Cellulose in Ionic Liquids - A Molecular Dynamics Study

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Barmak Mostofian

December 2014

© by Barmak Mostofian, 2014
All rights reserved.

*To family and friends!*

# Acknowledgements

I would like to express my gratitude to all my colleagues and friends at the UT/ORNL Center for Molecular Biophysics. I cannot imagine a better group of researchers to work with. In particular, I would like to thank my advisors Jeremy C. Smith and Xiaolin Cheng. Their unlimited support and guidance has been instrumental to me in reaching my academic goals. It was a pleasure to learn from them every single day and a privilege to be their student.

Strength of character, in compassion and honor, was both the standard and the goal. I could only try to find in myself that same strength to act appropriately.
Tara Fatemi, *It Was Nearing Midnight*

# Abstract

The use of ionic liquids for the dissolution of cellulose promises an alternative method for the thermochemical pretreatment of biomass that may be more efficient and environmentally acceptable than conventional techniques in aqueous solution. Understanding how ionic liquids act on cellulose is essential for improving pretreatment conditions and thus detailed knowledge of the interactions between solute and solvent molecules is necessary. Here, results from the first all-atom molecular dynamics simulation of an entire cellulose microfibril in 1-butyl-3-methylimidazolium chloride (BmimCl) are presented and the interactions and orientations of solvent ions with respect to glucose units on the hydrophobic and hydrophilic surfaces of the fiber are analyzed in detail, shedding light on the initiation stages of cellulose dissolution. Moreover, replica-exchange simulations of a single cellulose chain fully solvated in BmimCl and in water are performed for a total of around 13 $\mu$s in order to study the dynamics and thermodynamics of the end state of the dissolution.

The results indicate that chloride anions predominantly interact with cellulose hydroxyl groups and disrupt the intrachain O3H'$\cdots$O5 hydrogen bonds, which are essential for the integrity of cellulose fibers. The cations stack preferentially on the hydrophobic cellulose surface, governed by non-polar interactions with cellulose, which can stabilize detached cellulose chains by compen-

sating the interaction between stacked layers. Moreover, a frequently occurring intercalation of cations on the hydrophilic surface is observed, which by separating cellulose layers can also potentially facilitate the initiation of fiber disintegration. The single-chain simulations indicate that differences in cellulose solvation mechanisms between the two solvents exist. Although global size-related properties of the cellulose chain are comparable in the two solvents, local conformational properties of cellulose differ significantly between the BmimCl and water solutions. In general, the results indicate that solute-solvent interaction energies are more favorable and that the cellulose chain is more flexible in BmimCl than in water.

Taken together, the simulations explain how ionic liquids can facilitate cellulose dissolution: the synergistic action of anions and cations helps to initiate fiber deconstruction through specific interactions on the fiber surface and to solvate single cellulose chains through favorable solvent interactions and conformational flexibility.

# Table of Contents

**1 Introduction** **1**

**2 Background and Significance** **5**

2.1 Pretreatment of Lignocellulosic Biomass . . . . . 5

2.2 The Use of Ionic Liquids for Biomass Dissolution . . . . . 8

**3 Simulation Theory** **13**

3.1 Molecular Dynamics Simulation . . . . . 13

3.1.1 Molecular Mechanics Force Fields . . . . . 14

3.1.2 Equations of Motion . . . . . 15

3.2 Replica-Exchange Simulations . . . . . 16

3.3 Configurational Entropy Calculation . . . . . 18

**4 Molecular Dynamics Simulations of Cellulose Fibers** **20**

4.1 Solvent Structure around the Hydrophobic (200) Surface . . . 24

4.2 Solvent Structure around the Hydrophilic ($1\bar{1}0$) Surface . . . . 35

4.3 Solvent Structure around the Hydrophilic (010) Surface . . . . 40

**5 Replica-Exchange Simulations of Cellulose Chains** **50**

5.1 Cellulose Conformational Changes . . . . . 51

5.2 Cellulose-Solvent Interactions . . . . . 73

5.3 Cellulose Configurational Entropy . . . . . 79

**6 Conclusions** **82**

6.1 Future Perspectives . . . . . 84

**Bibliography** **86**

**Appendix** **115**

**Vita** **120**

# List of Figures


1 Molecular structure of cellulose. . . . . 7

2 Publications on ionic liquids. . . . . 9

3 Molecular structure of BmimCl. . . . . 11

4 Cellulose chain tilts and hydroxyl SASA contributions. . . . . 22

5 RDFs of solvent ions around cellulose surface monomers. . . . 25

6 RDFs of $Cl^-$ anions around cellulose surface hydroxyl groups. 27

7 Cellulose intrachain O3H'$\cdots$O5 hydrogen bond occupancies. . 28

8 Average spatial distributions of solvent with respect to a cellobiose unit. . . . . 29

9 Analysis of solvent cation orientations with respect to solute monomers on the (200) surface. . . . . 34

10 Solvent cation orientations. . . . . 36

11 Interaction potential energy components. . . . . 37

12 Analysis of solvent cation orientations with respect to solute monomers on the $(1\bar{1}0)$ surface. . . . . 41

13 Analysis of solvent cation orientations with respect to solute monomers on the (010) surface. . . . . 43

14 Average Coulombic energy between polar imidazole ring hydrogens and cellulose or $Cl^-$ anions. . . . . 46

15 Potential energy distributions of REMD simulations. . . . . 52
16 REMD time series. . . . . 53
17 Radius of gyration of cellulose in the two solvents. . . . . 54
18 End-to-end distance of cellulose in the two solvents. . . . . 56
19 Solvent-accessible surface area and solvent-excluded volume of cellulose in the two solvents. . . . . 56
20 Cellulose hydroxymethyl and intramolecular hydrogen bond profiles. . . . . 60
21 Cellulose temperature profiles of P(GG)/P(GT) and P(O3H'$\cdots$O5). 61
22 Free energy landscapes for P(GG)/P(GT) and P(O3H'$\cdots$O5). 63
23 $\phi$-$\psi$ scatter plots. . . . . 65
24 Free energy landscapes for RMSD($\phi, \psi$) and P(O3H'$\cdots$O5). . 68
25 Cellulose temperature profiles of P($^4C_1$) and $c_n$. . . . . 70
26 Spatial distributions of solvent. . . . . 75
27 Solute-solvent interaction energies. . . . . 77
28 Solute-solvent interaction energies per residue. . . . . 78
29 Cellulose configurational entropy. . . . . 80

# 1

# Introduction

The transition of energy production from petroleum and coal to carbon-neutral and renewable resources is one of the most demanding scientific tasks of our time[1–4]. Mainly due to the rapid growth of developing nations, the energy demand is projected to grow by more than 50% within the next 20 to 25 years[5]. As more than half of the world's energy consumption is currently used for transportation[6], the development of cost-effective methods for the generation of liquid transportation fuels is imperative[7–9]. In this context, biofuels (or bioethanol) have been identified as a sustainable alternative to fossil fuels[10, 11]. While so-called "first-generation" biofuels are mainly produced from sugar cane or corn starch and are often added to gasoline, the next generation of biofuels is supposed to be derived from the non-edible parts of plants, thus not competing with food resources, and to deliver greater energy efficiency[12–14].

Recently, some research has been focused on the development of biofuels through exploring additional biomass feedstocks[15, 16] and new strategies for biofuel production[17]. The goal is to develop advanced biorefineries that

routinely make cellulosic ethanol based on plant waste material, or lignocellulosic biomass, which is an abundant and renewable energy source[18–20]. In plant cell walls, cellulose fibers are attached to hemicellulose and lignin. With its dense arrangement and structural heterogeneity, lignocellulosic biomass is fairly resistant to hydrolysis, which has been referred to as "biomass recalcitrance"[21]. Thus, the efficient dissolution of cellulose to render it more accessible to hydrolyzing enzymes is key to an economical and sustainable biofuel production process.

Discovery and development of cellulose solvents have a long history[22]. Depending on the cellulose source and the desired modification, different types of solvents are employed, which are traditionally categorized based on whether they are aqueous or non-aqueous and derivatizing or non-derivatizing, as well as on their composition and type of interaction with cellulose[23–25]. It has been widely accepted that, in order to disintegrate the cellulose fibril structure, a good cellulose solvent must break the inter- and intramolecular hydrogen bonds between and within glucan chains[26, 27]. This general viewpoint, however, has recently been challenged in several reports claiming that the role of hydrophobic interactions in the solubility of cellulose had been underestimated[28, 29]. Indeed, potential of mean force calculations for the separation of cellooligomers have suggested that hydrophobic interactions contribute favorably to stabilizing a crystal-like stacked structure[30]. Moreover, when considering the intricate cellulose solubility problem, the amphiphilic character of cellulose polymers should not be overlooked. There is some agreement that the amphiphilic nature of cellulose affects its structural properties - a concept referred to as the "Lindman Hypothesis"[31]. Therefore, in order to simultaneously weaken the solute's hydrophobic and hydrogen bond interactions and facilitate dissolution in a more effective way, a good cellulose solvent

should have both polar and non-polar groups. In this context, ionic liquids (ILs) such as 1-butyl-3-methylimidazolium chloride (BmimCl), which has been shown to dissolve cellulose more efficiently than water[32], could potentially represent a type of solvent useful for the large-scale dissolution of lignocellulose in biorefineries.

In this study, all-atom molecular dynamics (MD) simulations are used to examine the dissolution of cellulose in BmimCl. MD simulations are generally useful for the analysis of interactions at atomic detail and they are particularly suited for the investigation of the non-derivatized dissolution of cellulose by ILs. To understand the enhanced cellulose dissolution ability of ILs compared to water, which may be controlled by favorable thermodynamics rather than kinetics[28], one has to investigate the interaction structures and energies of the two different solvents with cellulose. All-atom simulations of the entire dissolution process of cellulose in ILs, which involves the transition from the fiber state to the fully dissolved state of individual chains[33], is not currently feasible due to the large system size (generally >100,000 atoms) and the concomitant timescale limitation to the sub-$\mu$s regime. Therefore, atomistic MD simulations can only target specific cellulose-solvent interactions that may occur during the dissolution process. Moreover, due to the slow dynamics of the ILs, which are referred to as "sticky" or "sluggish" in the literature[34], it must be assured that the simulation converges and that the results are significant.

Here, the analysis of solute-solvent interactions of an entire cellulose fiber simulated in BmimCl at 450 K, which corresponds to the typical pretreatment temperature in ILs, is presented, thus shedding light on the initiation stages of cellulose dissolution. Moreover, replica-exchange MD simulations of a fully solvated single cellulose chain are performed in order to study the end state of the dissolution process. Analyzing the initial and final stages of cellulose dis-

solution in BmimCl and water may explain the more favorable dissolution in the IL. This study provides complementary information to previous MD simulations on the structure and dynamics of cellulose and the free energy of single chain detachment from cellulose microfibrils in aqueous and IL solutions[35–37].

This dissertation is organized as follows: chapter 2 introduces the significance of ILs as a solvent for lignocellulose pretreatment. The theoretical background of the applied MD simulations are described in chapter 3. The following two chapters present the reults on the simulations of cellulose fibers and single chains, respectively. The concluding chapter 6 includes potential perspectives for future studies.

# 2

# Background and Significance

## 2.1 Pretreatment of Lignocellulosic Biomass

Lignocellulose is a fibrous material mainly composed of the carbon-rich polymers cellulose, hemicellulose, and lignin. It provides strength and vascular function to plant cell walls[38, 39]. In particular, its cellulose content can be relatively high (up to 90% in cotton and 40%-50% in wood).

Cellulose forms microfibrils after synthesis in plants. As shown in Figure 1a, these contain 36 unbranched $\beta$-glucan chains arranged in parallel sheets or layers that are held together by regular intra- and intermolecular hydrogen bonds within the sheets and stacking interactions between them[40]. Cellulose microfibrils may aggregate to larger fibers estimated to be up to several nm in diameter and up to several $\mu$m in length[41–43]. Usually, more than half of the cellulose in plants is crystalline[44], in particular in the fiber core, and two different native forms can occur, I$\alpha$ and I$\beta$[45]. These two crystal allomorphs differ in their hydrogen bonding patterns[46, 47] and the I$\beta$ form is predominant in higher plants[44]. Treated cellulose can adopt even further al-

lomorphs such as cellulose II (mercerized cellulose)[48], cellulose III (ammonia treated)[49, 50], or cellulose IV (thermally treated)[51], which vary by their hydrogen bond network or chain association[52].

Cellulose I$\beta$ has a monoclinic unit cell[53] and, depending on the position in the unit cell (see Figure 1a), glucan chains are referred to as center or origin chains, which are non-equivalent in terms of their environment. Their parallel arrangement in layers results in hydrophilic and hydrophobic fiber surfaces depending on whether the hydroxyl groups of the glucose subunits are exposed on that surface or not. These fiber faces can be annotated by Miller indices as they correspond to lattice planes of the underlying crystal (see Figure 1a).

Technically, the production of cellulosic ethanol is a two-tier process: the degradation of lignocellulose to smaller compounds[54], which can be converted more readily, and the actual fermentation to ethanol[55]. The first step includes the pretreatment of biomass in order to disrupt plant cell walls and to expose the sugar containing material and the hydrolysis to fermentable mono- and disaccharides. Due to the above-mentioned recalcitrance of biomass, this step represents the main economic costs of biofuel production.

Optimization of pretreatment techniques has been the focus of many recent research efforts. Current thermochemical pretreatment methods use chemicals such as ammonia[56, 57], sulfites[58, 59], or dilute phosphoric or sulfuric acids[60, 61] for fiber expansion and disruption of the lignocellulose structure at elevated temperature and pressure. The use of any of these approaches depends on the source of biomass and its lignin and cellulose content[62, 63]. While any thermochemical pretreatment has its distinct advantage, for instance reducing cellulose crystallinity or increasing the separation of the lignocellulose components, they all produce side products that can be toxic or

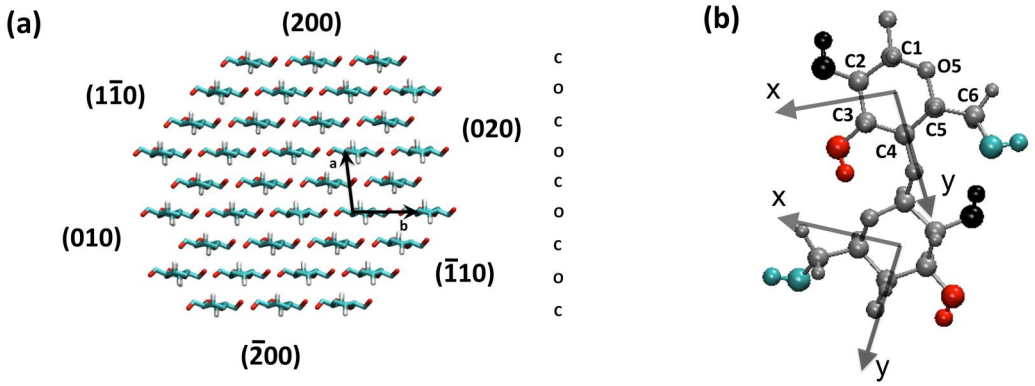



**Figure 1** Molecular structure of cellulose.

(a) A cross-section view of a 36-chain microfibril model of the cellulose I$\beta$ allomorph. Axes **a** and **b** define the monoclinic unit cell. The six faces are assigned Miller Indices. Layers consisting of center (C) or origin (O) chains are labeled on the right.
(b) The definition of the coordinate system on solute monomers for the orientation analysis. The reference coordinate system is defined for glucose monomers with the xy-plane as the plane of the pyranose ring (that is the plane defined by C2, C4, and O5). This view is along the negative z-axis. The y-axis is defined along the connection of the ring-center and the glycosidic bond. The x-axis is defined to be perpendicular to the y- and z-axes as shown here. For any monomer of a cellobiose unit, functional groups appear on different sides around the defined axes and thus both types are displayed here. The cellobiose hydroxyl groups of C2 (black), C3 (red), and C6 (cyan) are colored.

interfere with the subsequent step of hydrolysis[64, 65]. In this regard, more efficient and environmentally acceptable alternatives are desirable. For instance, a biological approach to pretreatment utilizes fungal peroxidases that can degrade lignin[66, 67]. Also, the use of recyclable ILs for the non-derivatizing dissolution of lignocellulose is a promising pretreatment method for a sustainable biofuel industry.

## 2.2 The Use of Ionic Liquids for Biomass Dissolution

ILs are generally defined as salts which melt below 100°C. They are usually liquid over a wide temperature range, which is thermodynamically favored because of the large size of the ions resulting in small lattice enthalpies and large entropy changes that favor the liquid state[68]. Due to their low volatility and thermal stability, ILs are suggested to be employed as reusable "green" solvents for several industrial processes such as catalysis, extraction, and synthesis[69–73]. The recent rise in popularity of ILs is reflected by the explosion of the number of scientific articles published on them as visualized in Figure 2. This number has increased from less than a thousand to over six thousand within the last ten years. Currently, ILs are being explored in a variety of applications ranging from the synthesis of small chemicals to the capture of greenhouse gases[34, 74, 75].

ILs had first been reported in 1914 by Walden[76] and their capability to dissolve cellulose was also documented for the first time as early as in the 1930s[77, 78]. However, the recent surge in investigating their potential as cellulose solvents has been triggered by the work of Swatloski et al. in 2002[79,

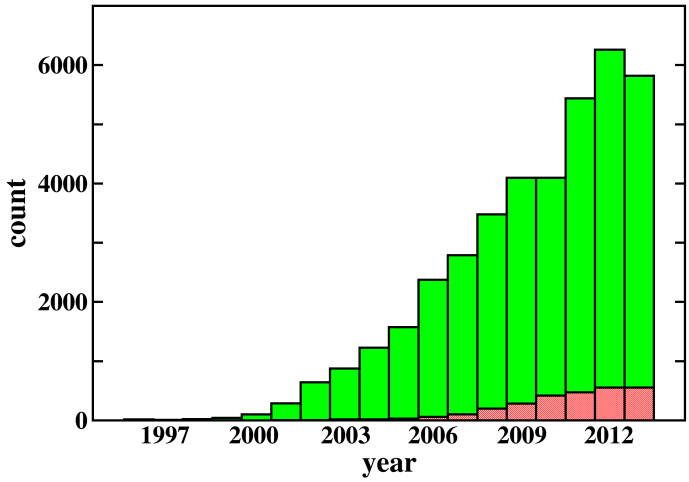



**Figure 2** Publications on ionic liquids.

Number of published articles on ILs (green bars) or on ILs and cellulose (red bars) per year. The counts are retrieved by Scifinder.

80], who found that high amounts of cellulose can be dissolved by BmimCl in a relatively short time. Since then, ILs have been exploited as solvents in cellulose research (see Figure 2), for instace for shaping the polysaccharide into fibers, films, or beads[81–84], for preparing derivatives through homogeneous chemical modifications[85–88], and last but no least, for the non-derivatized pretreatment of lignocellulosic biomass[89–93].

The application of ILs outperforms common methods of pretreatment, making cellulose more prone to deconstruction during the subsequent process of saccharification[94, 95]. Enhanced turnover rates of enzymatic hydrolysis arise probably from better accessibility of the polymer to enzymes after treatment with ILs[96, 97]. Characterization of the pretreated and reprecipitated

cellulose, via scanning electron microscopy and X-ray diffraction among other techniques, has indicated structural transformation[98, 99], which manifests itself as greater delignification and reduced crystallinity of cellulose. Depending on the source of biomass as well as on the pretreatment temperature, conversion of cellulose I to cellulose II upon IL treatment and recrystallization has been observed[100, 101]. The cellulose II allomorph is known to be more readily digestible by hydrolyzing enzymes[102, 103].

Within the large group of ILs, those with a cation containing a heterocyclic group and a short alkyl-chain are most effective in interacting with carbohydrate substrates in general and in dissolving cellulose in particular[89, 92, 104–107]. The ability of imidazolium-based ILs, such as BmimCl (Figure 3), to decompose cellulose fibers has been measured under various conditions[108, 109]. However, in order to better exploit ILs as an efficient and more economic solvent system for biomass pretreatment, the molecular mechanism by which they act on cellulose needs to be better understood and underlying distinct interaction patterns between solute and solvent have to be identified[110–114]. Shedding some light on this issue is of utmost interest since it would help design more sophisticated solvents for biomass degradation.

Recent NMR measurements of cellobiose and cellulose model compounds solvated in different ILs with various cation-anion pairs have revealed the predominant interactions between anions and the hydroxyl groups of the solutes[115, 116]. After numerous cellulose solubility studies, it has become widely accepted that the greater hydrogen bond basicity of the anion compared to water oxygens, that is its potential to make stronger hydrogen bonds with cellulose than do water oxygens, is important for efficient cellulose dissolution in ILs[92, 117]. In contrast, the cation's role in this process is still a matter of debate. While some researchers suggested a mechanism in which the cation is also engaged

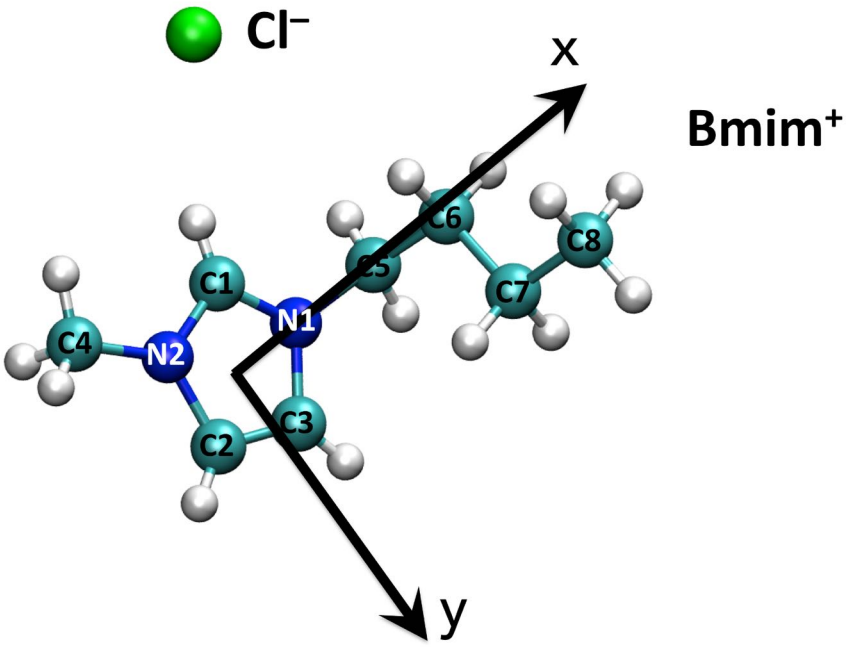



**Figure 3** Molecular structure of BmimCl.

The BmimCl ion pair consists of an organic cation with an imidazole-containing head group and a chloride anion. The coordinate system for the orientational analysis is defined for $Bmim^+$ cations with the x-y-plane as the plane of the imidazole ring (that is the plane defined by N1, N2, and C2). This view is along the negative z-axis. The x-axis is defined along the connection of the ring-center to the butyl group. The y-axis is defined to be perpendicular to the x- and z-axes as shown here.

in hydrogen bonding to cellulose hydroxyl groups[118], others emphasized the importance of the aromatic ring structure for its possible hydrophobic stacking interaction with the glucose rings[33]. Yet, another hypothesis raised the possibility that cations activate their counterions for cellulose dissolution[117].

All-atom MD simulations are a very useful tool for the analysis of site-site interactions and thus can help to shed some light on interactions between cellulose and ILs, which are relevant for dissolution. Recent studies on cellulose oligomers solvated in ILs have shown direct hydrogen bonding between the anions of different ILs and the cellulose hydroxyl groups[119] and different interaction energies between cations of varying alkyl chain length and cellulose[120]. Moreover, simulations have revealed more favorable solute-solvent interaction energies of glucan chains in IL solutions than in aqueous solutions[121]. Similarly, MD simulations of a cellulose oligomer in binary solvent systems with varying IL/water ratios characterized molecular interactions between cations, anions, water and cellulose during precipitation after IL pretreatment[122]. Recent simulations that were performed on crystalline cellulose bundles or microfibrils indicate that both the small anions and the bulky imidazole-containing cations can penetrate the arrangement of glucan chains[123, 124] and that the cation may even interact more strongly with cellulose during its deconstruction[125].

# 3

# Simulation Theory

## 3.1 Molecular Dynamics Simulation

Numerical simulations have become the third pillar of science, connecting experiments and theory. In particular, molecular modeling and simulation methods deal with the study of materials and biologcial systems, covering a range of size scales from the sub-atomic level of quantum mechanics to the micrometer scale of bulk materials[126]. On the atomistic level of molecular mechanics, Monte-Carlo and Molecular Dynamics can be used to predict the thermodynamic and dynamic properties of the simulated system based on the principles of equilibrium and non-equilibrium statistical mechanics[127].

In fact, MD simulations are now one of the main tools for the computational study of biomolecules, for they provide complementary information to experimental results on the dynamical motions that underlie biological function[128]. Moreover, MD allows the investigation of molecular regions and atomic interactions at a time and size resolution that is not accessible to experimental techniques. In recent years, MD simulations have significantly contributed to

understanding a large variety of biological processes such as protein folding, molecular recognition mechanisms, or ion transport through transmembrane channels[129–132]. Also, the study of solvent structure around a biomolecular solute and the analysis of solute-solvent interactions benefited from MD simulations[133, 134].

### 3.1.1 Molecular Mechanics Force Fields

Empirical molecular mechanics energy functions (or force fields) describe the energy of a molecular system as a function of the nuclear coordinates alone, which is appropriate based on the Born-Oppenheimer Approximation[135]. Equation (3.1) shows the functional form of a general force field, where the total potential energy of the system, $E(\mathbf{r}^N)$, is given as the sum over all bonded and non-bonded interactions. The former consists of the summation of harmonic interaction potentials for any bond, angle, or torsion, and the latter represents van der Waals and Coulomb interaction energies of all atom pairs separated by three or more covalent bonds.

$$E(\mathbf{r}^N) = \sum_{bonds} \frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{angles} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 + \sum_{torsions} \frac{V_n}{2}(1 + cos(n\omega_i - \gamma_i)) + \sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}\right) \tag{3.1}$$

Although relatively simple functions are used to describe the contributions from processes such as bond or angle stretching, force fields can estimate the potential energy quite well even for molecular systems with a significant number of atoms. However, they do not allow the simulation of chemical bond

creation or cleavage or the simulation of electronic polarization. Both limitations can be alleviated for specific circumstances by the use of hybrid QM/MM simulation software[136] or special purpose force fields[137, 138], respectively, which further augment the applicability of MD simulations.

Force field parameters, such as equilibrium bond lengths $l_i$ or force constants $k_i$, are usually derived from quantum mechanical calculations and refined upon comparing results to experiments[139]. For computational efficiency, non-bonded interactions are truncated and evaluated with neighbor lists[140]. For long-range electrostatic interactions, however, this approach may result in artifacts that violate the conservation of the total potential energy[141]. Thus, electrostatics are usually treated by the calculation of an extra term, for instance with the Particle Mesh Ewald (PME) method where the global charge distribution is calculated in reciprocal space[142].

One key feature of molecular mechanics force fields is their transferability, which allows the application of one set of parameters to a range of problems. In the present work, BmimCl parameters are used that were derived based on the AMBER force field[143] and the TIP3P water model[144] along with the GLYCAM06 parameters[145] for cellulose molecules. More details on the simulation parameters can be found in Appendix A).

### 3.1.2 Equations of Motion

In MD, a time-dependent trajectory is generated with coordinates and velocities of all particles of the underlying system by solving equations embodied in Newton's second law of motion $\frac{d^2\mathbf{r}_i}{dt^2} = \frac{\mathbf{F}_i}{m_i}$, where $\mathbf{F}_i$ is the force exerted on particle $i$, $m_i$ is its mass, and their ratio is the accelaration on particle $i$, $\mathbf{a}_i$. Forces can be expresed as the gradient of the potential energy: $\mathbf{F}_i = -\nabla_i E(\mathbf{r}^N)$.

Various numerical algorithms have been developed for integrating the equations of motion, all of which assume that the particles' coordinates, velocities, and accelerations can be approximated by a Taylor series[146–148]. One such method is the Verlet algorithm[140]:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2 + ... \tag{3.2}$$

$$\mathbf{r}(t-\delta t) = \mathbf{r}(t) - \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2 - ... \tag{3.3}$$

Addition of (3.2) and (3.3) gives:

$$\mathbf{r}(t+\delta t) = 2\mathbf{r}(t) - \mathbf{r}(t-\delta t) + \mathbf{a}(t)\delta t^2 \tag{3.4}$$

The above algorithm shows that MD is a deterministic process since the coordinates (and velocities) are derived from the previous ones in the trajectory. In order to achieve numerical stability and to avoid errors in the calculation of forces in atomistic systems, the typical time step $\delta t$ has to be smaller than the fastest motions, that is the vibrations of hydrogens: $\delta t \leq 1fs$. A larger time step of $\delta t = 2fs$ can be achieved by constraining the position of hydrogen atoms to the equiibrium bond lengths with the corresponding heavy atoms[149, 150].

## 3.2 Replica-Exchange Simulations

In order to obtain an accurate canonical distribution from simulation, molecular conformational space has to be sampled adequately. For many biologically relevant molecules this is problematic due to the ruggedness of their free energy landscape and the kinetic trapping in local minima of the simulated system.

Various algorithms have been developed to overcome this issue, examples being the multicanonical method[151], simulated tempering[152], or the replica-exchange method[153]. In the replica-exchange approach, simulations of the same system (replicas) are run at different temperatures. At regular time intervals, replicas at neighboring temperatures are exchanged with a given probability, $P_{ex}$. Replica exchange results in a random walk in temperature space which induces a random walk in potential energy space, thus allowing escape from free energy minima and improving phase space sampling.

For replica-exchange molecular dynamics (REMD), it has been shown[154] that the exchange probability $P_{ex}$ is independent of the kinetic energy of the system as the corresponding velocities are scaled with $\sqrt{T_m/T_n}$ after transition of a replica from temperature $m$ to the neighboring temperature $n$ (and scaled with $\sqrt{T_n/T_m}$ after transition of a replica from temperature $n$ to $m$):

$$P_{ex} = min\{1, exp[(\beta_m - \beta_n)(E_{(m)} - E_{(n)})]\}, \tag{3.5}$$

where $E_{(m)/(n)}$ are the potential energies (which only depend on the configuration of the system) for replicas at temperatures $m$ and $n$, respectively, and $\beta_{m/n} = 1/k_B T_{m/n}$. In particular, for simulations in the isobaric-isothermal (NPT) ensemble, which are performed at a constant temperature $T_i$ and at a constant pressure $P_i$, this probability is corrected with energy terms related to the change in PV-work[155]:

$$P_{ex} = min\{1, exp[(\beta_m - \beta_n)(E_{(m)} - E_{(n)}) - (\beta_m P_m - \beta_n P_n)(V_{(m)} - V_{(n)})]\}, \tag{3.6}$$

where $V_{(m)/(n)}$ are the system volumes for replicas at temperatures $m$ and $n$, respectively. The correction term is small compared to the potential energy.

Thus, the overlap of neighboring replica potential energy distributions is a qualitative measure of the replica exchange acceptance or the exchange ratio between neighboring replicas.

In REMD, the number of replicas required to span a given temperature range at constant $P_{ex}$ scales with $\sqrt{f}$, where $f$ is the total number of degrees of freedom in the molecular system[156]. The temperature range is chosen such that energy minima can be sampled at the lowest temperature and the system can escape those minima at the highest temperature. In the present study of solvated cellulose, it is meaningful to choose a temperature range that also includes the typical cellulose pretreatment temperatures in water and in BmimCl. More details on the REMD simulation parameters can be found in Appendix B).

## 3.3 Configurational Entropy Calculation

The most widely used analytical method for the calculation of conformational entropies from simulations is the quasi-harmonic (QH) approximation that estimates the absolute entropy of a molecule as the sum of entropy contributions from its degrees of freedom[157–160]. The QH method has been applied for the analysis of configurational entropy of various molecular systems, such as peptides, DNA, and hydrocarbons[161–163]. It is known that the QH analysis overestimates the entropy value due to artifacts arising from anharmonicity and pairwise correlations[164, 165], and thus correction terms and various alternative approaches have recently been proposed[166–171].

Based on a multivariate harmonic approximation of the configurational

probability distribution, the absolute intramolecular entropy is estimated as:

$$S = \frac{1}{2}Nk_B + \frac{1}{2}k_B[ln(2\pi)^N + ln|\mathbf{C}|] \tag{3.7}$$

Here, $N$ is the number of degrees of freedom, $k_B$ is the Boltzmann constant, and $\mathbf{C}$ is the covariance matrix of the (internal) coordinates $\mathbf{q}$, i.e. $\mathbf{C} =< (\mathbf{q} - < \mathbf{q} >) \bigotimes (\mathbf{q} - < \mathbf{q} >) >$, where angular brackets indicate canonical ensemble averaging and $\mathbf{a} \bigotimes \mathbf{b}$ is the matrix with components $ij$ equal to $a_i b_j$. According to Di Nola[172], the entropy consists of an uncorrelated diagonal part, $S_{diag}$, and an expression for the correlation correction, $S_{corr}$:

$$S_{diag} = \frac{1}{2}Nk_B + \frac{1}{2}k_B[ln(2\pi)^N + \sum_i ln(C_{ii})] \tag{3.8}$$

$$S_{corr} = \frac{1}{2}k_B ln(\frac{|\mathbf{C}|}{\prod_i C_{ii}}) \tag{3.9}$$

Instead of applying the quasi-harmonic approximation, the diagonal part can be replaced with the exact entropy evaluation for every degree of freedom $i$, which gives:

$$S = -k_B \sum_i \int dq_i P(q_i) ln(P(q_i)) + \frac{1}{2}k_B ln(\frac{|\mathbf{C}|}{\prod_i C_{ii}}) \tag{3.10}$$

# 4

# Molecular Dynamics Simulations of Cellulose Fibers

The starting cellulose structure for the fiber simulation is a microfibril crystal that consists of 36 cellulose chains arranged in nine layers (see Figure 1a). During the MD simulation several changes in the conformation of the cellulose fiber are observed. For example, the cellulose fiber swells and twists from the beginning of the simulation. These changes, together with the different orientations of hydroxymethyl groups on center chains compared to origin chains, have been observed before in simulations of I$\beta$-cellulose fibers[35, 173–177]. The swelling is quantified in the current simulation by an increase in the distance between neighboring unit cells in the fiber core by $\sim$0.3 Å on average along the cell dimension **a**. The observed fiber twist in the simulation is a further feature that has been discussed in the literature[175, 178, 179] and it has been concluded that twisted microfibrils represent an appropriate model for cellulose I$\beta$[180]. A calculation for the central layer of glucan chains shows that the fiber twists around its polymerization axis by $\sim$20° on average in the

current MD simulation. In addition, the center and origin chains tilt clockwise or counter-clockwise, respectively, with respect to the fiber's polymerization axis, as observed in previous fiber simulations[175]. Analysis of the tilt angle for any cellulose chain reveals an average value of approximately +10° for center chains and -10° for origin chains, respectively (Figure 4). This different tilt of the center and origin chains gives rise to certain hydroxyl groups being more solvent-exposed than others on the same fiber surface. The fact that the solvated fiber behaves as observed before proves the convergence and appropriateness of the current simulation for the analysis of cellulose-IL interactions.

The simulated cellulose fiber has six faces (see Figure 1a); two of these, the (200) and ($\bar{2}$00) surfaces, are generally referred to as hydrophobic, while the remaining four surfaces, (1$\bar{1}$0), ($\bar{1}$10), (010), and (020), all expose the equatorial hydroxyl groups to the surroundings and are thus referred to as hydrophilic. Parallel surfaces from opposite sides of the fiber are denoted as corresponding surfaces as their shape is identical to each other and different from the other surfaces. There are three pairs of corresponding surfaces in the given fiber structure. Any two corresponding surfaces expose hydroxyl groups or ring hydrogens to the solvent in the same way. To explore the behavior of solvent and its possible effect on a cellulose fiber, contacts between BmimCl and the cellulose surfaces are characterized. In the following, the structure of $Bmim^+$ and $Cl^-$ in proximity to only one surface from any pair of corresponding surfaces is discussed, namely the hydrophobic (200) and the hydrophilic (1$\bar{1}$0) and (010) faces (see Figure 1a). As the analysis of spatial density distributions in earlier MD studies has revealed, the solvent arranges itself in a symmetric fashion around cellulose fibers during the initiation phase of dissolution[35, 176, 181]. It has been shown that solvent components such as $Bmim^+$ and $Cl^-$ occupy distinct sites on cellulose surfaces, for example the space between two cellulose

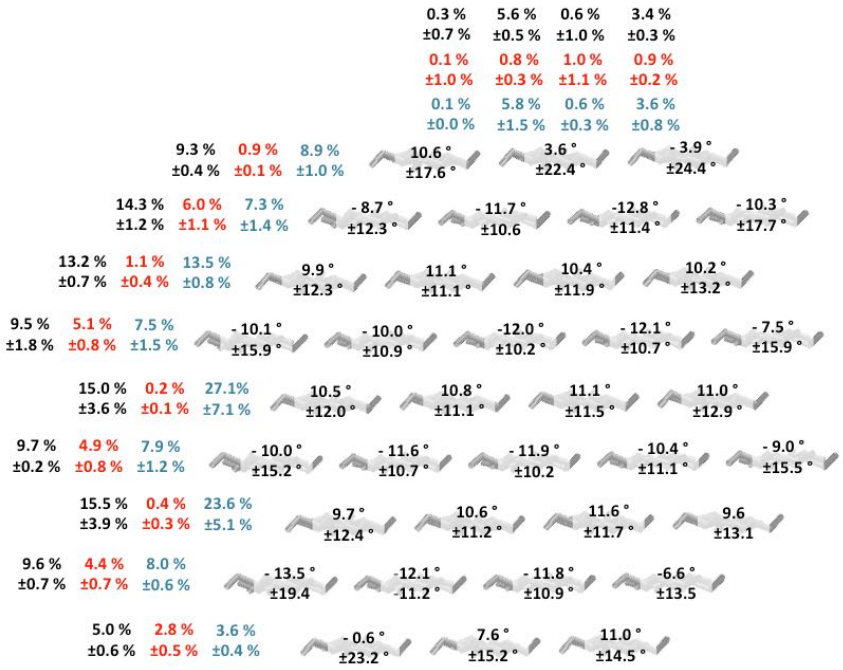



**Figure 4** Cellulose chain tilts and hydroxyl SASA contributions.

Cellulose chain tilt angles for every glucan chain averaged over the simulation are shown in degrees on the corresponding chain. Average contribution of hydrogens from the solvent-exposed hydroxyl group C2 (black, on the left/top), C3 (red, in the middle), and C6 (cyan, on the right/bottom) to the corresponding monomer's SASA are shown next to the respective surface glucan chain.

chains, with the same density on any two corresponding surfaces. Therefore, it can be presumed that similar interaction between cellulose and BmimCl is taking place at the analogous positions on corresponding surfaces, and considering only one surface from any corresponding surface pair in this analysis thus suffices to provide information on all crucial solute-solvent interactions. It should also be noted that the symmetrical distribution of solvent around the cellulose fiber indicates convergence of the MD simulation for the analysis of solute-solvent interactions.

Site-site radial distribution functions (RDFs or g(r)) are instrumental for identifying dominant short-range contacts or mid-range structural order in liquids. The RDFs presented here correspond to the probability of observing one site at a specific distance from another site. They are calculated for $Cl^-$ with respect to solvent-exposed hydrogen atoms from the C2-, C3-, and C6-hydroxyl groups (highlighted in Figure 1b), which are the sites of greatest $Cl^-$ accumulation. Figure 5a reveals that highest peaks, and thus the strongest interactions, are between the anions and the hydrophilic (110) surface, with a site-site separation of 2-3 Å. In contrast, the RDFs of the ring and methyl hydrogens of the cation with respect to the solvent-exposed oxygen atoms of the cellulose hydroxyl groups show broad profiles of low intensity (Figure 5b), suggesting that $Bmim^+$ does not accumulate around hydroxyl groups of the cellulose surfaces. However, this analysis of site-site RDFs does not exclude the possibility that the rather bulky cation may have a preferred way of approaching and aligning itself at cellulose surfaces. Indeed, g(r) profiles for the center of mass of the imidazole ring atoms with respect to the pyranose rings of the central glucan chains on different cellulose surfaces, reveal a minimum at around 7 Å, particularly for the $(1\bar{1}0)$ and the (010) surfaces (Figure 5c), indicating that a $Bmim^+$ solvation region exists around the cellulose surface.

Hence, the orientation and interaction energy of those $Bmim^+$ cations located within 7 Å of surface pyranose ring centers is analyzed with respect to the cellulose in order to reveal if any preferred arrangments of the molecules exist. In the remainder of this chapter, the orientations of $Bmim^+$ relative to the glucose units on the three different surfaces of the cellulose fiber are presented, together with RDFs for $Cl^-$ in greater detail.

## 4.1 Solvent Structure around the Hydrophobic (200) Surface

To assess the extent to which $Cl^-$ anions interact with the hydrophobic (200) surface, RDFs with respect to the hydrogens of the hydroxyl groups in the three surface chains are computed (Figure 6a, b). In order to focus on interactions specific to this surface, not all hydroxyls from the two border chains were considered as reference sites, as these may also be assigned to a neighboring surface. Hence, only those hydroxyls on the two border chains that point toward the middle chain on (200) were included, while on the middle chain all hydroxyls were considered. Moreover, RDFs with respect to hydroxyl groups pointing toward the neighboring (020) surface are combined in Figure 6a while those toward $(1\bar{1}0)$ are combined in Figure 6b. The reason for averaging these RDFs is that the individual profiles were found to be of comparable shape. As shown in these two plots, $Cl^-$ preferentially appears around the C2- and C6-hydroxyls pointing toward the $(1\bar{1}0)$ surface (dotted black and solid cyan profiles in Figure 6b). The g(r) profiles for the corresponding C3-hydroxyl (dashed red in Figure 6b) as well as for all hydroxyls extending toward the (020) surface (Figure 6a) show a much smaller peak. The reason for this dif-

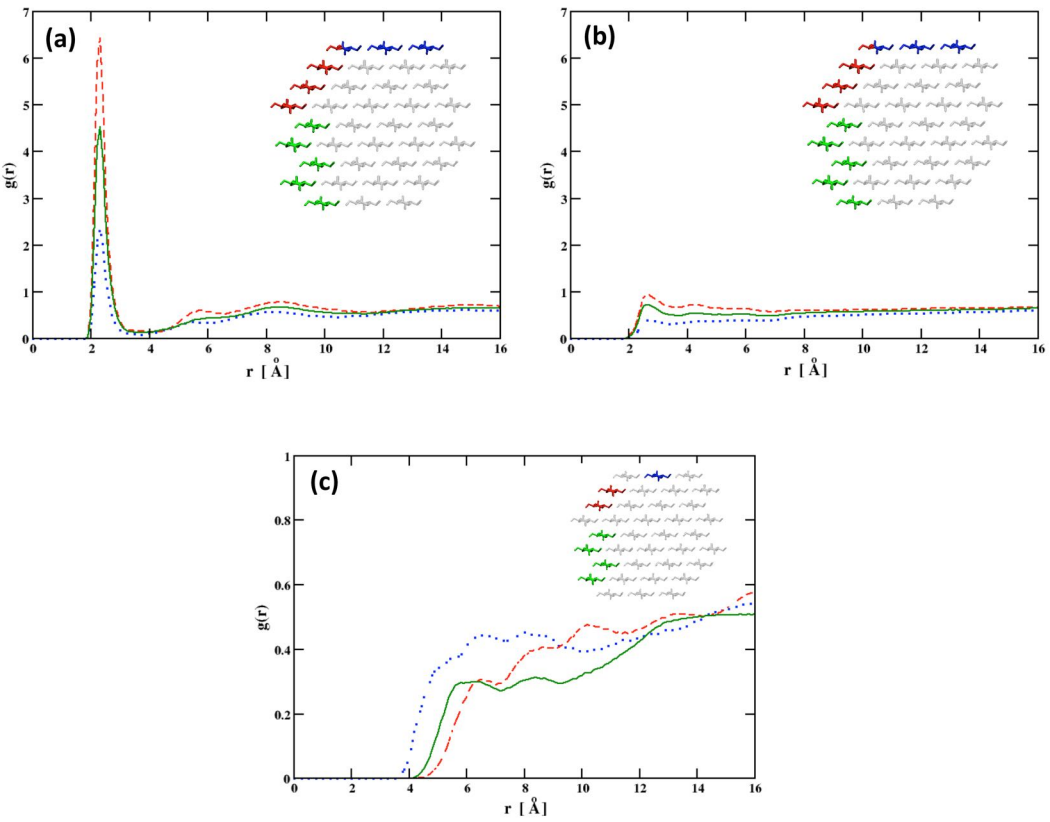



**Figure 5** RDFs of solvent ions around cellulose surface monomers.

(a) Site-site RDFs between $Cl^-$ anions and hydrogen atoms from the C2-, C3-, and C6-hydroxyls on the cellulose surface.
(b) Site-site RDFs between methyl- and ring-hydrogens from $Bmim^+$ cations and oxygen atoms from the C2-, C3-, and C6-hydroxyls on the cellulose surface.
(c) RDFs of the center of mass of imidazole rings with respect to the center of mass of pyranose rings. Only cellulose chains on a fiber surface center are taken into consideration in order to exclude border effects with neighboring surfaces.
In all profiles, RDFs are shown for the (200) (dotted blue), the $(1\bar{1}0)$ (dashed red), and the (010) surface (solid green), respectively.

ferential interaction pattern within one layer of cellulose chains lies in the fact that, during the course of the simulation, the center and origin chains tilt clockwise or counter-clockwise, respectively (see Figure 4). As a result, the (200) surface chains tend to expose hydroxyl groups on one side to the solvent while hiding the other side from the solvent. Nevertheless, even among the solvent-exposed C3-hydroxyls, the RDF shows a relatively lower peak (dashed red profile in Figure 6b) suggesting that the intrachain O3H'$\cdots$O5 hydrogen bonds (where the prime labels the residue toward the reducing end) remain largely intact. Indeed, the occupancy of these hydrogen bonds over the course of the entire simulation is only slightly lower than that in the fiber core (Figure 7).

The differential interaction of $Cl^-$ with C2- and C6-hydroxyls on the hydrophobic surface is also illustrated by computing three-dimensional spatial densities (Figure 8a-c). The contours in Figure 8 correspond to regions around a surface cellobiose unit where the average $Cl^-$ concentration during the given simulation is 5-10 times the $Cl^-$ concentration in bulk Bmimcl (green solid contours) or the average ring-hydrogen concentration is 2-3 times larger than its bulk concentration (orange meshed contours). Chains bordering on neighboring surfaces show solvent density not only on top but also on the respective sides of the glucan chains in these averaged distribution images.

The spatial distributions in Figure 8a-c also reveal that, unlike the $Cl^-$ anions, $Bmim^+$ cations do not have specific interactions with atoms on the (200) surface, as no ring hydrogen density peaks (meshed contours) are visible around any surface sites. On the other hand, the orientational analysis for $Bmim^+$ indicates that the cations do have a preferred alignment at this cellulose surface. The corresponding histogram (Figure 9a) shows several prevailing regions in orientation space which are clustered into six groups $(\alpha, \beta, \gamma, \delta, \epsilon, \eta)$.

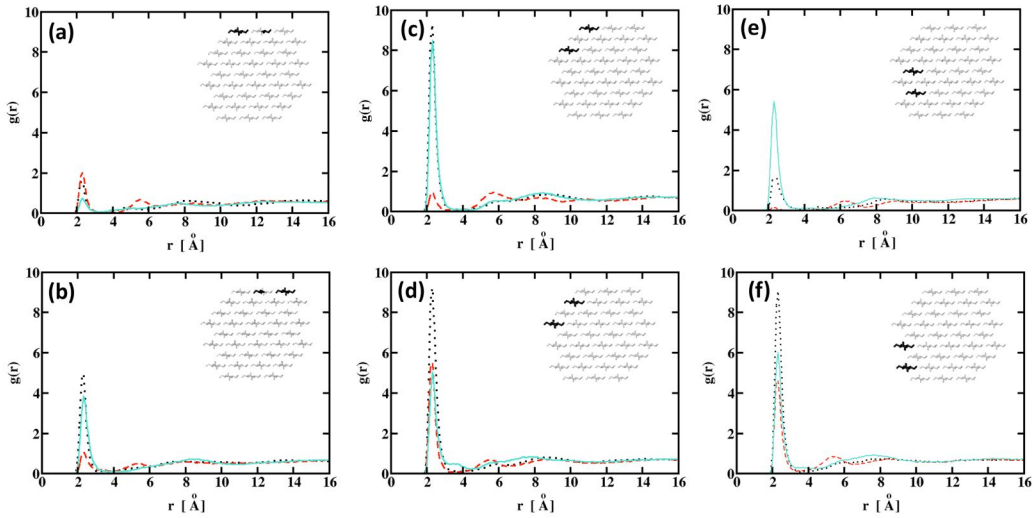



**Figure 6** RDFs of $Cl^-$ anions around cellulose surface hydroxyl groups.

Each profile describes a site-site RDF between $Cl^-$ and hydrogen atoms of the C2- (dotted black), C3- (dashed red), or C6-hydroxyl (solid cyan) on the highlighted gluan chains (inset). RDFs are shown with respect to (a) hydroxyl groups on the (200) surface showing toward the (020) surface, (b) hydroxyl groups on the (200) surface showing toward the $(1\bar{1}0)$ surface, (c) center chain hydroxyl groups on the $(1\bar{1}0)$ surface, (d) origin chain hydroxyl groups on the $(1\bar{1}0)$ surface, (e) center chain hydroxyl groups on the (010) surface, and (f) origin chain hydroxyl groups on the (010) surface.

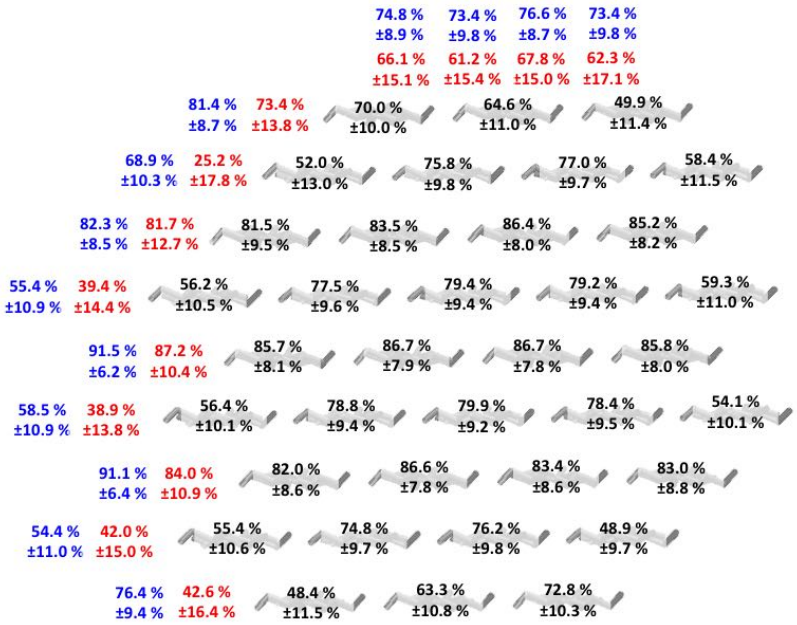



**Figure 7** Cellulose intrachain O3H'$\cdots$O5 hydrogen bond occupancies.

The percentage for only those hydrogen bonds on the (200), (1$\bar{1}$0), (010) surfaces, is given next to the corresponding glucan chain for the simulation in BmimCl (red, on the bottom/right) or in water (blue, on the top/left). In this study, a hydrogen bond is considered to be formed when the donor-acceptor distance is smaller than 3.5 Å and the hydrogen-donor-acceptor angle is smaller than 60°.

**Figure 8** Average spatial distributions of solvent with respect to a cellobiose unit.

Solvent-exposed C2- (black), C3- (red), and C6-hydroxyls (cyan) are highlighted. Contours correspond to 5-10 times the $Cl^-$ bulk concentration (green solid surfaces) and to 2-3 times the ring-hydrogen bulk concentration (orange meshed surfaces), respectively. Distributions are given for chains on (a-c) the (200), (d-f) the $(1\bar{1}0)$, and (g-j) the (010) surfaces.

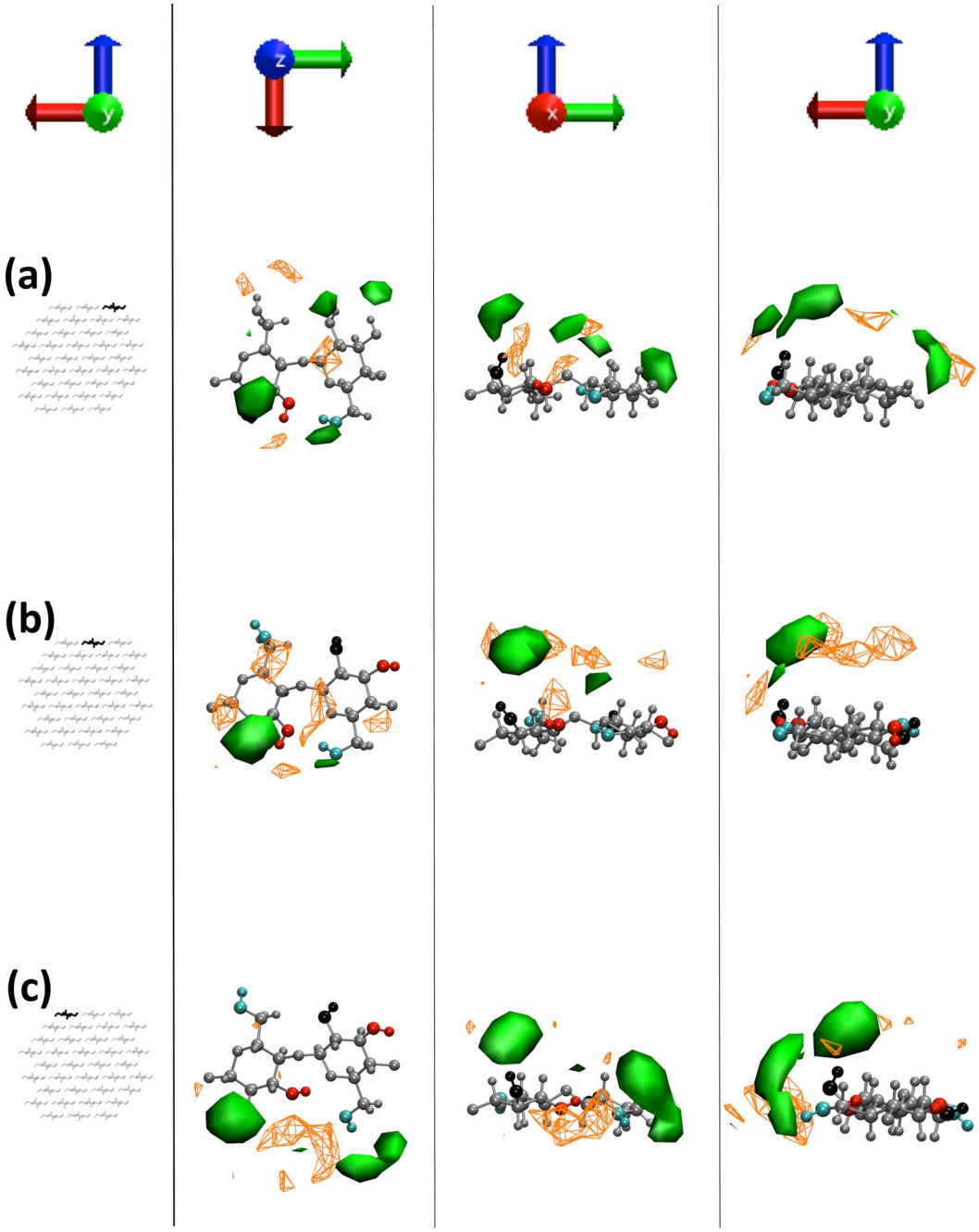



Figure 8, continued.

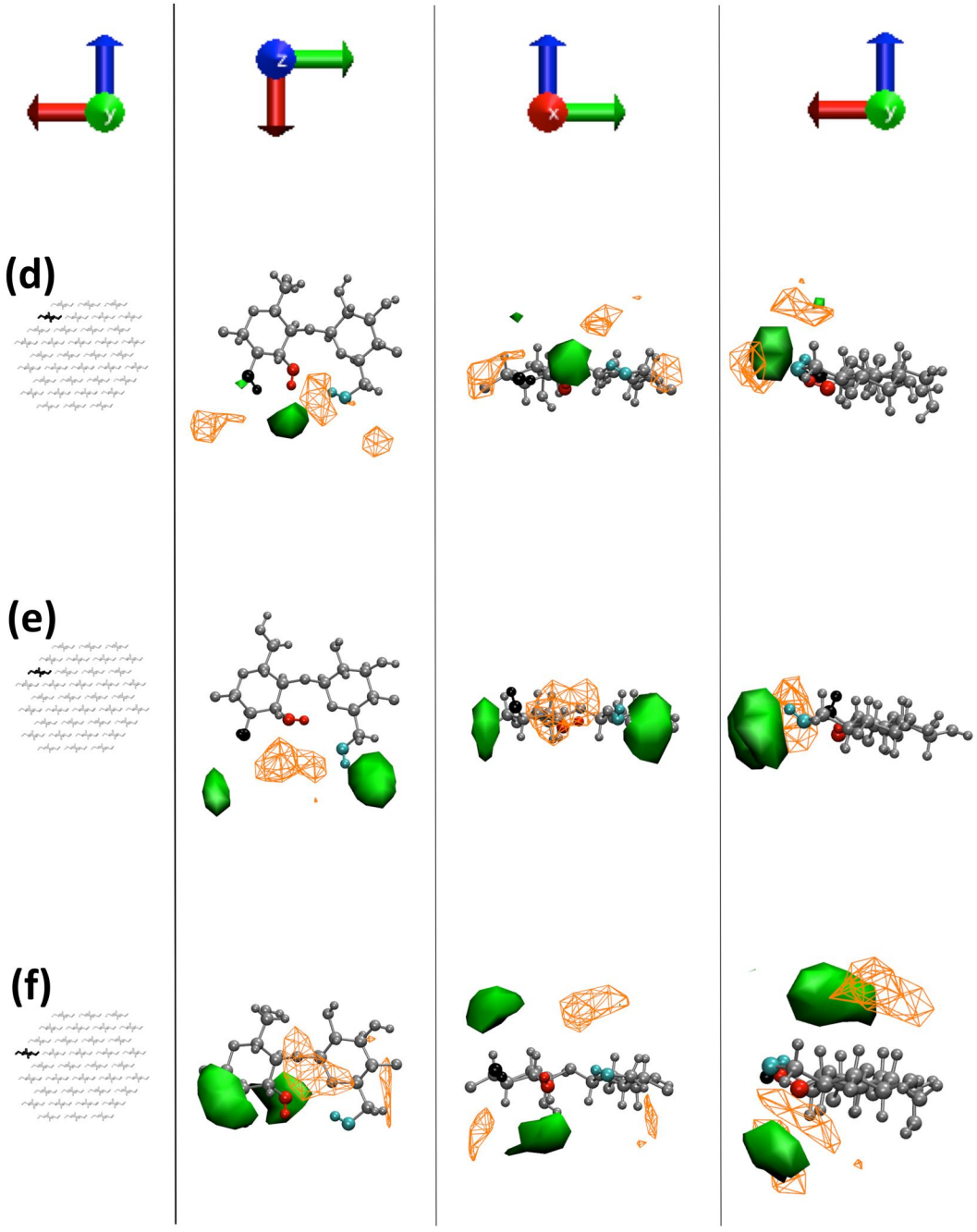



Figure 8, continued.

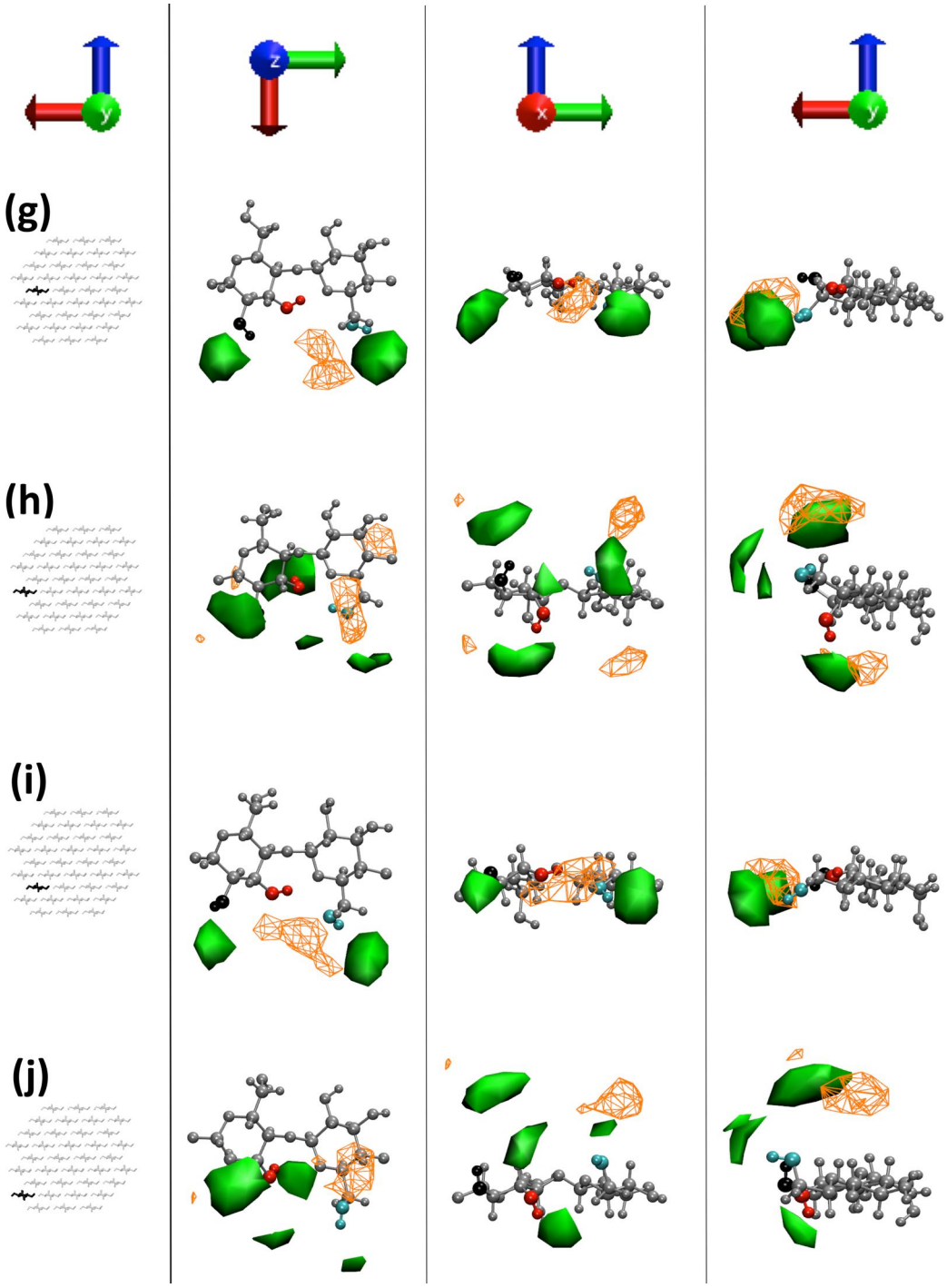



Figure 8, continued.

Peaks in this three-dimensional histogram correspond to specific orientations of the cation rings with respect to nearby surface pyranose rings as the three dimensions correspond to three rotation axes (see explanation in Appendix A)). Figure 9b visualizes the positions of all cation ring-centers from the six preferred orientations with respect to the nearest pyranose ring center in a top and a side view of the (200) surface. The bulk of these most prevalent orientations is found on top of the glucose rings, indicating that $Bmim^+$ avoids positioning in between neighboring strands or over the glycosidic bonds. Instead, the cation is mainly positioned over the pyranose rings in the favored orientations determined with respect to the glucose monomer.

In Figure 9c, the average arrangement and separation with reference to the closest glucose ring is visualized for structures from the two most populated groups of preferred orientations in Figure 9a. Averages of all six high-occupancy orientation clusters are given in Figure 10a. $Bmim^+$ cations are found to align themselves by stacking with the imidazole rings parallel to the closest glucose rings. The $Bmim^+$ tail and head groups do not have preferred directions, and this is also indicated by the observation that $\theta_z$ adopts several values for the highlighted groups while $\theta_x$ is either $\sim 0^\circ$ or $\sim 180^\circ$ and $\theta_y$ is $\sim 0^\circ$, signifying parallel arrangement of the imidazole and glucose ring planes. For all high-occurrence orientations, the separation of the imidazole ring center from the closest pyranose ring center is strictly between 3 and 4 Å. The above results indicate that cation rings stack against the hydrophobic (200) surface. To gain a quantitative understanding of this stacking interaction, the interaction potential energy between any $Bmim^+$ cation and the cellulose fiber is decomposed into Coulombic ($E_{Coul}$) and van der Waals ($E_{vdW}$) contributions (see chapter 3.1.1) for each of the preferred orientations which are shown in Figure 11a. On average, the computed $E_{vdW}$ of below -10 kcal/mol

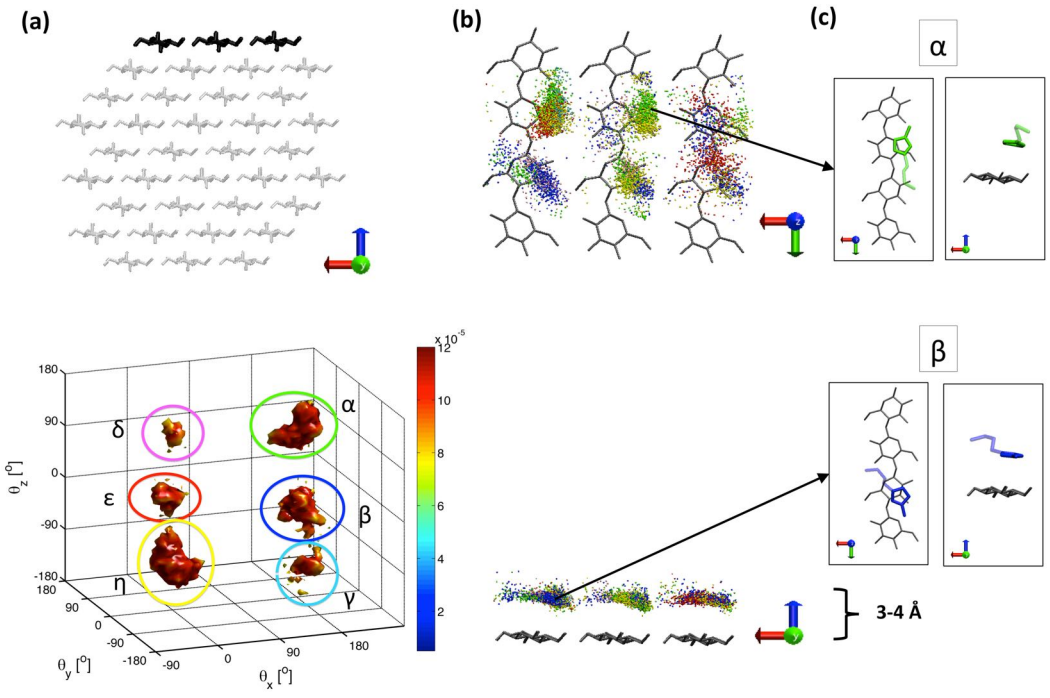



**Figure 9** Analysis of solvent cation orientations with respect to solute monomers on the (200) surface.

(a) Orientations are presented for cations around the cellulose (200) surface (top). Six peaks in the histogram are identified and shown as distinct regions in orientation space ($\alpha$, $\beta$, $\gamma$, $\delta$, $\epsilon$, $\eta$) (bottom).
(b) The positions of all cation ring-centers with respect to closest glucose monomers are visualized in a top view and a side view of the surface. Colors correspond to histogram peaks in (a).
(c) The average orientation and separation for the two most populated histogram peaks ($\alpha$ and $\beta$) is shown explicitly.

is clearly lower than $E_{Coul}$. This indicates that the stacking of cation rings at this interface involves mainly non-polar interaction between the cellulose and the $Bmim^+$ cations. This may compensate stacking interactions between cellulose layers, facilitating the transition from a compact crystal structure to detached cellulose chains. Indeed, X-ray diffraction studies on fibers from ramie cell walls treated with 1-ethyl-3-methylimidazolium acetate (EmimAc), an IL with a cation similar to BmimCl, have shown a reduced and broadened (200) reflection peak compared to untreated cellulose[112], interpreted as a reduction of cellulose stacking interactions due to IL treatment. In the current simulations, it is noticeable that both the $Cl^-$ anions and $Bmim^+$ cations are engaged in attacking the hydrophobic (200) surface in distinct, and yet complementary, ways.

## 4.2 Solvent Structure around the Hydrophilic $(1\bar{1}0)$ Surface

The fiber $(1\bar{1}0)$ face is referred to as hydrophilic because hydroxyl and hydroxymethyl groups on this surface are exposed to the solvent. Since individual RDF profiles of $Cl^-$ with respect to hydroxyl hydrogens on center or origin chains are of comparable shape, they are combined into one plot for the center chains (Figure 6c) and one for the origin chains (Figure 6d). These profiles show that $Cl^-$ avoids the C3-hydroxyl while accumulating around the C2- and C6-hydroxyl groups on center chains. For the origin chains, g(r) also peaks at $\sim$3 Å for $Cl^-$ around hydrogens on C3, indicating direct interaction. This is further confirmed by the much lower intrachain O3H'$\cdots$O5 hydrogen bond occupancy for solvent-exposed origin chains on this surface (Figure 7),

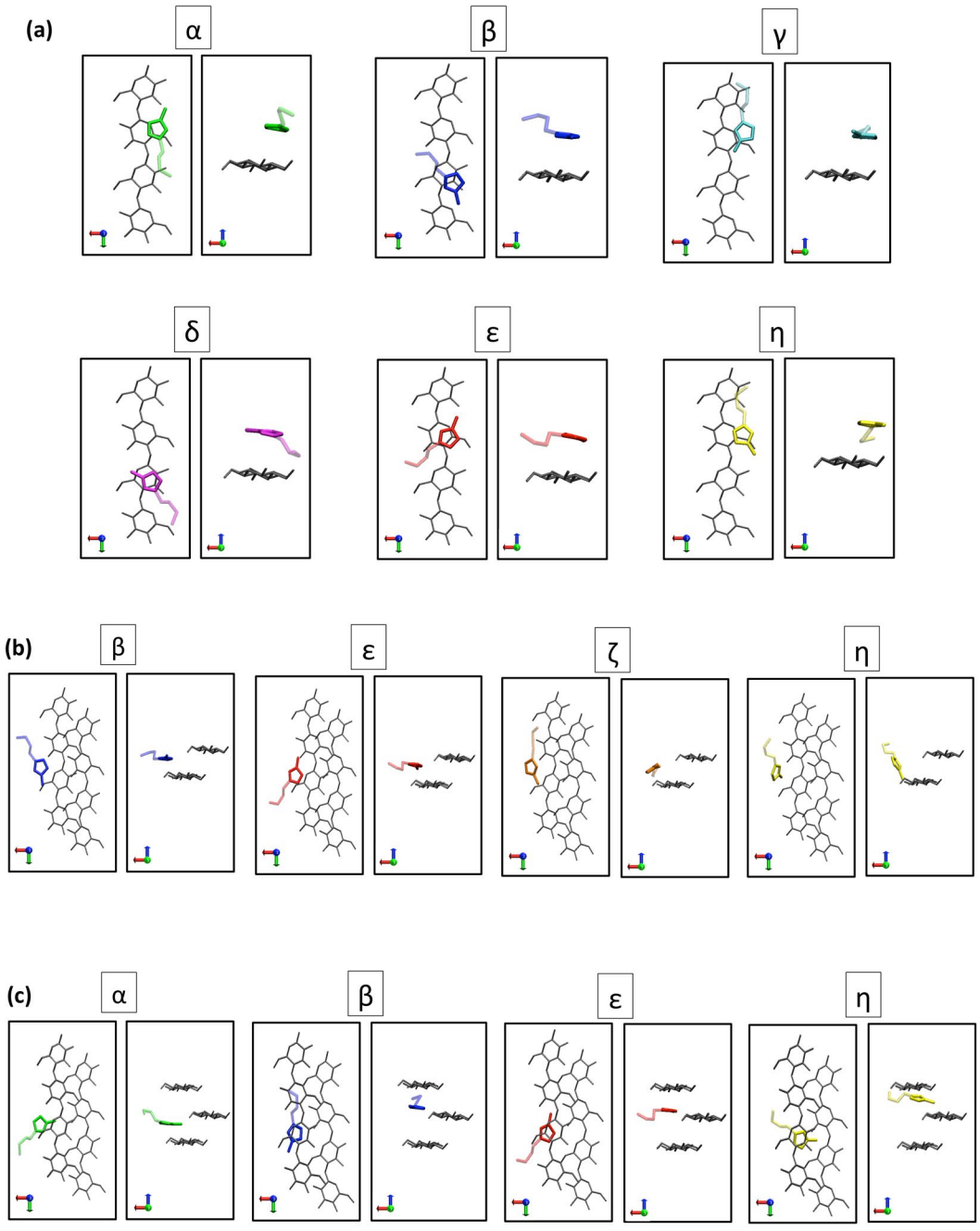



**Figure 10** Solvent cation orientations.

Average orientations and separations of solvent cations are given with respect to the closest glucose monomer for all histogram peaks for (a) the (200), (b) the $(1\bar{1}0)$, and (c) the (010) surface, respectively.

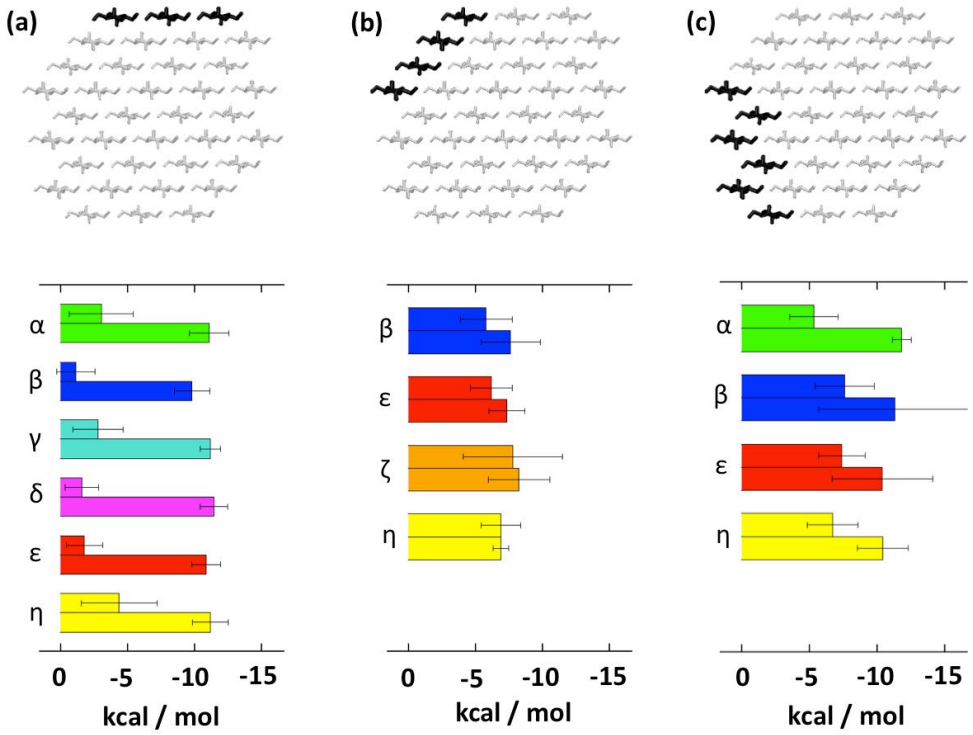



**Figure 11** Interaction potential energy components.

Decomposition of the interaction potential energy between solvent cations and cellulose into Coulombic (top bar) and van der Waals (bottom bar) energy; cations assigned to the histogram peaks in Figs. 9, 12, or 13 are taken into consideration for (a) the (200), (b) the $(1\bar{1}0)$ and (c) the (010) surface, respectively.

the mean percentage values being 25.2% and 39.4%, respectively, for the two origin chains while for the center chains the corresponding occupancy values reach ∼80%, comparable to those for the fiber core. Calculating the average solvent-accessible surface area (SASA) per monomer for surface hydroxyl groups leads to the same conclusion, i.e. that C3-hydroxyl groups on center chains are not available for direct interaction with the solvent. The SASA is obtained by an algorithm that uses a sphere of a particular radius to "probe" the surface of a molecule (the sphere radius for the current analysis was the radius of a $Cl^-$ anion, i.e. 1.81 Å)[182]. This measure is commonly used to quantify the extent of interaction between solute and solvent and it can also be decomposed into contributions from different parts of the solute molecule. As shown in Figure 4, hydrogen atoms from the C3-hydroxyl group contribute to the SASA of the center chains five times less than to the origin chains while the corresponding SASA values for the C2- or C6-hydroxyls remain fairly constant between the different chains. This analysis shows how accessible these surface hydroxyls can be to solvent molecules if not hydrogen bonded to the solute as in the case of the C3 group on center chains.

While in the initial crystal structure the two types of glucan chains differ only with respect to their chemical environment, the clockwise and counterclockwise tilting in the simulation (Figure 4) as well as the differential hydrogen bonding (Figure 7) result in characteristic hydrophilic surfaces with alternating center and origin chains exposing distinct sites for interaction with solute molecules. As a result, $Cl^-$ anions and $Bmim^+$ cations surrounding the solvent-exposed hydroxyl groups arrange differently on different chains of the $(1\bar{1}0)$ surface.

Spatial distributions give an averaged three-dimensional illustration of the interaction between anions and C3-hydroxyls on origin chains and the absence

of $Cl^-$ density contours at corresponding positions of center chains (Figure 8c-f). Interestingly, these images of average density also reveal that $Bmim^+$ cations concentrate along the cellulose strands of the $(1\bar{1}0)$ surface, albeit at lower probabilities than $Cl^-$. $Bmim^+$ tends to occupy the space on the cellulose surface in which $Cl^-$ is less accumulated, that is, in particular around the hydrogen-bonded C3-hydroxyl on the center chains (Figure 8c, e), and to some degree around the C2- and C6-hydroxyl groups on the origin chains (Figure 8d). Hence, anions and cations appear in an alternating fashion along the chains of the hydrophilic $(1\bar{1}0)$ surface, and in particular its center chains. The bottom origin chain on this surface shows density contours mainly beneath and above the pyranose ring plane (Figure 8f) due to its intermediate character as a border chain to the neighboring surface.

In Figure 12a, four relatively concentrated regions imply strong orientational preference for $Bmim^+$ on the $(1\bar{1}0)$ surface. These preferred orientations are mainly found in between neighboring cellulose strands on the $(1\bar{1}0)$ surface (Figure 12b, c). For the two most prevalent orientations (denoted $\beta$ and $\epsilon$), the average arrangement and separation from the nearest pyranose ring center show that the cation indeed intercalates between neighboring chains. Moreover, in both orientations, the imidazole rings are aligned in a similar way with respect to the fiber surface. The head or tail group of $Bmim^+$ points to opposite directions along the main fiber axis and the C1-H1 group of the imidazole ring points away while the C2-H2 and C3-H3 groups face the inter-chain space of the hydrophilic surface in both orientations. The two remaining cation orientations (see Figure 10b) are of much lower probability and may be considered as intermediate structures for a transition between orientations $\beta$ and $\epsilon$. Decomposition of the interaction potential energy between $Bmim^+$ and cellulose reveals comparable $E_{Coul}$ and $E_{vdW}$ values (Figure 11b). The average

Coulombic contribution is 4.2 kcal/mol lower and the average van der Waals energy is 3.3 kcal/mol higher than the corresponding interaction energies for the hydrophobic (200) surface. This indicates that cations can also interact with the hydrophilic surface of cellulose through favorable electrostatic interaction.

It should be noted that the intercalation of $Bmim^+$ cations is facilitated by the above-mentioned clockwise and counter-clockwise tilt of center and origin chains, respectively, which results in surface residues and functional groups moving against (when the origin layer is on top) or away from each other (when the center layer is on top). As a result, two grooves open on this surface that further help $Bmim^+$ to penetrate. This intercalation of $Bmim^+$ and the selective disruption of O3H'$\cdots$O5 hydrogen bonds on the origin chains by $Cl^-$ both indicate that cations and anions of the IL can interact with the hydrophilic $(1\bar{1}0)$ surface in distinct ways. These interactions may be primary steps in the dissolution of a cellulose fiber.

## 4.3 Solvent Structure around the Hydrophilic (010) Surface

The fiber (010) face contains voids which can be occupied by solvent molecules. The voids are defined by adjoining origin chains above and beneath as well as the side groups from the center chains. Consequently, this fiber surface does expose pyranose ring hydrogens to the solvent, but due to the greater exposure of hydroxyl groups, it is also referred to as hydrophilic. The bottom center chain, which borders on the hydrophobic $(\bar{2}00)$ face, is neglected in the following analysis of solvent distributions due to its intermediate character

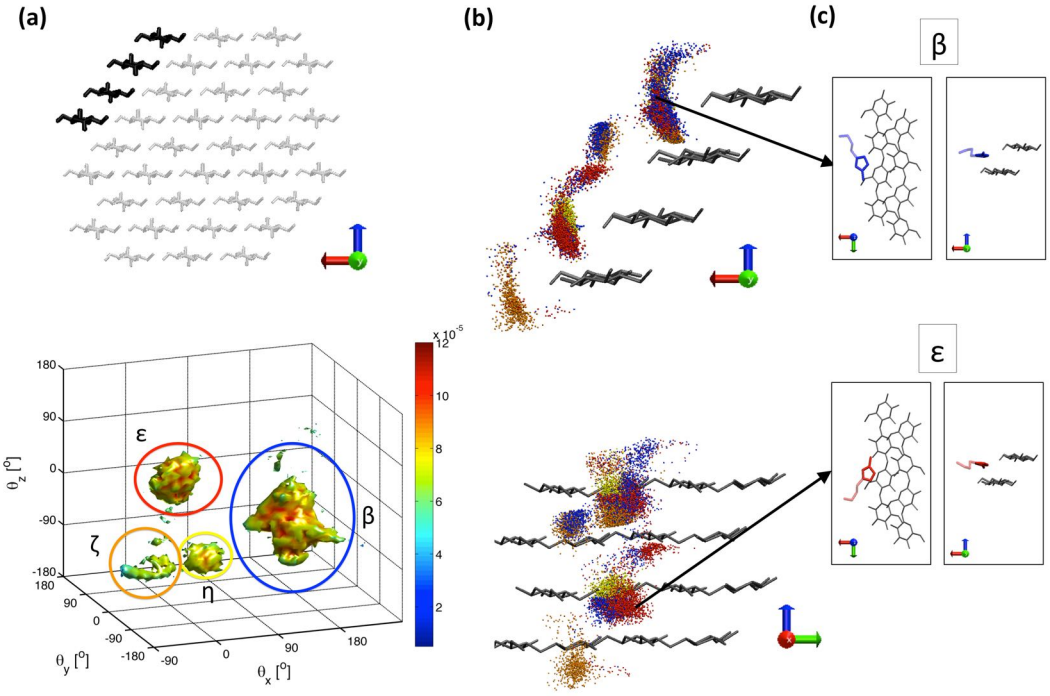



**Figure 12** Analysis of solvent cation orientations with respect to solute monomers on the $(1\bar{1}0)$ surface.

(a) Orientations are evaluated for cations around the cellulose $(1\bar{1}0)$ surface (top). Four peaks in the histogram are identified and shown as distinct regions in orientation space ($\beta$, $\epsilon$, $\zeta$, $\eta$) (bottom).
(b) The positions of all cation ring-centers with respect to closest glucose monomers are visualized in two different side views of the surface. Colors correspond to histogram peaks in (a).
(c) The average orientation and separation for the two most populated histogram peaks ($\beta$ and $\epsilon$) is shown explicitly.

as a border chain between two different surfaces. RDFs for $Cl^-$ are similar to the other hydrophilic surface. On the center chains, the anions avoid the C3-hydroxyl hydrogen and accumulate mainly around the C6-hydroxyl, while on the protruding origin chains the RDFs show strong peaks for $Cl^-$ around all hydroxyl groups (Figure 6e, f). The hydrogen on the C6-hydroxyl contributes significantly more to the SASAs of the center chains (on average 25.9%) than the C2- and C3-hydroxyls (15.3% and 0.3%) (Figure 4), which may be the reason for the higher peak of C6-hydroxyls in Figure 6e. Similar to the interactions between BmimCl and cellulose on the hydrophilic $(1\bar{1}0)$ surface, occupancies indicate that the binding of $Cl^-$ to C3-hydroxyl hydrogens disrupts the intrachain hydrogen bonding O3H'$\cdots$O5 to a greater extent on protruding origin chains than on buried center chains (Figure 7). Here as well, the anions and cations accumulate in an alternating way around side groups of center chains on this surface (Figure 8g-j), which is further enhanced by the presence of the surface voids.

Finally, analyzing the arrangement of $Bmim^+$ cations around the (010) surface reveals frequently occurring orientations, of which the two most preferred (orientations $\beta$ and $\epsilon$ in Figure 13a) are identical to those on the previously discussed hydrophilic $(1\bar{1}0)$ surface, which means that the C1-H1 group of the imidazole ring extends into the solvent. On the hydrophilic (010) surface, however, cations adopt their preferred orientations upon intercalating into the available voids (Figure 13b, c). The average interaction energies between the cellulose and $Bmim^+$ show comparable Coulombic interactions as on the other hydrophilic $(1\bar{1}0)$ surface and also comparable van derWaals interactions as on the hydrophobic (200) surface (Figure 11c). The stronger $E_{vdw}$ at this surface compared to the other hydrophilic surface is due to the interaction of cations with the non-polar pyranose rings above and beneath the surface voids.

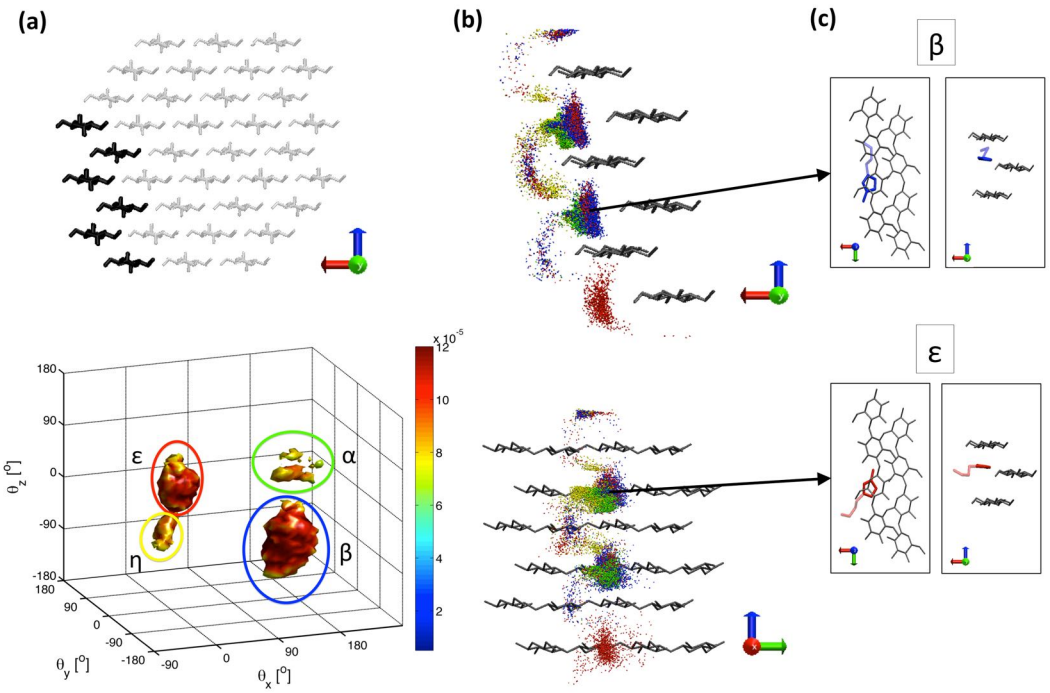



**Figure 13** Analysis of solvent cation orientations with respect to solute monomers on the (010) surface.

(a) Orientations are evaluated for cations around the cellulose (010) surface (top). Four peaks in the histogram are identified and shown as distinct regions in orientation space ($\alpha$, $\beta$, $\epsilon$, $\eta$) (bottom).
(b) The positions of all cation ring-centers with respect to closest glucose monomers are visualized in two different side views of the surface. Colors correspond to histogram peaks in (a).
(c) The average orientation and separation for the two most populated histogram peaks ($\beta$ and $\epsilon$) is shown explicitly.

The significantly higher probability of $Bmim^+$ cations with the C1-H1 group of imidazole rings pointing away from both hydrophilic surfaces examined suggests that these orientations may serve a distinct function. Recent spectroscopic analyses assign a greater acidity to the H1 proton on IL imidazole rings[183, 184] which means that this site can be considered a better hydrogen bond donor than the other imidazole protons. Therefore, it may seem surprising that the H1 proton points away from the solute in the preferred $Bmim^+$ orientations instead of hydrogen bonding to the hydrophilic cellulose surfaces where exposed hydroxyl oxygens can accept hydrogen bonds. One plausible reason for this may be that all imidazole ring hydrogens are relatively polar due to the delocalized electron system such that the combined interactions of hydrogens H2 and H3 with the hydrophilic fiber surfaces would outweigh the respective interactions of H1 with those surfaces. Alternatively, the observed orientations could also be stabilized through polar interactions of $Bmim^+$ ring hydrogens with $Cl^-$. As previously reported[143, 185–187], anions prefer to accumulate around the more polar H1 proton of imidazole. This fact is consistent with the presence of a sharp peak at a smaller distance in the RDF of $Cl^-$ around H1 compared to RDFs of $Cl^-$ around H2 or H3 in simulations of bulk BmimCl (data not shown). Coordination numbers ($c_n$) derived through the integration of the RDF profiles from zero up to the first minimum indicate the preferred interaction of $Cl^-$ with H1 ($c_n = 1.4$) rather than with H2 ($c_n = 1.2$) or with H3 ($c_n = 1.1$). Thus, the intercalation of cations observed in the fiber simulation, in which H1 preferentially faces the solvent while H2 and H3 are oriented toward cellulose, may arise from an interplay of the polar ring hydrogens interacting with anions on one side and surface hydroxyls on the other side.

In order to examine whether the affinity of polar imidazole ring hydrogens

to $Cl^-$ or to cellulose surface hydroxyls is the reason for the preferred $Bmim^+$ orientations on the hydrophilic surfaces, the average electrostatic contribution to the interaction potential between these hydrogens and the hydrophilic fiber surfaces or the solvent $Cl^-$ in the respective intercalating orientations ($\beta$ and $\epsilon$) is computed and compared to the corresponding values in the structures in which H1 faces the hydrophilic surfaces and H2 and H3 extend toward solvent (Figure 14). These opposite orientations are not visible as peaks in the orientation histograms of Figures 12 and 13 because they are sparsely populated (a few hundred structures were evaluated for each of the two surfaces) compared to the preferred orientations. The calculations reveal, for both of the preferred structures on hydrophilic $(1\bar{1}0)$ and (010) surfaces, the average Coulombic interactions between $Cl^-$ and H1 are greater than the sum of such interaction energies for H2 and H3 with the fiber surface. Also, for opposite intercalating orientations, the average interaction energies of H2 and H3 with $Cl^-$ are clearly higher than those between H1 and the respective fiber surface. However, the average sum of the Coulombic interaction energies for polar imidazole hydrogens and fiber surfaces or $Cl^-$ anions is comparable between structures taken from the preferred orientations and those taken from the opposite orientations (on average, between -16.2 and -19.2 kcal/mol for all six analyzed orientations from the two hydrophilic surfaces). Hence, it can be excluded that the preferred orientations observed on the hydrophilic surfaces arise from a higher affinity of the acidic imidazole H1 site for $Cl^-$ or the combination of H2 and H3 interactions with surface hydroxyl groups.

The dominant $Bmim^+$ orientation found on the hydrophilic surfaces can possibly also explain why the alkyl chain length has to be limited to 4-6 carbons for imidazolium-based ILs to be effective in cellulose dissolution[92, 120]. The length of the alkyl chain determines the amphiphilic character of the solvent.

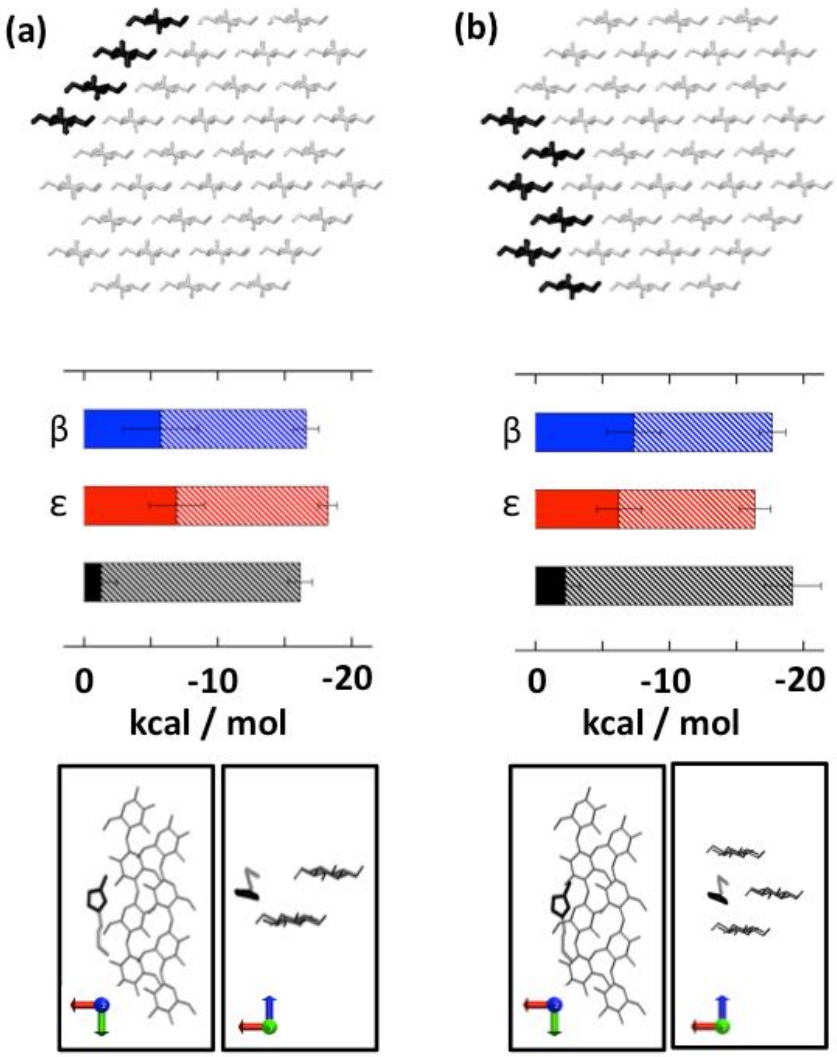



**Figure 14** Average Coulombic energy between polar imidazole ring hydrogens and cellulose or $Cl^-$ anions.

Cations assigned to histogram peak orientations $\beta$ (blue) and $\epsilon$ (red) as well as an opposite orientation (black) are taken into consideration for (a) the $(1\bar{1}0)$, and (b) the (010) surface, respectively. Interaction energies with cellulose are shown as solid bars, and those with $Cl^-$ are shown as shaded bars in addition to the corresponding solid bar. The bottom panel visualizes the average separation and orientation for structures from the set of opposite orientations.

The cation studied here contains a butyl tail and the present results suggest that the alkyl chain is restricted to roughly this number, due to structural reasons based on the observed preferred orientations at the fiber $(1\bar{1}0)$ and (010) faces in which the cation tail is positioned parallel to the glucan chain. Cations occurring in proximity to any cellulose surface have an average head to tail distance of ∼7.5 Å, which is comparable to the value for cations in bulk BmimCl. This value is smaller than the length of a cellobiose unit (∼10.5 Å between linker oxygens) and just smaller than the separation of neighboring C2- and C6-hydroxyls on a glucan chain (∼8 Å), which is the preferred site of $Bmim^+$ on center chains of hydrophilic surfaces as discussed above (see Figure 8). As a consequence, a longer alkyl tail would prevent cations from appearing next to neighboring cellobioses at the same time and would also interfere with $Cl^-$ anions at their cellulose interaction sites, hindering them from penetrating the fiber. The solvent structure around the fiber captured in this study, an alternating occurrence of cations and anions along the glucan chains on the hydrophilic surfaces (in particular the center chains) could not be realized with a longer cation alkyl group.

Inter- and intrachain hydrogen bonds are crucial for the structural rigidity of cellulose fibers. Analysis of the simulation suggests a disrupting effect of $Cl^-$ anions on solvent-exposed intrachain O3H'$\cdots$O5 hydrogen bonds of the fiber, occurring mainly on the hydrophilic surfaces. On origin chains in particular, the probability of hydrogen bond formation is clearly lower than on center chains, as anions interact with the relevant C3-hydroxyls on these chains. These results demonstrate the capability of ILs with a small anion to break cellulose hydrogen bonds and thus to initiate dissolution. In order to assess the effectiveness of $Cl^-$ in disrupting cellulose hydrogen bonds, a comparison of the intrachain hydrogen bond formation of cellulose solvated in

BmimCl with that of cellulose solvated in water would be desirable. To this end, a similar MD simulation of a cellulose fiber solvated in water is performed (see Appendix A)). This simulation is run for 100 ns at a temperature of 375 K, which corresponds to the boiling temperature of water at which the pretreatment of cellulosic biomass is usually performed. As shown in Figure 7, the percentage of intramolecular O3H'$\cdots$O5 hydrogen bonds on origin glucan chains is clearly higher in the water simulation (54%-68% on average) than in the BmimCl simulation (25%-42% on average) indicating that the IL is more effective in disrupting solvent-exposed cellulose hydrogen bonds than is water. It should be noted that the lower probability of formed hydrogen bonds in BmimCl may arise from the two simulations being performed at different temperatures, as the formation of hydrogen bonds in cellulose has been shown to be temperature-dependent[188, 189]. Even so, the present study shows the superiority of ILs over water as a pretreatment solvent for cellulose since they provide a liquid environment at elevated temperatures capable of interfering more severely with cellulose hydrogen bonds, which is not feasible for water at ambient pressure.

The analysis of $Bmim^+$ orientations with respect to different cellulose surfaces revealed two distinct patterns of interaction: stacking on cellulose strands on the hydrophobic surface and intercalation in between cellulose strands on hydrophilic surfaces. The analysis indicates how structural features of the solute and the solvent cation contribute to the favorable interactions with the two. Stacking of $Bmim^+$ on the hydrophobic fiber surface takes place as surface pyranose rings are approached by the planar aromatic imidazole ring. Intercalation of the planar cation in between cellulose chains is facilitated by a differential tilting of center and origin chains on the $(1\bar{1}0)$ surface and by voids on the (010) surface. The stacking interactions with the solvent

can facilitate cellulose dissolution by substituting the same kind of interaction between neighboring cellulose layers. It is, however, also conceivable that the intercalation could enhance the disintegration of cellulose fibers directly through interference with non-polar interactions of neighboring layers and indirectly by promoting the penetration of $Cl^-$ anions into the fiber interior where they can attack intrachain as well as interchain hydrogen bonds essential for fiber integrity. This may provide an explanation how ILs can facilitate cellulose dissolution through interactions on the hydrophilic and hydrophobic fiber surfaces.

The present simulation reveals how both components of BmimCl interact with the amphiphilic cellulose in the early stages of fiber dissolution. The distinct orientations $Bmim^+$ can adopt on the different surfaces highlight the versatility of imidazolium-based cations, which help to facilitate the disintegration of cellulose layers. Moreover, taking into consideration that $Cl^-$ anions interact extensively with solvent-exposed hydroxyl groups, it is conceivable that cellulose fiber deconstruction is achieved by BmimCl through the combination of polar anions penetrating the fiber and disrupting hydrogen bonds and nonpolar cations loosening layers of the cellulose structure via non-polar stacking interactions and intercalation. The present results thus suggest a synergistic approach to cellulose dissolution.

# 5

# Replica-Exchange Simulations of Cellulose Chains

To study the end state of cellulose dissolution in BmimCl and in water, REMD simulations of a fully solvated single chain are performed in both solvents. The REMD enhanced-sampling method allows more rapid convergence of the dynamics of the simulated systems and greater exploration of conformational variation than standard MD. As described in chapter 3.2, the overlap of replica potential energy distributions is a qualitative measure of the replica exchange acceptance. Figure 15 shows these distributions for all replicas in both REMD simulations. Distributions of replicas separated by a simulation temperature of 25 K (Figure 15a) or by 15 K (Figure 15b), respectively, are highlighted in color. These potential energy distributions have a Gaussian shape with a linearly decreasing mean and standard deviation upon increasing temperature (with the exception of the BmimCl simulations at low temperature). According to Kim et al.[190], a gap between two energy distributions would be characteristic of a phase transition region, which is clearly not the case in

Figure 15. A sufficient number of replica exchanges in either REMD simulation is indicated by the overlap of any two neighboring distributions.

In Figure 16 time series of replica exchanges are shown for the four temperatures 375 K, 400 K, 425 K, and 450 K. These graphs indicate a random walk in the replica space of both systems that is equivalent to a random walk in the corresponding temperature space. The intense exchange of neighboring replicas signifies the effectiveness of the REMD. The average replica exchange acceptance at any temperature of the REMD simulations is 15%-18%, which corresponds to an exchange event every 5-10 ps (see Appendix B) for more details).

Based on these results, it can be concluded that the REMD has been performed properly in both cellulose solutions, allowing the sampling of many different cellulose conformations. In the following, thermodynamic properties at temperatures relevant to the cellulose pretreatment process (e.g. T = 375 K in water or T = 450 K in BmimCl) are analyzed and the temperature dependence of these properties is also examined.

## 5.1 Cellulose Conformational Changes

Through the enhanced sampling, a large number of cellulose structures are explored in both BmimCl and water. A basic parameter that quantifies the corresponding overall size of a polymer is the radius of gyration, $R_g$. This parameter is often used in polymer physics to describe the dimensions of a polymer chain. In Figure 17, all sampled cellulose structures are characterized by $R_g$ calculated using the glycosidic bond oxygens (O4):

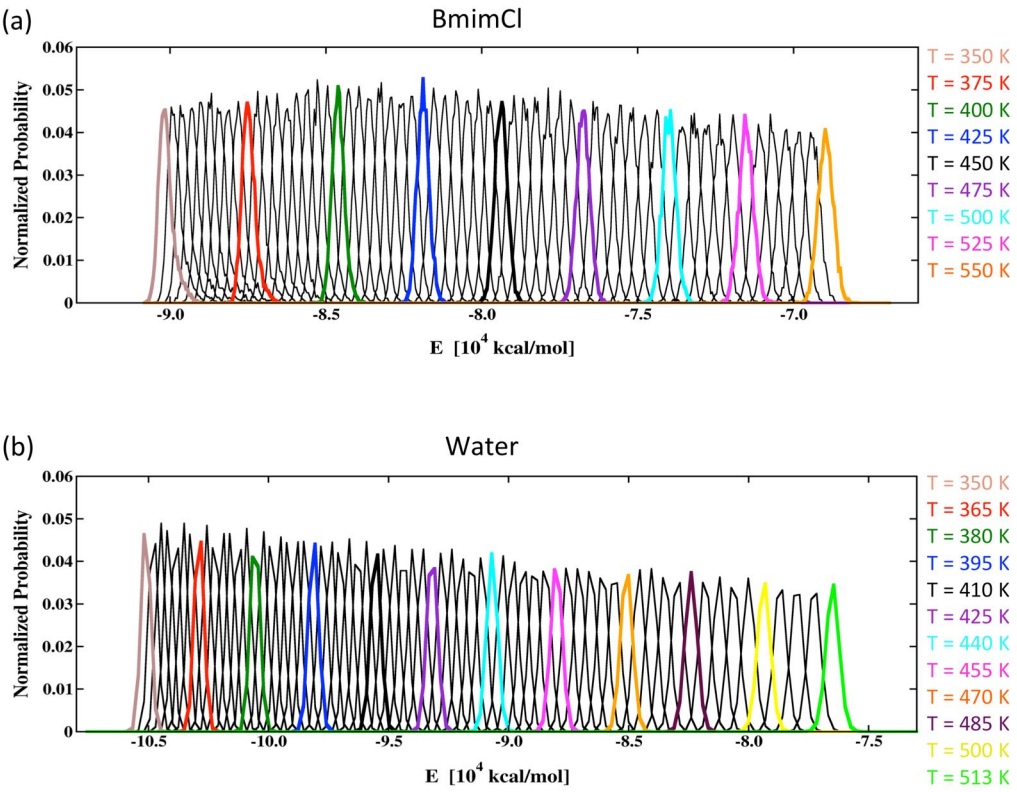



**Figure 15** Potential energy distributions of REMD simulations.

The distribution of potential energy at any REMD simulation temperature (a) in BmimCl and (b) in water. Colored distributions correspond to the simulation temperature indicated on the right, the black distributions correspond to all other temperatures.

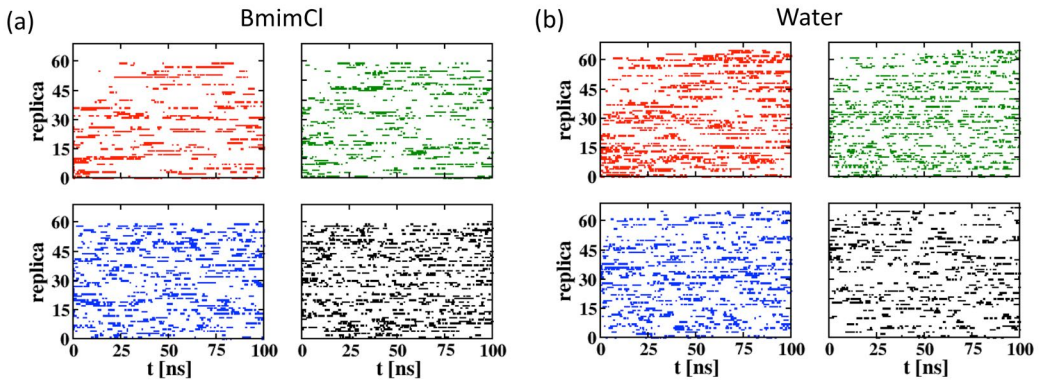



**Figure 16** REMD time series.

Time series of replica exchanges for the four temperatures 375 K (top left), 400 K (top right), 425 K (bottom left), and 450 K (bottom right) for REMD simulations (a) in BmimCl and (b) in water.

$$R_g = \sqrt{\frac{1}{N_g}\sum_{i}^{N_g}(r_i(O4) - r_{mean}(O4))^2} \tag{5.1}$$

where $r_{mean}(O4)$ is the average position of the oxygen atom and $N_g = 9$ is the number of glycosidic bonds.

The sampled cellulose conformations range from elongated decamer structures to collapsed configurations (see Figure 18b). A further parameter that describes the dimensions of a polymer is the distance between its two ends (or end-to-end distance, $d$). Temperature profiles based on $R_g$ (Figure 17a) or on $d$ (Figure 18a) both indicate that in both solvents the overall size of the cellulose decamer reduces with increasing temperature. The polysaccharide does not collapse to a compact structure with increasing temperature, but rather gradually reduces its stiffness and thus samples more conformations of smaller overall size at higher T. This behavior is qualitatively different from that in

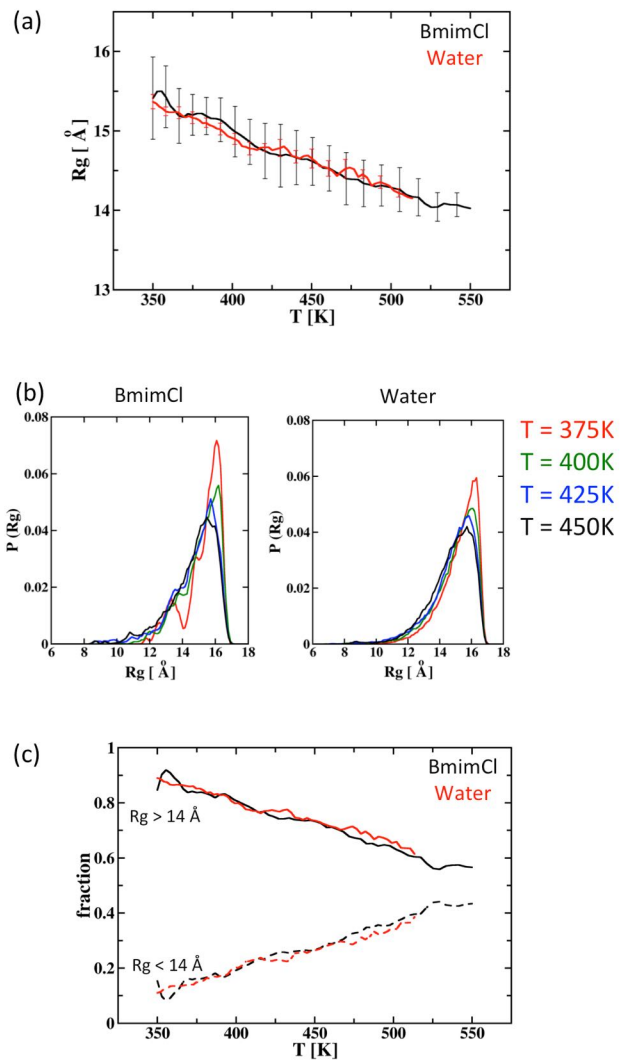



**Figure 17** Radius of gyration of cellulose in the two solvents.

(a) Temperature profile of the radius of gyration, $R_g$, calculated using the cellulose linker oxygens O4 in BmimCl (black) and in water (red).
(b) Distribution of $R_g$, P($R_g$), for cellulose structures in BmimCl (left) and in water (right) binned with $\Delta R_g = 0.1$ Å at the four temperatures T = 375 K, 400 K, 425 K, and 450 K. Colors correspond to the given temperatures (see far right). In both solvents about three quarters of all structures are extended with $R_g > 14$ Å.
(c) Fraction of extended ($R_g > 14$ Å solid lines) and compact ($R_g < 14$ Å dashed lines) structures for cellulose in BmimCl (black) and in water (red).

proteins, which denature and expand upon increasing temperature and fold into a compact native state structure at lower, physiological temperatures. Protein folding is hypothesized to occur with a hydrophobic collapse event, during which a hydrophobic solvent-separated protein core is formed[191]. The temperature at which this event happens can be measured by the change in a polymer's solvent-accessible surface area, or SASA as explained in chapter 4.2, or by its solvent-excluded volume, V, which is the volume defined by the solute's van der Waals surface that is not solvent-accessible[192, 193]. Negligible changes in SASA and in V of cellulose (Figure 19) both confirm that a sharp transition from an extended to a compact conformation does not occur.

The fact that the decrease of the radius of gyration or the end-to-end distance as a function of the temperature is comparable for both solvents means that this temperature-dependent conformational change is not driven by specific interactions of the solute with water molecules or with BmimCl ions. However, in BmimCl both profiles show greater variation at any temperature as indicated by the larger error bars, suggesting higher flexibility of cellulose in the IL than in water. Based on the value of $R_g$, any given cellulose decamer structure can be broadly classified as compact ($R_g < 14$ Å) or extended ($R_g > 14$ Å). According to this classification, in both solvents at any temperature the majority of all sampled cellulose structures is extended (Figure 17b). Interestingly, while ∼90% of cellulose structures are extended at the lowest simulation temperature, only ∼55% are extended at the highest temperatures (Figure 17c). This shows that the REMD approach broadly explores conformational space, sampling both compact and extended cellulose conformations at any temperature.

The cellulose decamer behaves like a semi-rigid rod polymer at lower temperatures, which is characteristic at length scales below a polymer's persistence

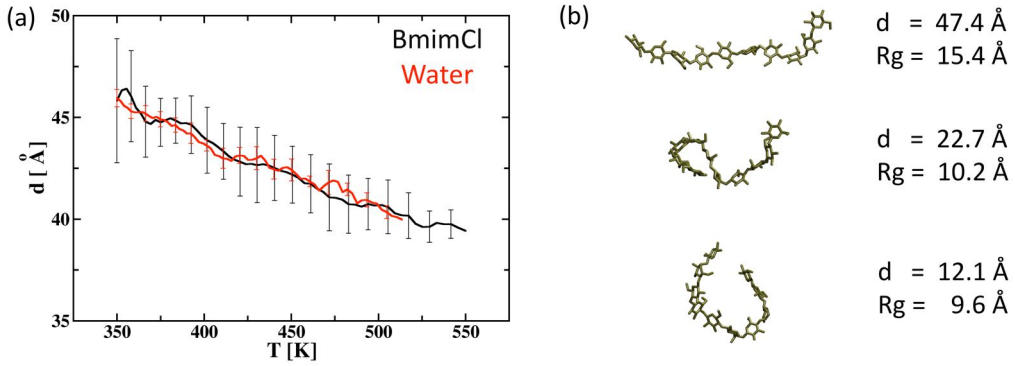



**Figure 18** End-to-end distance of cellulose in the two solvents.

(a) The average end-to-end distance of the cellulose decamer as a function of the temperature for REMD simulations in BmimCl (black) and in water (red). (b) Three representative cellulose structures with their corresponding end-to-end distance and radius of gyration.

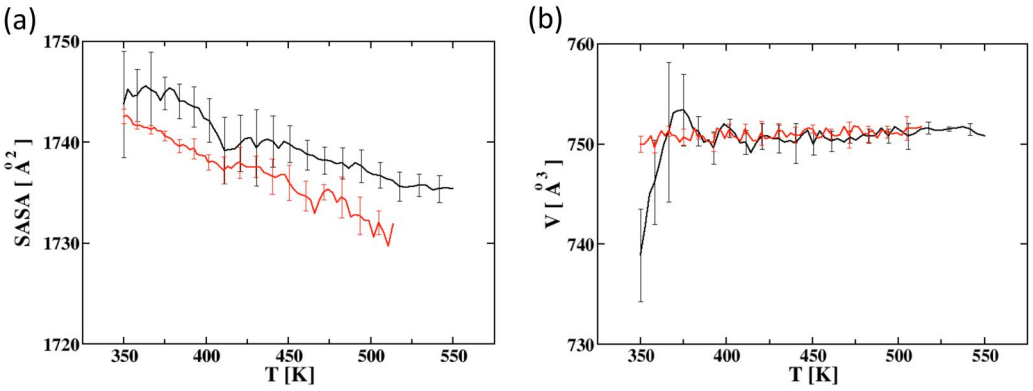



**Figure 19** Solvent-accessible surface area and solvent-excluded volume of cellulose in the two solvents.

(a) Solvent-accessible surface area of the cellulose decamer as a function of the temperature for REMD simulations in BmimCl (black) and in water (red). (b) Solvent-excluded volume of the cellulose decamer as a function of the temperature for REMD simulations in BmimCl (black) and in water (red).

length, $L_P$[194]. $L_P$ is a property quantifying the stiffness of a polymer. The value of $L_P$ of cellulose in aqueous solution at room temperature is reported to be between ∼100 Å (based on a study of polymer bending rigidity[195]) and ∼150 Å (based on a study of relative bond orientations[196]). Both values are obtained from the extrapolation of small cellulose fragment simulations in water. Experiments on cellulose in different solutions result in the same order of magnitude for $L_P$[197–199]. These values for $L_P$ at room temperature correspond to a cellulose chain with more than twenty glucose monomers, which is longer than the decamer chain under study with the contour length $L_C = 52$ Å. With the given REMD simulations, however, $L_P$ can be estimated at elevated temperatures. $L_P$ is derived from the decrease of correlation between glucose units (given by the cosine of the angle between them) as a function of separation. With this approach, in the present REMD simulations $L_P$ for the decamer is indeed estimated to be greater than 10 glucose units at T < 400 K but decreases to 5 glucose units (∼26 Å) at T = 500 K in both solvents (data not shown). This analysis indicates that the cellulose polymer is less rigid and thus assumes more compact structures at elevated temperatures than at lower temperatures.

Except for the greater flexibility of the solute in BmimCl, the present results on overall structural parameters such as the radius of gyration and the persistence length do not reveal a significant difference between the two solvents. However, in addition to these parameters, the rotamer conformations of the hydroxymethyl groups and the glycosidic bonds as well as the ability to form intramolecular hydrogen bonds or to maintain the six-membered ring of its glucose units in the favorable chair conformation are investigated.

The hydroxymethyl group of a glucose unit can adopt stable staggered rotamers, referred to as GG, GT, and TG corresponding to the trans and gauche

conformations of the O5-C5-C6-O6 and C4-C5-C6-O6 torsions. The angle $\omega$ is defined as the former torsion. The distributions of $\omega$ at distinct temperatures for simulations in BmimCl and in water are given in Figure 20a. The TG conformation is populated at less than 1%. The GG and GT populations differ significantly between the BmimCl and water solutions. On average, $\omega$ adopts the GG conformation in around 50% of the cellulose configurations in water at the lowest temperature (T = 350 K) and in around 40% at the highest temperature (T $\approx$ 513 K). Accordingly, the population values for the GT conformation range from about 50% to about 60% of cellulose structures with increasing simulation temperature in water. In BmimCl, however, the corresponding values for the GG and GT populations change from around 30% and 70% (at T = 350 K) to around 20% and 80% (T = 550 K), respectively, depending on the temperature. This result is similar to that found in previous simulations of cellulose oligomers of different lengths solvated in water or the IL 1-ethyl-3-methylimidazolium acetate (EmimAc)[121]. Together, these results suggest that the preference of the GT over the GG conformation of the hydroxymethyl groups is higher in cellulose-dissolving ILs, such as EmimAc or BmimCl, than in water. Thus, the ratio P(GG)/P(GT) may serve as an order parameter that distinguishes between the different structures the cellulose chain can adopt in the two solvents. The temperature dependence of this ratio is visualized as an average per glucose monomer in Figure 21a. P(GG)/P(GT) is higher for cellulose in water (0.7-0.8) than in BmimCl (0.2-0.5). In the cellulose-water simulations, the value of P(GG)/P(GT) decreases slightly with increasing T for all the hydroxymethyl groups except for the terminal group at the nonreducing end, which lacks a neighboring monomer on one side. In contrast, in the cellulose-BmimCl simulations P(GG)/P(GT) generally increases with T. It is worth noting that the corresponding distribu-

tions of this ratio also follow the same trends if only compact ($R_g < 14$ Å) or extended ($R_g > 14$ Å) cellulose chains are considered (data not shown).

In crystalline cellulose, intramolecular hydrogen bonds are formed between O3H'$\cdots$O5, O3H'$\cdots$O6, O2H$\cdots$O6', and O6H'$\cdots$O2 of neighboring monomers (where the prime labels the residue toward the reducing end)[46]. The average occupancies of these hydrogen bonds as a function of simulation temperature are given in Figure 20b. The only hydrogen bond that is formed to any significant degree is O3H'$\cdots$O5: all other hydrogen bond occupancies are less than 1%. With increasing temperature, the average O3H'$\cdots$O5 occupancy, P(O3H'$\cdots$O5), decreases from around 20% to 15% in BmimCl, while in water this value decreases from around 38% to 30%. Temperature profiles for individual O3H'$\cdots$O5 hydrogen bond occupancies along the cellulose chain are shown in Figure 21b. With the exception of the cellulose-BmimCl simulations at low temperatures (350-400 K), P(O3H'$\cdots$O5) assumes similar values for individual hydrogen bonds at any given temperature in either solvent.

The above results reveal clear differences between cellulose chains in BmimCl and in water in terms of the hydroxymethyl rotamer conformation and intramolecular hydrogen bonding. However, these two properties may be coupled to each other, as the hydroxymethyl conformation of a glucose residue could depend on the hydrogen bond state of its O5 atom. Thus, it is desirable to know how the conformational state of a hydroxymethyl rotamer changes with the adjacent O3H'$\cdots$O5 hydrogen bond state in both solvents. It is found that in more than 75% of all cases in which the O3H' hydrogen atom is bound to a solvent water oxygen or $Cl^-$ anion, and not to the O5 atom of the neighboring residue, the adjacent hydroxymethyl group assumes the GT conformation. This indicates that the intramolecular O3H'$\cdots$O5 hydrogen bond is coupled to the hydroxymethyl conformation. In fact, the difference

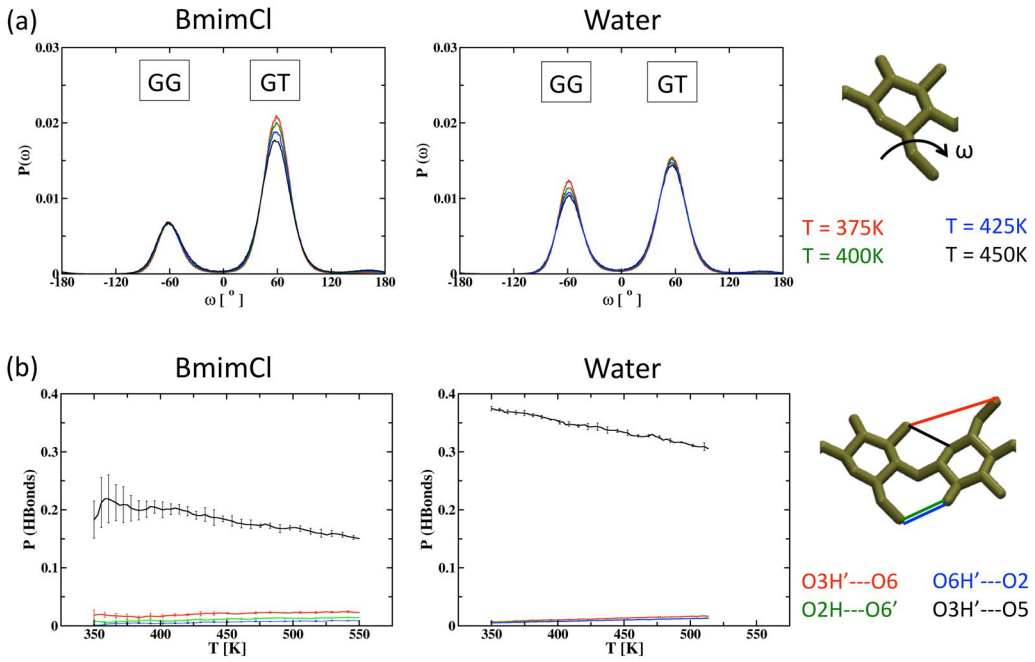



**Figure 20** Cellulose hydroxymethyl and intramolecular hydrogen bond profiles.

(a) Probability distributions of the cellulose hydroxymethyl angle $\omega$ (displayed on the far right) in BmimCl (left) and in water (right) for the four temperatures T= 375 K, 400 K, 425 K, and 450 K. The two preferred conformations around $\omega = -60°$ and $\omega = 60°$ are referred to as GG and GT, respectively. Colors correspond to the given temperatures (see far right).
(b) Temperature profiles of occupancies for cellulose intramolecular hydrogen bonds O3H'$\cdots$O6, O2H$\cdots$O6, O6H'$\cdots$O2, and O3H'$\cdots$O5 (displayed on the far right) between any two neighboring glucose units in BmimCl (left) and in water (right). Only the O3H'$\cdots$O5 hydrogen bond is formed to a significant degree. Colors correspond to the given hydrogen bonds (see far right).

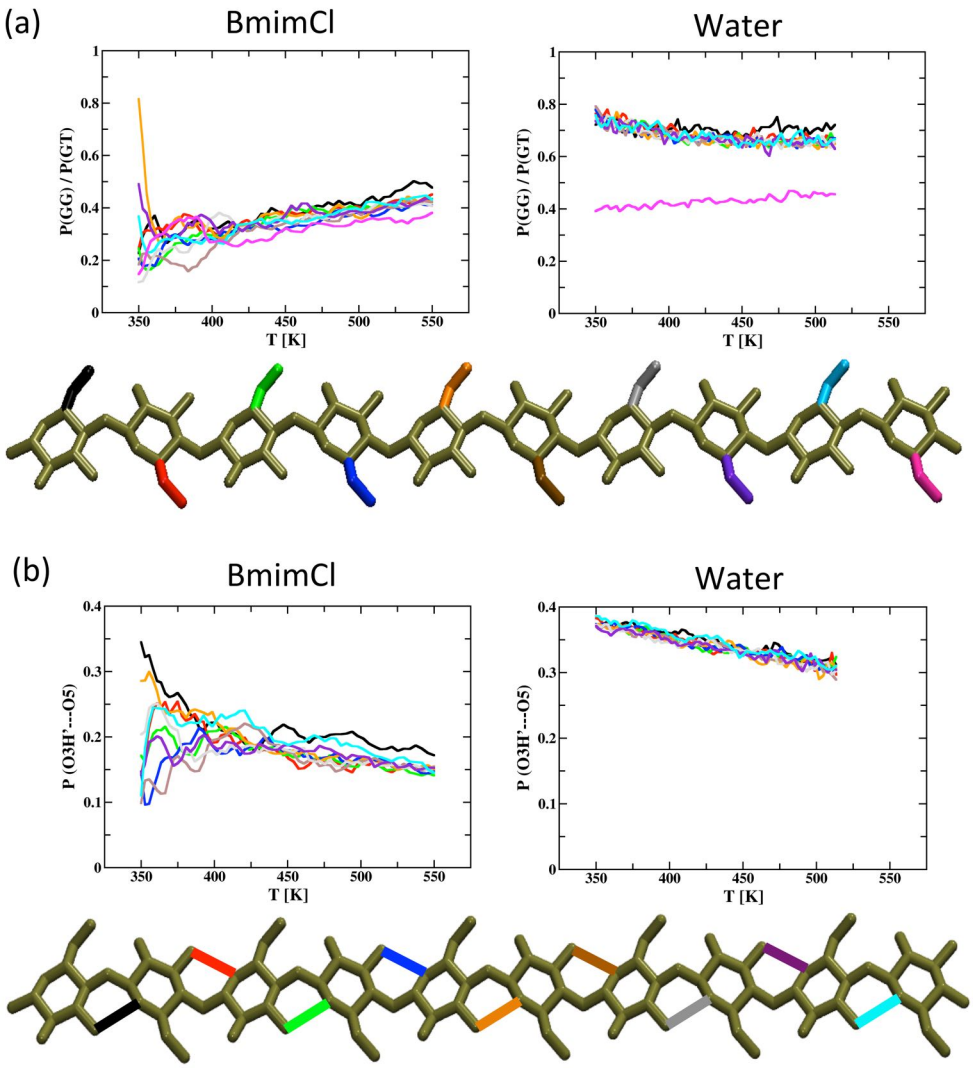



**Figure 21** Cellulose temperature profiles of P(GG)/P(GT) and P(O3H'$\cdots$O5).

Profiles for (a) the ratio of $\omega$ populations P(GG)/P(GT) and (b) the intramolecular O3H'$\cdots$O5 hydrogen bond occupancies P(O3H'$\cdots$O5) are given as a function of the temperature T. Colors correspond to the hydroxymethyl group or hydrogen bond given below the corresponding graph.

in $\omega$ populations between the two solvents, with $\omega$ assuming the GT conformation more frequently in BmimCl than in water, as discussed above, may be indirectly controlled by the greater hydrogen-bond basicity of $Cl^-$ anions, that is their potential to make stronger hydrogen bonds, than water oxygens.

In Figure 22, the correlation of the above two order parameters, P(GG)/P(GT) and P(O3H'$\cdots$O5), is visualized in two-dimensional free energy landscapes based on population histograms for simulations at the corresponding cellulose pretreatment temperatures, i.e. at T = 375 K in water and at T = 450 K in BmimCl. For these and all following free energy landscapes, it should be mentioned that a value of 0.1% corresponds to over 100 structures and that the standard deviation in the free energy basins is 0.01%-0.02%. The most populated region of the free energy landscape in Figure 22 is split into two basins along the ratio P(GG)/P(GT). The two basins in the BmimCl simulations ($IL_{1/2}$) contain 23%-28% (depending on the simulation temperature) of the sampled cellulose structures while only about 15% fall in the basins of the corresponding water simulations ($W_{1/2}$). However, the two energy basins cannot be distinguished from each other based on global features of the cellulose structures they contain. For instance, both extended and compact structures are found in either basin and individual structures have comparable average potential energies and average solute-solvent interaction energies. Hence, cellulose structures with similar global properties can differ in local structural properties such as the order parameters P(GG)/P(GT) and P(O3H'$\cdots$O5), the latter possibly leading to their different dissolution abilities.

The two torsion angles around the cellulose glycosidic bond, $\phi$ for C4'-O4'-C1-O5 and $\psi$ for C5'-C4'-O4'-C1, define the backbone structure of the polysaccharide. Two-dimensional scatter plots of $\phi$ and $\psi$ reveal three distinct regions of stable conformations, which have previously been referred to as

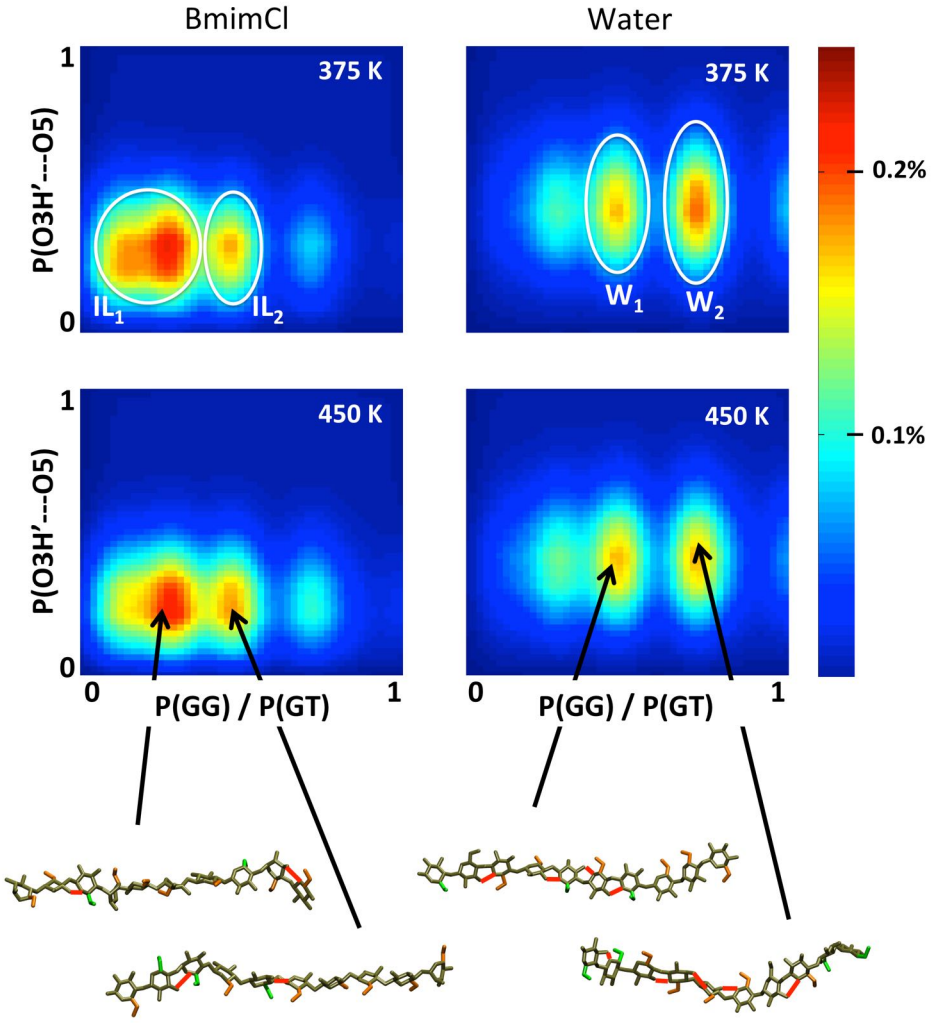



**Figure 22** Free energy landscapes for P(GG)/P(GT) and P(O3H'$\cdots$O5).

Free energy landscape for the order parameters P(GG)/P(GT) and P(O3H'$\cdots$O5) for simulations in BmimCl (left) and in water (right) at T = 375 K (top) and at T = 450 K (bottom). The two distinct basins in each plot are referred to as $IL_{1/2}$ and $W_{1/2}$. Representative structures for each basin are given on the bottom (O3H'$\cdots$O5 hydrogen bonds are indicated in red; hydroxymethyl conformations are indicated in green (GG) and in orange (GT)).

basins O, A, and B (see Figure 23a)[195]. The populations of these different basins are similar for any individual glycosidic bond of the cellulose decamer: more than 80% can be attributed to basin O, less than 20% to basin B, and a negligible amount to basin A.

Comparison of $\phi$-$\psi$ plots reveals differences of populations in basins O and B between cellulose in BmimCl and water (Figure 23a). For example, a small population occurs in basin B of the BmimCl simulations, but not in that of the water simulations at $\phi \approx 90°$ and $\psi \approx 120°$. This additional $\phi$-$\psi$ conformation is mainly assumed by glycosidic bonds at the cellulose chain ends, which are more flexible in BmimCl than in water. In basin O, difference plots at T = 375 K and at T = 450, respectively, highlight two distinct sub-regions which are preferably populated in water ("water basin", blue) or in BmimCl ("BmimCl basin", red). Representative structures from these two areas of the $\phi$-$\psi$ plot are also illustrated in Figure 23b. Regardless of the simulation temperature and solvent environment, the average O3H'$\cdots$O5 separation is 4.2 Å for any cellobiose unit that falls in the BmimCl basin but only 2.5 Å for those in the water basin. The corresponding average hydrogen bond occupancies are less than 1% in the BmimCl basin but around 40% in the water basin, once more showing the destructive effect of the IL on the intramolecular O3H'$\cdots$O5 hydrogen bond. This result indicates a correlation between this hydrogen bond and the backbone conformation. A similar population difference in basin O of the $\phi$-$\psi$ plot between water and IL solvated cellulose oligomers has been reported previously[121]. These authors found three distinct basins in basin O based on much shorter MD simulations ($<$ 5 ns).

The glycosidic bond torsions $\phi$ and $\psi$ describe the orientation of two neighboring glucose units and so can be referred to as internal coordinates for the cellulose backbone. Hence, the root-mean-square deviation (RMSD) in $\phi$-$\psi$ space

**Figure 23** $\phi$-$\psi$ scatter plots.

(a) Two-dimensional scatter plots for the glycosidic bond torsions $\phi$ and $\psi$ binned with $(\Delta\phi,\ \Delta\psi) = (1^\circ,\ 1^\circ)$ for simulations in BmimCl (left) and in water (right) at the two temperatures 375 K (top) and 450 K (bottom). The three different basins O, A, and B are indicated in the top left scatter plot.
(b) $\phi$-$\psi$ difference maps (BmimCl - Water) for the most populated basin O ($-180^\circ < \phi/\psi < 0^\circ$) at both temperatures. Two example cellobiose structures characterize the difference between structures from the BmimCl basin (red) and those from the Water basin (blue).

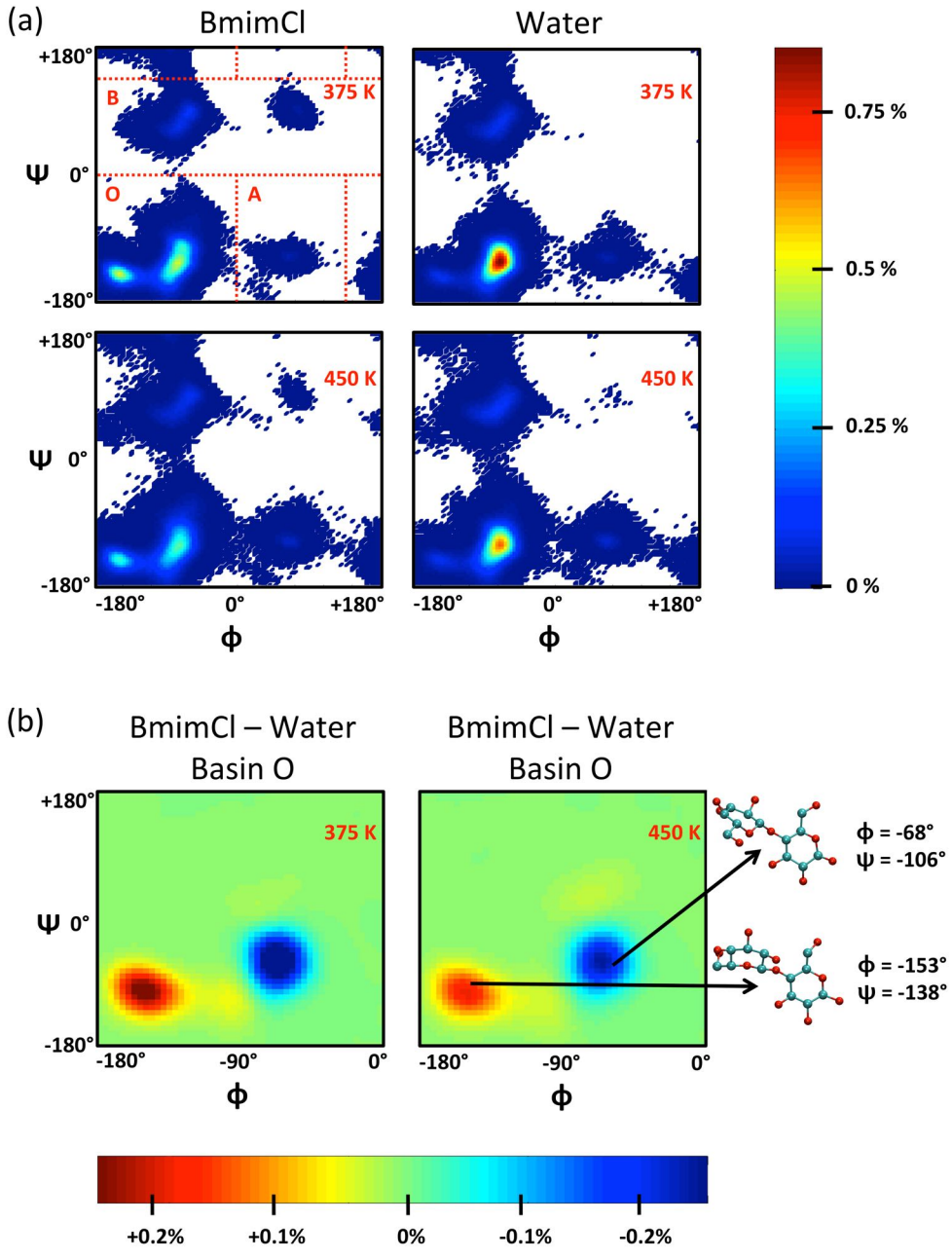



Figure 23, continued.

is a metric sensitive to structural changes of the cellulose. Chain-averaged $\phi$-$\psi$ RMSDs are defined as:

$$RMSD(\phi,\psi) = \sqrt{\frac{1}{N_g}\sum_{i=1}^{N_g}(\phi_i(t)-\phi_i(t_0))^2 + (\psi_i(t)-\psi_i(t_0))^2} \qquad (5.2)$$

where the reference structure at $t_0$ is the REMD starting structure after the initial NVT-equilibration with $\phi_i(t_0) = 83°$ and $\psi_i(t_0) = 32°$ for any glycosidic bond $i$ and $N_g = 9$ is the number of glycosidic bonds.

Figure 24a shows free energy landscapes using RMSD($\phi,\psi$) and P(O3H'$\cdots$O5) as the order parameters for T = 375 K and for T = 450 K. In all landscapes, there is a clear separation between a basin at RMSD($\phi,\psi$) $\approx 45°$ ($IL_1$ or $W_1$) and another at RMSD($\phi,\psi$) $\approx 90°$ ($IL_2$ or $W_2$). Extended cellulose decamers ($R_g > 14$ Å) populate the basins $IL_1/W_1$ for the most part, while compact structures ($R_g < 14$ Å) generally fall into $IL_2/W_2$, as the distribution of basin-specific $R_g$ values indicates (Figure 24b). Thus, consistent with the previous results on the temperature dependence of cellulose chain length (see Figure 17), the basins $IL_1/W_1$ depopulate with increasing temperature as more compact structures are sampled. Therefore, the chain-averaged RMSD of $\phi$-$\psi$ is shown to be a suitable parameter for identifying distributions of compact and extended cellulose structures.

Another conformational characteristic concerns the pyranose rings, which can adopt several conformations referred to as chair, boat, skew, half-chair, and envelope[200]. While the most stable ring conformation of D-glucopyranose is the $^4C_1$-chair[201], other conformations are adopted by ring flipping or puckering. The $^4C_1$ population, P($^4C_1$), is derived by calculating the angles of the three improper torsions C4-O5-C2-C1, O5-C2-C4-C3, and C2-C4-O5-C5. Fol-

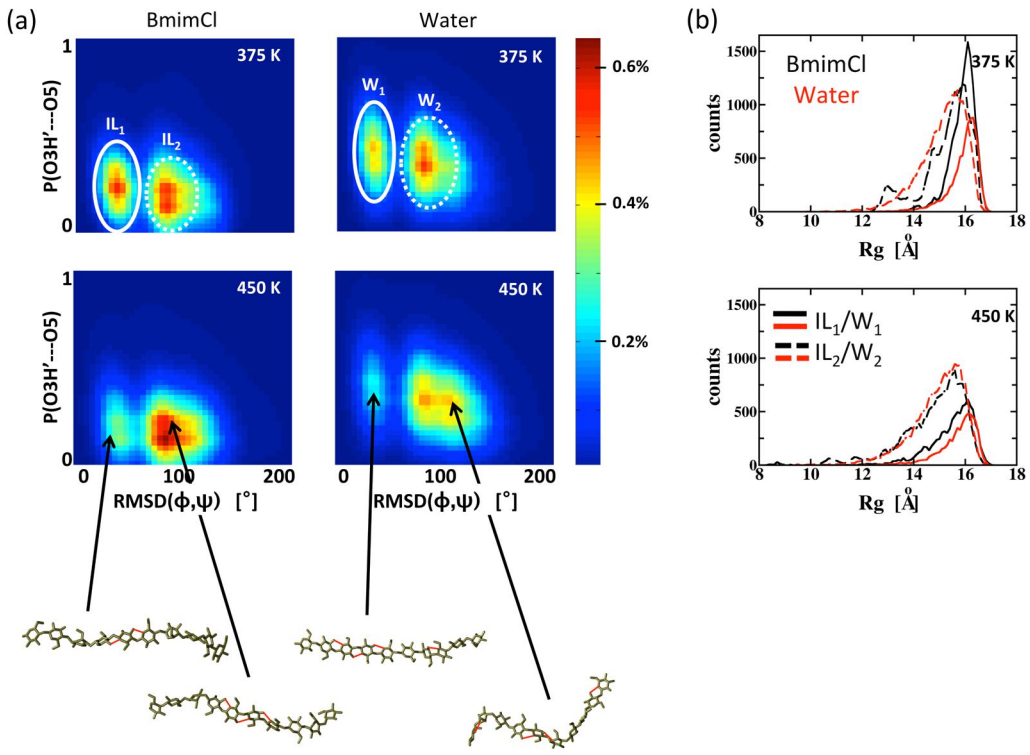



**Figure 24** Free energy landscapes for RMSD($\phi, \psi$) and P(O3H'$\cdots$O5).

(a) Free energy landscape for the order parameters RMSD($\phi, \psi$) and P(O3H'$\cdots$O5) for simulations in BmimCl (left) and in water (right) at T = 375 K (top) and at T= 450 K (bottom). The two distinct basins in each plot are referred to as $IL_{1/2}$ and $W_{1/2}$. Representative structures for each basin are given on the bottom (O3H'$\cdots$O5 hydrogen bonds are indicated in red). (b) The total count of cellulose structures in BmimCl (black) and in water (red) binned with $\Delta R_g = 1$ Å for the four basins $IL_1$ (solid black), $IL_2$ (dashed black), $W_1$ (solid red), and $W_2$ (dashed red) at T = 375 K (top) and at T = 450 K (bottom).

lowing Ref. [202], the $^4C_1$ conformation is characterized by a value of 35° for these improper torsions (a tolerance value of +/- 15° is applied in this analysis). The temperature dependence of P($^4C_1$) is shown in Figure 25a. The deviation of the pyranose ring conformation from the stable $^4C_1$ is increased at elevated temperatures. Clearly, the ring flipping occurs more often in BmimCl than in water. This is in accord with other results presented in this chapter supporting the fact that the cellulose chain is more flexible in BmimCl than in water.

The REMD simulations are performed on a decamer as the study of longer cellulose chains would have required larger simulation systems and a much greater number of replicas. However, since cellulose solubility is independent of chain length for DP $>$ 8[31], the present results derived from simulations of cellulose 10-mers should be applicable to longer chains. Nevertheless, it is useful to determine whether the current results hold for a larger cellulose chain. To this end, additional simulations of cellulose 20-mers are performed at three different temperatures around the corresponding pretreatment temperatures in water (T = 350 K, 375 K, 400 K) and in BmimCl (T = 400 K, 450 K, 500 K). While overall structural parameters, such as the radius of gyration or the solvent-accessible surface area, are about two times greater for the 20-mer compared to the 10-mer at any of the given temperatures, all the other cellulose conformations analyzed in this study have comparable relative populations as shown in Table 1. This validates the fact that meaningful conclusions for the cellulose dissolution abilities of the two solvents can be drawn from the results based on simulations of a decamer.

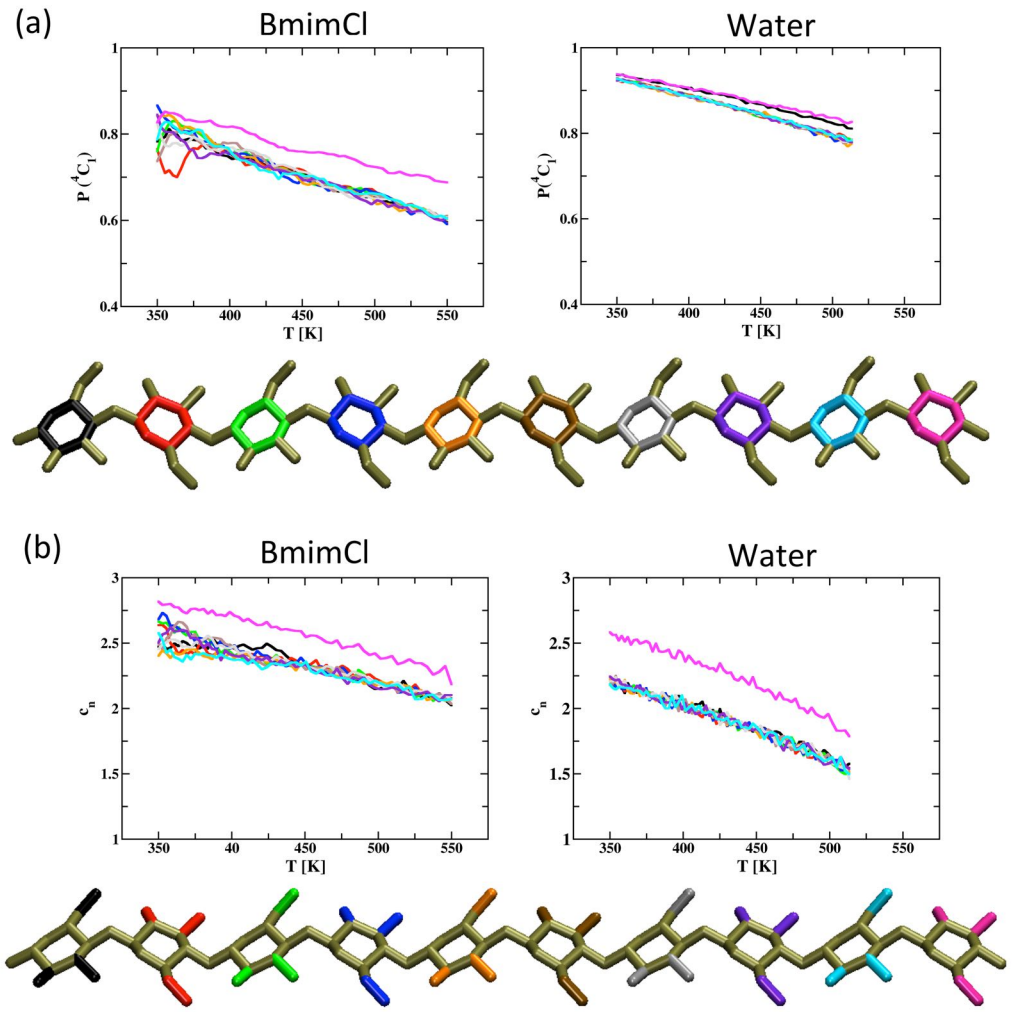



**Figure 25** Cellulose temperature profiles of P($^4$C$_1$) and c$_n$.

Profiles for (a) the percentage of $^4$C$_1$-chair conformation and (b) the average Cl$^-$ anion or water oxygen coordination numbers c$_n$ around cellulose hydroxyl groups are given as a function of the temperature T. Colors correspond to the pyranose ring or hydroxyl groups given below the corresponding graph.

**Table 1** Comparison of cellulose conformations between REMD simulations (top row) and additional MD simulations (bottom row) in water (at 350 K, 375 K, and 400 K) and BmimCl (at 400 K, 450 K, and 500 K). Percentages are given for the pair of glycosidic bond angles, ($\phi$, $\psi$), to fall into basin O, A, or B, the hydroxymethyl angle, $\omega$, to adopt GG or GT conformations, the backbone hydrogen bond O3H'$\cdots$O5 to be formed, and the pyranose ring to be in $^4C_1$-chair conformation.

| | Water Simulation | | | BmimCl Simulation | | |
|---|---|---|---|---|---|---|
| Temperature | 350 K | 375 K | 400 K | 400 K | 450 K | 500 K |
| P(($\phi$, $\psi$) $\epsilon$ O) [%] | 90.7 ± 0.9 | 88.8 ± 0.7 | 86.2 ± 1.0 | 88.2 ± 4.4 | 86.6 ± 2.4 | 83.4 ± 1.4 |
| | 91.3 ± 1.4 | 88.7 ± 0.9 | 81.8 ± 1.7 | 91.5 ± 5.3 | 86.9 ± 4.2 | 81.0 ± 3.6 |
| P(($\phi$, $\psi$) $\epsilon$ A) [%] | 0.9 ± 0.0 | 1.2 ± 0.3 | 1.8 ± 0.4 | 1.9 ± 1.3 | 2.3 ± 0.7 | 2.7 ± 0.7 |
| | 1.1 ± 0.0 | 1.3 ± 0.7 | 1.4 ± 0.7 | 0.8 ± 0.5 | 1.9 ± 0.9 | 3.2 ± 1.1 |
| P(($\phi$, $\psi$) $\epsilon$ B) [%] | 8.4 ± 1.0 | 10.0 ± 0.6 | 12.0 ± 0.9 | 9.9 ± 3.9 | 11.2 ± 2.4 | 13.9 ± 1.5 |
| | 7.5 ± 1.3 | 10.0 ± 0.4 | 16.8 ± 1.3 | 7.7 ± 4.9 | 11.2 ± 3.6 | 15.8 ± 2.8 |
| P($\omega$ = GG) [%] | 41.2 ± 0.7 | 40.0 ± 0.2 | 39.1 ± 0.2 | 22.9 ± 0.8 | 25.6 ± 0.9 | 26.7 ± 0.2 |
| | 41.8 ± 0.9 | 40.2 ± 0.4 | 37.9 ± 0.8 | 21.8 ± 2.1 | 21.8 ± 3.0 | 26.7 ± 0.9 |
| P($\omega$ = GT) [%] | 58.2 ± 0.7 | 59.2 ± 0.2 | 60.0 ± 0.2 | 75.8 ± 0.9 | 72.0 ± 1.0 | 70.0 ± 0.2 |
| | 57.7 ± 0.9 | 59.0 ± 0.4 | 61.1 ± 0.8 | 76.8 ± 2.5 | 76.8 ± 3.0 | 70.0 ± 0.8 |
| P(O3H'$\cdots$O5) [%] | 37.7 ± 0.4 | 36.7 ± 0.4 | 34.5 ± 0.4 | 20.3 ± 2.0 | 18.2 ± 1.4 | 17.0 ± 1.3 |
| | 37.6 ± 0.9 | 36.7 ± 0.7 | 34.5 ± 0.7 | 18.8 ± 4.1 | 18.2 ± 4.0 | 15.5 ± 3.0 |
| P($^4C_1$) [%] | 92.8 ± 0.5 | 91.0 ± 0.7 | 89.1 ± 0.8 | 76.8 ± 1.9 | 70.8 ± 2.0 | 66.2 ± 2.3 |
| | 92.8 ± 0.8 | 90.6 ± 0.9 | 88.1 ± 0.9 | 76.6 ± 2.3 | 70.3 ± 2.9 | 66.1 ± 2.2 |

## 5.2 Cellulose-Solvent Interactions

A good cellulose solvent must make favorable interactions with both the polar and the non-polar parts of the amphiphilic solute. As shown in chapter 4 and in previous studies[124], imidazolium-based ILs are capable of breaking hydrogen bonds and separating chains in large cellulose aggregates. It is desirable to know to what extent these interactions take place in the dissolved state of an individual cellulose chain. To this end, the binding of $Cl^-$ anions and water molecules to the polar cellulose hydroxyl groups C2, C3, and C6 is first investigated. These sites form hydrogen bonds with water oxygens (average separation of 1.8 Å) or with $Cl^-$ anions (average separation of 2.4 Å). Average coordination numbers of a hydroxyl group by water oxygens or $Cl^-$ anions can be derived from site-site radial distribution functions by integrating the corresponding RDF up to the first minimum. The numbers describe the probability of binding and they reveal that the C2-hydroxyl groups are most likely bound (at 350 K, 95% probability for $Cl^-$ and 85% for water oxygen), followed by the C6- (85%, 80%) and the C3-hydroxyls (80%, 55%). Moreover, the probabilities indicate that $Cl^-$ anions bind to these groups stronger than water molecules do, consistent with the notion that the hydrogen bond basicity of the anion determines its dissolution ability. The fact that C3-hydroxyls are the least solvent coordinated of the three hydroxyl groups is due to the formation of intramolecular O3H'$\cdots$O5 hydrogen bonds as discussed in the previous chapter. This suggests that the higher coordination of hydroxyl groups by $Cl^-$ than by water molecules results in greater disruption of the O3H'$\cdots$O5 hydrogen bonds, thus leading to the cellulose chain being more flexible in BmimCl.

The sum of coordination numbers, $c_n$, around all hydroxyls of individual

glucose units is given in Figure 25b. The binding of solvent to the terminal monomer is strong because of the availability of its O3H hydrogen, which cannot form an intramolecular O3H'$\cdots$O5 hydrogen bond. The decline of $c_n$ with temperature is consistent with the gain in entropy upon solvent liberation. Interestingly, the water oxygen profile, which has a value of less than 1.5 at the highest temperature, decreases more sharply than the $Cl^-$ profile. The coordination number of 2 to 2.5 $Cl^-$ anions per glucose unit agrees well with previous results[36, 203]. The higher coordination numbers and the slower decay of the $Cl^-$ profile compared to water show that the anion binds to cellulose more strongly.

Three-dimensional spatial distributions underline the stronger interaction of $Cl^-$ versus water molecules with cellulose (Figure 26). Similar to the distribution of $Cl^-$ around the cellulose fiber discussed in chapter 4 (see Figure 8), the anions occupy distinct positions around the hydroxyl groups of C2, C3, and C6 at densities five times higher than that of bulk BmimCl. Density distributions for water oxygens show an average accumulation of water with densities up to 1.8 times its bulk density mainly around the C2-/C3-hydroxyls and around the glucose rings. This analysis confirms the relatively high density of $Cl^-$ around hydroxyls indicating strong hydroxyl-anion binding, which may contribute to the more effective dissolution of cellulose in BmimCl than in water.

As shown in the previous chapter, the cation $Bmim^+$ adopts preferred orientations with respect to proximal surface monomers on cellulose fibers. No specific orientations are observed in the present simulations of a single cellulose chain. Spatial distributions of atoms from the $Bmim^+$ head group show density peaks around the glucose unit at values of 1.5 times the value of bulk BmimCl (Figure 26). These peaks are located on top of the glucose

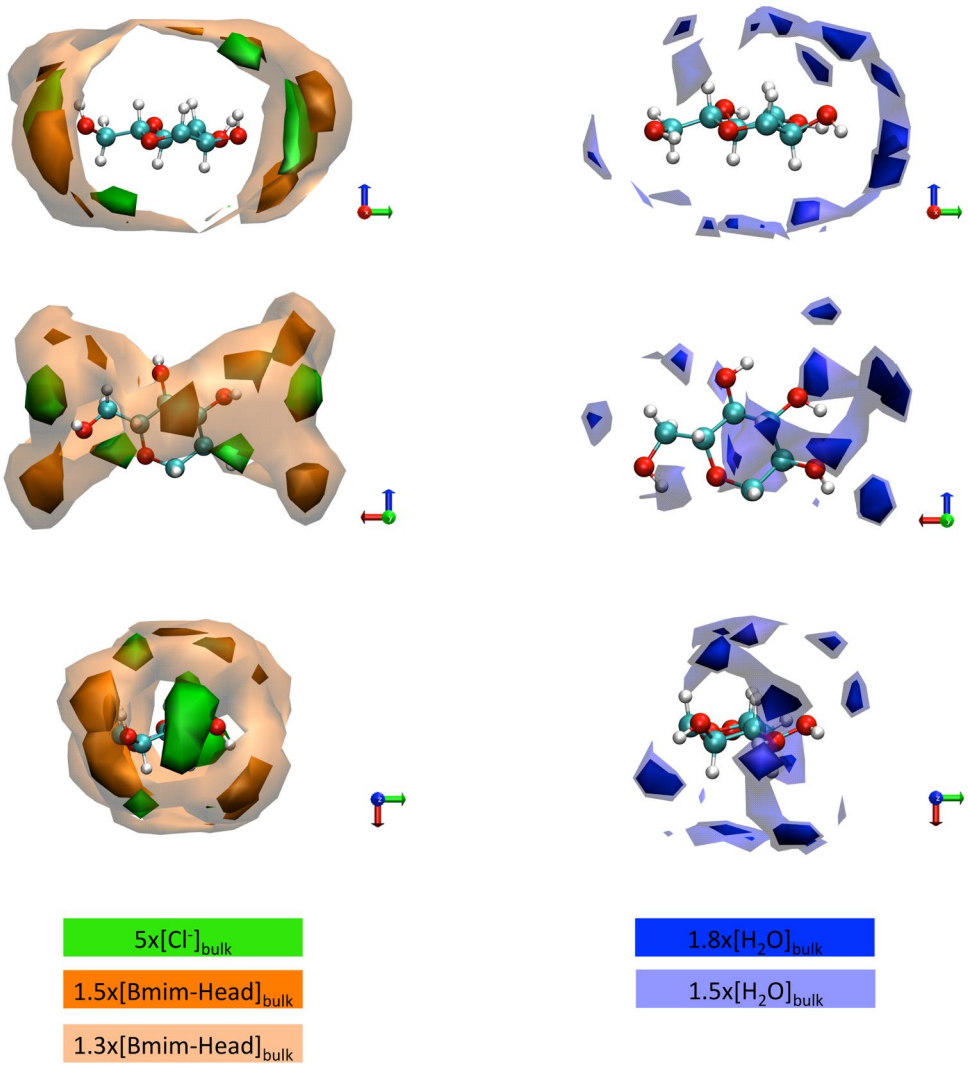



**Figure 26** Spatial distributions of solvent.

Average three-dimensional spatial distributions of Cl (green contours), $Bmim^+$ (orange contours), and water (blue contours) with respect to glucose subunits. The same distributions are visualized in three different orientations.

ring and near the equatorially positioned hydroxyls, but no distinct site-site interactions are found.

Gross et al.[203] have recently reported clearly higher densities of BmimCl ions and water molecules around cellulose chains. In their study, $Cl^-$ anions were found around cellulose hydroxyls with average densities up to 15 times that of bulk BmimCl and even $Bmim^+$ cations and water molecules were found to accumulate above and below the glucose rings, with average densities of 3.5 to 4 times that of their bulk values. The reason for their values being significantly higher than those found here is probably that in Ref. [203] the cellulose chains were simulated with restraints on the terminal residues (and on all residues during equilibration) to keep the chain extended and thus effectively flat during the entire simulation, which lasted only 30 ns. The present REMD simulations, in contrast, are performed on an unrestrained cellulose chain.

To further investigate the interaction between solute and solvent, the interaction energies between the solvated cellulose decamer and its solvent molecules are decomposed into Coulomb ($E_{Coul}$) and van der Waals ($E_{vdW}$) contributions. Figure 27 shows that cellulose-BmimCl interactions are more favorable than those of cellulose with water. While all solute-solvent interactions become less favorable with temperature, $E_{Coul}$ for the cellulose-water interaction is the most affected. However, at the optimal cellulose pretreatment temperature, that is at 375 K in water and at 425-450 K in BmimCl, the corresponding cellulose-solvent Coulomb interaction energies are similar ($E_{Coul} \approx$ -350 kcal/mol).

Consistent with the above results on $Cl^-$ coordination of the cellulose hydroxyl groups, $E_{Coul}$ is significantly stronger for cellulose-$Cl^-$ than for cellulose-$Bmim^+$ in the BmimCl solution. However, $E_{vdW}$ is positive for the cellulose-$Cl^-$ interaction but negative for the cellulose-$Bmim^+$ interaction. This corre-

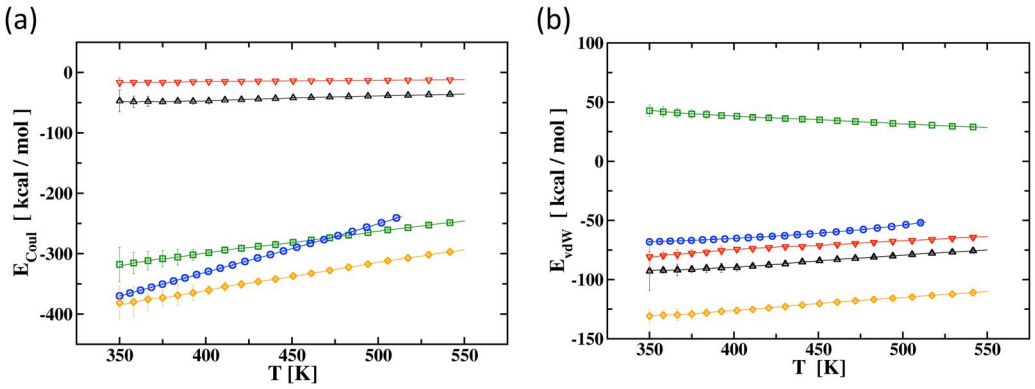



**Figure 27** Solute-solvent interaction energies.

Decomposition of solute-solvent interaction energies into (a) Coulomb and (b) van der Waals contributions. Profiles are given for interactions between the entire cellulose molecule and water (blue circles), BmimCl (orange diamonds), only the $Cl^-$ anion (green squares), only the $Bmim^+$ cation head group (black upward triangles), and only the $Bmim^+$ cation tail group (red downward triangles).

sponds to favorable interaction between cellulose and both the methyl-imidazole head group and the aliphatic tail group of the cation. In fact, all $E_{Coul}$ and $E_{vdW}$ profiles are comparable for the two $Bmim^+$ moieties.

Additionally, monomer-wise separation of these solute-solvent interaction energies reveals that $E_{Coul}$ lies between -37 to -27 kcal/mol/glucose in BmimCl and between -35 to -22 kcal/mol/glucose in water depending on the temperature (Figure 28). The terminal monomers have lower electrostatic interaction energies due to the exposure of the additional polar hydroxyl groups at C1 and C4. The corresponding van der Waals interaction energy is about -12 to -10 kcal/mol/glucose in BmimCl and about -7 to -5 kcal/mol/glucose in water.

These findings show that cellulose chain solvation is facilitated by the interplay of both IL components. The larger cation has more favorable van der

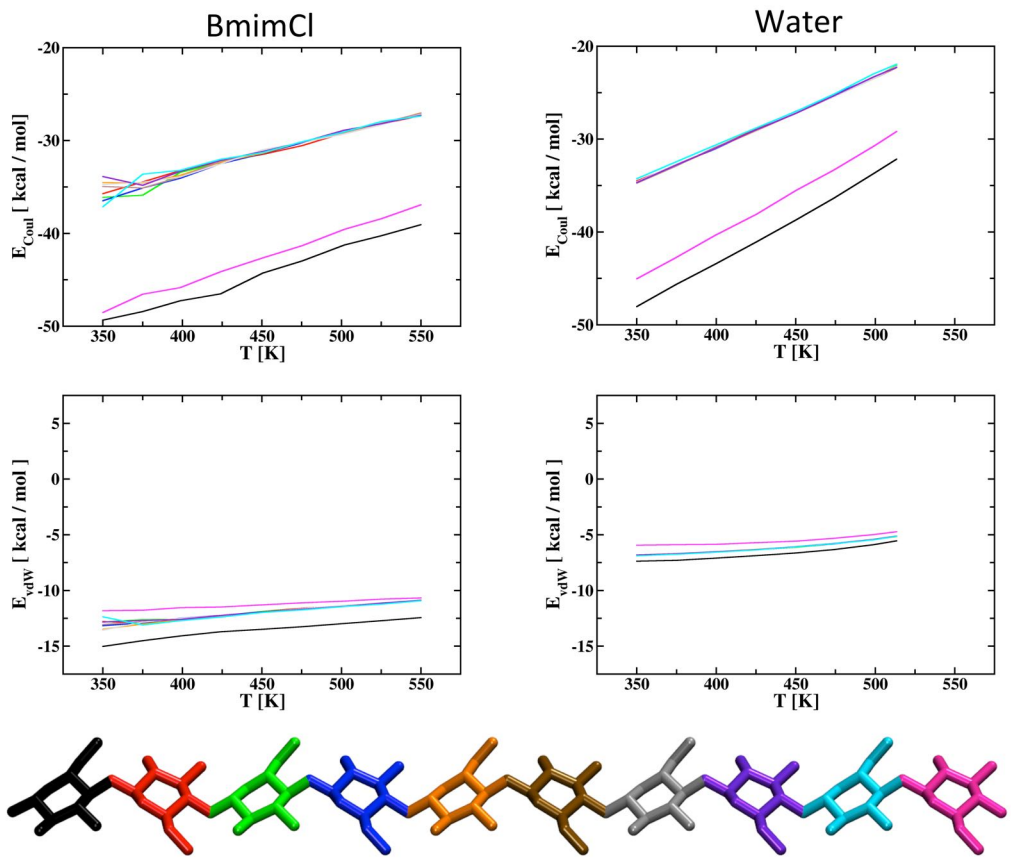



**Figure 28** Solute-solvent interaction energies per residue.

Residue-wise decomposition of solute-solvent interaction energies into Coulomb (top) and van der Waals (bottom) contributions for REMD simulations in BmimCl (left) and in water (right). Colors of the temperature profiles correspond to the glucose unit given below.

Waals interactions with cellulose than does water, and the smaller anion has more favorable Coulomb interactions. The presented interaction energies between the cellulose decamer and its solvent environment, taken together with the analysis of distinct interaction sites between solute and solvent molecules, all imply a more favorable enthalpy of single chain solvation in BmimCl than in water.

## 5.3 Cellulose Configurational Entropy

To understand the differential dissolution of cellulose in the two solvents, in addition to the enthalpy contribution the difference in entropy has to be addressed. The above results on cellulose conformational distributions have indicated the solute to be more dynamic in the IL than in water. Therefore, it is of interest to estimate the absolute configurational entropy of the chain in the two solvents. Here, the QH approach, introduced in chapter 3.3, is used mainly as a means of roughly estimating the conformational flexibility of cellulose in the two solvents. A model is applied that evaluates first-order entropies independently based on their probability distribution and that determines correlations based on the QH method (see equation (3.10))[204]. The entropy is evaluated in torsional space resulting in 120 degrees of freedom for the entire cellulose decamer. This approach, combined with enhanced phase-space sampling, is expected to provide robust results on qualitative differences in cellulose conformational flexibility between the two solvents.

Figure 29a shows the temperature profiles for the configurational entropies calculated using all torsional angles of the cellulose chain in BmimCl and in water. Clearly, above ∼425 K the entropy is higher for cellulose in BmimCl than in water, indicating that the solute is more dynamic in the IL than in wa-

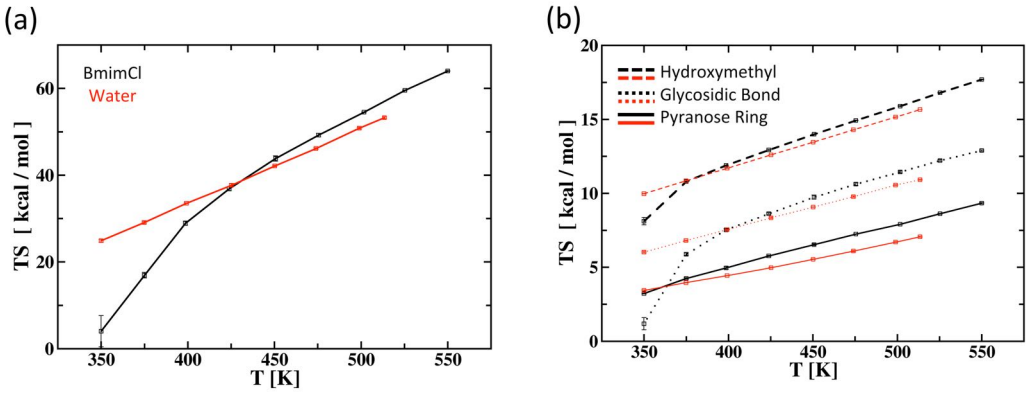



**Figure 29** Cellulose configurational entropy.

(a) Configurational entropy contribution to the free energy of solvation of cellulose in BmimCl (black) and in water (red) based on all cellulose torsions. (b) Configurational entropy contribution to the free energy of solvation of cellulose in BmimCl (black) and in water (red) based on torsions from the hydroxymethyl group (dashed) or from the glycosidic bond(dotted) or from the pyranose ring (solid) only.

ter at elevated temperatures. Performing the same calculation using only the two glycosidic bond torsions (C4'-O4'-C1-O5 and C5'-C4'-O4'-C1), only those two associated with the hydroxymethyl groups (O5-C5-C6-O6 and C5-C6-O6-H6O), or only two such torsions from the pyranose rings (C5-C4-C3-C2 and C2-C1-O5-C5) allows to compare the difference in the configurational entropy with respect to these components. This analysis corresponds to diagonalizing a submatrix of the covariance matrix obtained for the entire cellulose chain while neglecting all the off-diagonal block correlations. This corresponds to an approximate decomposition as the sum of the estimated entropy values for the individual components will not add up to the total entropy of the entire chain. Nevertheless, this allows to make a qualitative assessment of the conformational flexibility of individual cellulose parts in the two solvents. As shown in

Figure 29b, the value of the configurational entropy contribution is highest for the hydroxymethyl groups followed by the glycosidic bonds and the pyranose rings in either solvent. This suggests that the hydroxymethyl rotamer is the most flexible part and the pyranose ring structure is the least flexible part of a glucose unit. Interestingly, all three moieties are estimated to be more flexible in BmimCl than in water. Taken together, this analysis confirms that cellulose is more dynamic in the IL than in water, and that this higher mobility is present in all parts of the glucose unit.

In a previous study, the intramolecular contribution to the entropy change upon cellulose dissolution was found to be more favorable in water than in BmimCl[205], which seems to be in contrast to the current result. However, in Ref. [205] the authors used a two-phase thermodynamic model, in which the entropy contribution from intramolecular degrees of freedom was derived from the density of states of an ideal solid system based on sub-ns simulations. Thus, the conformational dynamics captured for the entropy calculation take place at a much smaller timescale in the previous work[205] than in the present study, where the intramolecular entropy is estimated from configurational probability distributions derived from 100 ns of REMD.

The present REMD simulations suggest that the end state of cellulose dissolution, i.e. a fully solvated single cellulose chain, is more favored in BmimCl than in water. Both Coulombic and van der Waals interaction energies between solute and solvent are shown to be more favorable and the configurational entropy of cellulose is estimated to be higher in BmimCl than in water. These results may describe a driving force for the efficient dissolution of cellulose in the IL.

# 6

# Conclusions

One of the major obstacles to overcome in the production of lignocellulose-based biofuels is the efficient pretreatment of biomass, which includes the dissolution and disintegration of cellulose fibers. In order to shed some light on the processes that may control cellulose dissolution in ionic liquids, the initial and final stages of cellulose dissolution are analyzed by all-atom MD simulations of cellulose solvated in BmimCl. This study provides some relevant insight: the simulation of a cellulose fiber indicates that the different ions act synergistically on the different fiber surface to initiate the dissolution process, and REMD simulations of single cellulose chains in BmimCl and in water suggest that a driving force for the efficient dissolution of cellulose in BmimCl may arise from favorable solute-solvent interactions coupled with an increase in cellulose conformational flexibility.

In the fiber simulation in BmimCl selective accumulation of $Cl^-$ anions is found around the solvent-exposed hydroxyl groups of the glucan chains. This contact disrupts intrachain hydrogen bonds on the hydrophilic surface origin chains. In contrast, $Bmim^+$ cations undergo ring stacking interactions

with the cellulose hydrophobic surface and intercalate between chains on the hydrophilic surface. Hence, overall, BmimCl solvates cellulose by interacting with both the hydroxyl groups and the hydrophobic side of the polysaccharide, consistent with the Lindman Hypothesis[29, 31], and this may explain why ILs act as ideal solvents for the amphiphilic cellulose molecule. These results are complementary to those from previous simulations on cellulose-IL interactions, in which active roles during dissolution for both anions and cations had been suggested[120, 125].

The REMD simulations of a cellulose decamer are performed between 350 K and 550 K, which is necessary for the sampling of a variety of cellulose structures in a solvent such as BmimCl. The results indicate that, although global size-related properties of the cellulose chain are comparable in the two solvents, differences are found in the flexibility and conformations of cellulose. For example, although in both solvents the temperature profile of $R_g$ is similar, the temperature profiles of P(GG)/P(GT) or P(O3H'$\cdots$O5) are not. Also, cellulose $\phi$-$\psi$ scatter plots reveal distinct BmimCl and water basins. Furthermore, higher conformational mobility of cellulose in BmimCl is indirectly suggested by a wider free energy basin for the order parameters P(GG)/P(GT) and P(O3H'$\cdots$O5), or by a larger variance in their corresponding temperature profiles, and also by the greater flipping of pyranose rings. The difference in conformational flexibility of cellulose chains between BmimCl and water solutions is qualitatively assessed by the computation of configurational entropies. All the above-mentioned differences in local cellulose structural properties, together with the more favorable solute-solvent interaction energies, in BmimCl compared to water may lead to the greater dissolution ability of the IL.

In conclusion, this study provides some information on molecular interactions relevant for the initiation of cellulose dissolution in BmimCl and some

information on the differences in the dynamics and thermodynamics of a fully solvated single chain in BmimCl and in water. The use of MD simulations, and in particular the use of REMD for the single chain simulations, has proven to be very suitable for such an investigation of cellulose dissolution.

## 6.1 Future Perspectives

As mentioned above, the results presented here shed light only on the initial and the final stages of the dissolution process by ILs as large-scale structural changes of the entire fiber remain out of reach due to computational limitation. However, such large structural changes of cellulose may be induced by specific solute-solvent interactions as shown in a recent experimental study, in which cellulose I$\beta$ was treated with liquid ammonia[206, 207]. Therefore, future simulation studies that try to investigate the impact of ILs on cellulose may focus on identifying distinct interaction patterns during experimentally identified different stages of dissolution. In particular, since it is believed that cellulose fibers are hydrated to a certain degree when exposed to the IL solvent in the experiments, MD simulations of hydrated fibers in ILs could reveal to what extent the results of the present study correspond to those from a mixed solvent system. The simulation of a single cellulose chain in such a binary solvent has been performed before, providing some results on the interaction betwen cellulose, water, and the IL 1-ethyl-3-methylimidazolium acetate (EmimAc)[122].

Clearly, to assess the general validity of the present results for other cellulose-dissolving ILs, simulations of cellulose in ILs such as EmimAc have to be run and analyzed in the same way as the cellulose fiber and cellulose chains in the current work. In particular, it would be interesting to see how ILs with cations

of different tail length interact with the hydrophilic faces of a fiber because the distinct interaction pattern between the IL and the cellulose fiber described in chapter 4 probably does not emerge with larger cations. It would be equally interesting to investigate whether the conformational flexibility of single cellulose chains in other ILs corresponds to that observed in BmimCl as described in chapter 5. ILs that are known to be less efficient in the dissolution of cellulose than BmimCl, e.g. such with an allyl instead of an alkyl substituent, can be used as the solvent environment in future cellulose simulation studies in order to evaluate the significance of the results obtained in the present work.

Finally, in order to realize the full potential of ILs in biomass treatment, more research needs to be done to elucidate the effect of ILs on biomass that includes other components, such as lignin and hemicellulose. It is known that ILs can dissolve all the three major biomass components from a variety of sources, e.g. from wood, cotton, or bamboo, which allows their subsequent use for advanced materials in a sustainable industry[89]. However, the implementation of this into a new technology still requires the development of some separation processes and the utilization of ILs as a reusable green solvent. A fundamental knowledge of the driving force for the more efficient dissolution of lignocellulose in ILs than in water will enable us to improve the biomass pretreatment process, which is an essential step in the production of biomaterials and biofuels.

# Bibliography

[1] Hoffert, M. I. et al. (2002) Advanced Technology Paths to Global Climate Stability: Energy for a Greenhouse Planet. *Science 298*, 981–987.

[2] Barnett, J. (2003) Security and climate change. *Global Environmental Change 13*, 7–17.

[3] Kerr, R. A. (2007) The Looming Oil Crisis Could Arrive Uncomfortably Soon. *Science 316*, 351.

[4] Bridge, G., Bouzarovski, S., Bradshaw, M., and Eyre, N. (2013) Geographies of energy transition: Space, place and the low-carbon economy. *Energy Policy 53*, 331–340.

[5] Graham-Rowe, D. (2011) Agriculture: Beyond food versus fuel. *Nature 474*, S6–S8.

[6] Escobar, J. C., Lora, E. S., Venturini, O. J., Yáñez, E. E., Castillo, E. F., and Almazan, O. (2009) Biofuels: Environment, technology and food security. *Renewable and Sustainable Energy Reviews 13*, 1275–1287.

[7] Lynd, L. R. (1996) Overview and Evaluation of Fuel Ethanol from Cellulosic Biomass: Technology, Economics, the Environment, and Policy. *Annual Review of Energy and the Environment 21*, 403–465.

[8] Huber, G. W., Iborra, S., and Corma, A. (2006) Synthesis of Transportation Fuels from Biomass: Chemistry, Catalysts, and Engineering. *Chem. Rev. 106*, 4044–4098.

[9] Carroll, A., and Somerville, C. (2009) Cellulosic Biofuels. *Annual Review of Plant Biology 60*, 165–182.

[10] Savage, N. (2011) Fuel options: The ideal biofuel. *Nature 474*, S9–S11.

[11] Robbins, M. (2011) Policy: Fuelling politics. *Nature 474*, S22–S24.

[12] Sims, R. E. H., Mabee, W., Saddler, J. N., and Taylor, M. (2010) An overview of second generation biofuel technologies. *Bioresource Technology 101*, 1570–1580.

[13] Fairley, P. (2011) Introduction: Next generation biofuels. *Nature 474*, S2–S5.

[14] Lee, R. A., and Lavoie, J.-M. (2013) From first- to third-generation biofuels: Challenges of producing a commodity from a biomass of increasing complexity. *Animal Frontiers 3*, 6–11.

[15] Simmons, B., Loque, D., and Blanch, H. (2008) Next-generation biomass feedstocks for biofuel production. *Genome Biology 9*, 242+.

[16] Somerville, C., Youngs, H., Taylor, C., Davis, S. C., and Long, S. P. (2010) Feedstocks for Lignocellulosic Biofuels. *Science 329*, 790–792.

[17] Langan, P., Gnanakaran, S., Rector, K. D., Pawley, N., Fox, D. T., Cho, D. W., and Hammel, K. E. (2011) Exploring new strategies for cellulosic biofuels production. *Energy Environ. Sci. 4*, 3820–3833.

[18] Schubert, C. (2006) Can biofuels finally take center stage? *Nature Biotechnology 24*, 777–784.

[19] 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., and Tschaplinski, T. (2006) The Path Forward for Biofuels and Biomaterials. *Science 311*, 484–489.

[20] Nigam, P. S., and Singh, A. (2011) Production of liquid biofuels from renewable resources. *Progress in Energy and Combustion Science 37*, 52–68.

[21] Himmel, M. E., Ding, S.-Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W., and Foust, T. D. (2007) Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production. *Science 315*, 804–807.

[22] Liebert, T. (2010) Cellulose Solvents - Remarkable History, Bright Future. *1033*, 3–54.

[23] Cuissinat, C., Navard, P., and Heinze, T. (2008) Swelling and dissolution of cellulose. Part IV: Free floating cotton and wood fibres in ionic liquids. *Carbohydrate Polymers 72*, 590–596.

[24] Heinze, T., and Liebert, T. (2001) Unconventional methods in cellulose functionalization. *Progress in Polymer Science 26*, 1689–1762.

[25] Philipp, B., Lukanoff, B., Schleicher, H., and Wagenknecht, W. (1986) Homogene Umsetzung an Cellulose in organischen Lösemittelsystemen. *Z. Chem. 26*, 50–58.

[26] BeMiller, J. N., and Whistler, L. In *Food Chemistry*; Fennema, O. R., Ed.; CRC: New York, 1996; pp 157–224.

[27] Zhang, L., Ruan, D., and Gao, S. (2002) Dissolution and regeneration of cellulose in NaOH/thiourea aqueous solution. *J. Polym. Sci. B Polym. Phys. 40*, 1521–1529.

[28] Lindman, B., Karlström, G., and Stigsson, L. (2010) On the Mechanism of Dissolution of Cellulose. *Journal of Molecular Liquids*

[29] Medronho, B., Romano, A., Miguel, M., Stigsson, L., and Lindman, B. (2012) Rationalizing cellulose (in)solubility: reviewing basic physicochemical aspects and role of hydrophobic interactions. *Cellulose 19*, 581–587.

[30] Bergenstråhle, M., Wohlert, J., Himmel, M. E., and Brady, J. W. (2010) Simulation studies of the insolubility of cellulose. *Carbohydrate Research 345*, 2060–2066.

[31] Glasser, W., Atalla, R., Blackwell, J., Malcolm Brown, R., Burchard, W., French, A., Klemm, D., and Nishiyama, Y. (2012) About the structure of cellulose: debating the Lindman hypothesis. *Cellulose 19*, 589–598.

[32] Zhu, S., Wu, Y., Chen, Q., Yu, Z., Wang, C., Jin, S., Ding, Y., and Wu, G. (2006) Dissolution of cellulose with ionic liquids and its application: a mini-review. *Green Chem. 8*, 325–327.

[33] Heinze, T., Schwikal, K., and Barthel, S. (2005) Ionic Liquids as Reaction Medium in Cellulose Functionalization. *Macromol. Biosci. 5*, 520–525.

[34] Wasserscheid, P., and Welton, T. In *Ionic Liquids in Synthesis*; Wasserscheid, P., and Welton, T., Eds.; Wiley, 2007; Vol. 2.

[35] Matthews, J., Skopec, C., Mason, P., Zuccato, P., Torget, R., Sugiyama, J., Himmel, M., and Brady, J. (2006) Computer simulation studies of microcrystalline cellulose I. *Carbohydrate Research 341*, 138–152.

[36] Youngs, T. G. A., Hardacre, C., and Holbrey, J. D. (2007) Glucose Solvation by the Ionic Liquid 1,3-Dimethylimidazolium Chloride: A Simulation Study. *J. Phys. Chem. B 111*, 13765–13774.

[37] Ramadugu, S. K., Chung, Y.-H., Xia, J., and Margulis, C. J. (2009) When Sugars Get Wet. A Comprehensive Study of the Behavior of Water on the Surface of Oligosaccharides. *J. Phys. Chem. B 113*, 11003–11015.

[38] Cosgrove, D. J. (2005) Growth of the plant cell wall. *Nat Rev Mol Cell Biol 6*, 850–861.

[39] Ding, S.-Y., and Himmel, M. E. (2006) The Maize Primary Cell Wall Microfibril: A New Model Derived from Direct Visualization. *J. Agric. Food Chem. 54*, 597–606.

[40] Klemm, D., Heublein, B., Fink, H.-P., and Bohn, A. (2005) Cellulose: Fascinating Biopolymer and Sustainable Raw Material. *ChemInform 36*, no.

[41] Ohad, I., and Danon, D. (1964) ON THE DIMENSIONS OF CELLULOSE MICROFIBRILS. *The Journal of Cell Biology 22*, 302–305.

[42] Ha, M.-A., Apperley, D. C., Evans, B. W., Huxham, I. M., Jardine, W. G., Viëtor, R. J., Reis, D., Vian, B., and Jarvis, M. C. (1998) Fine structure in cellulose microfibrils: NMR evidence from onion and quince. *The Plant Journal 16*, 183–190.

[43] Leppänen, K., Andersson, S., Torkkeli, M., Knaapila, M., Kotelnikova, N., and Serimaa, R. (2009) Structure of cellulose and microcrystalline cellulose from various wood species, cotton and flax studied by X-ray scattering. *Cellulose 16*, 999–1015.

[44] Foston, M., Hubbell, C., Davis, M., and Ragauskas, A. (2009) Variations in Cellulosic Ultrastructure of Poplar. *BioEnergy Research 2*, 193–197.

[45] Atalla, R. H., and van der Hart, D. L. (1984) Native Cellulose: A Composite of Two Distinct Crystalline Forms. *Science 223*, 283–285.

[46] Nishiyama, Y., Langan, P., and Chanzy, H. (2002) Crystal Structure and Hydrogen-Bonding System in Cellulose I from Synchrotron X-ray and Neutron Fiber Diffraction. *Journal of the American Chemical Society 124*, 9074–9082.

[47] Nishiyama, Y., Sugiyama, J., Chanzy, H., and Langan, P. (2003) Crystal Structure and Hydrogen Bonding System in Cellulose I from Synchrotron X-ray and Neutron Fiber Diffraction. *J. Am. Chem. Soc. 125*, 14300–14306.

[48] Halonen, H., Larsson, P., and Iversen, T. (2013) Mercerized cellulose biocomposites: a study of influence of mercerization on cellulose supramolecular structure, water retention value and tensile properties. *Cellulose 20*, 57–65.

[49] Wada, M., Chanzy, H., Nishiyama, Y., and Langan, P. (2004) Cellulose IIII Crystal Structure and Hydrogen Bonding by Synchrotron X-ray and Neutron Fiber Diffraction. *Macromolecules 37*, 8548–8555.

[50] Wada, M., Nishiyama, Y., and Langan, P. (2006) X-ray Structure of AmmoniaCellulose I: New Insights into the Conversion of Cellulose I to Cellulose IIII. *Macromolecules 39*, 2947–2952.

[51] Wada, M., Heux, L., and Sugiyama, J. (2004) Polymorphism of Cellulose I Family: Reinvestigation of Cellulose IVI. *Biomacromolecules 5*, 1385–1391.

[52] Ciolacu, D., and Popa, V. I. In *Cellulose: Structure and Properties, Derivatives and Industrial Uses*; Lejeune, A., and Deprez, T., Eds.; Nova Science Publisher Inc.: United States, 2010; pp 1–38.

[53] Atalla Rajai, H. *The Structures of Cellulose*; American Chemical Society, 1987; Vol. 340; pp 1–14.

[54] Chundawat, S. P. S., Beckham, G. T., Himmel, M. E., and Dale, B. E. (2011) Deconstruction of Lignocellulosic Biomass to Fuels and Chemicals. *Annual Review of Chemical and Biomolecular Engineering 2*, 121–145.

[55] Lin, Y.-C., and Huber, G. W. (2009) The critical role of heterogeneous catalysis in lignocellulosic biomass conversion. *Energy Environ. Sci. 2*, 68–80.

[56] Balan, V., Bals, B., Chundawat, S., Marshall, D., and Dale, B. In *Biofuels*; Mielenz, J. R., Ed.; Humana Press, 2009; Vol. 581; pp 61–77.

[57] Lau, M., Gunawan, C., and Dale, B. (2009) The impacts of pretreatment on the fermentability of pretreated lignocellulosic biomass: a comparative evaluation between ammonia fiber expansion and dilute acid pretreatment. *Biotechnology for Biofuels 2*, 61+.

[58] Tian, S., Luo, X. L., Yang, X. S., and Zhu, J. Y. (2010) Robust cellulosic ethanol production from SPORL-pretreated lodgepole pine using an adapted strain Saccharomyces cerevisiae without detoxification. *Bioresource Technology 101*, 8678–8685.

[59] Shuai, L., Yang, Q., Zhu, J. Y., Lu, F. C., Weimer, P. J., Ralph, J., and Pan, X. J. (2010) Comparative study of SPORL and dilute-acid

pretreatments of spruce for cellulosic ethanol production. *Bioresource Technology 101*, 3106–3114.

[60] Romero, I., Moya, M., Sánchez, S., Ruiz, E., Castro, E., and Bravo, V. (2007) Ethanolic fermentation of phosphoric acid hydrolysates from olive tree pruning. *Industrial Crops and Products 25*, 160–168.

[61] Geddes, C. C., Peterson, J. J., Roslander, C., Zacchi, G., Mullinnix, M. T., Shanmugam, K. T., and Ingram, L. O. (2010) Optimizing the saccharification of sugar cane bagasse using dilute phosphoric acid followed by fungal cellulases. *Bioresource Technology 101*, 1851–1857.

[62] Mosier, N. (2005) Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresource Technology 96*, 673–686.

[63] Alvira, P., Tomás-Pejó, E., Ballesteros, M., and Negro, M. J. (2010) Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review. *Bioresource Technology 101*, 4851–4861.

[64] Wyman, C. E., Dale, B. E., Elander, R. T., Holtzapple, M., Ladisch, M. R., and Lee, Y. Y. (2005) Coordinated development of leading biomass pretreatment technologies. *Bioresource Technology 96*, 1959–1966.

[65] Geddes, C. C., Nieves, I. U., and Ingram, L. O. (2011) Advances in ethanol production. *Current Opinion in Biotechnology 22*, 312–319.

[66] ten Have, R., and Teunissen, P. J. M. (2001) Oxidative Mechanisms

Involved in Lignin Degradation by White-Rot Fungi. *Chem. Rev. 101*, 3397–3414.

[67] Hammel, K., and Cullen, D. (2008) Role of fungal peroxidases in biological ligninolysis. *Current Opinion in Plant Biology 11*, 349–355.

[68] Krossing, I., Slattery, J. M., Daguenet, C., Dyson, P. J., Oleinikova, A., and Weingärtner, H. (2006) Why Are Ionic Liquids Liquid? A Simple Explanation Based on Lattice and Solvation Energies. *J. Am. Chem. Soc. 128*, 13427–13434.

[69] Rogers, R. D., and Seddon, K. R. (2003) Ionic Liquids–Solvents of the Future? *Science 302*, 792–793.

[70] Hagiwara, H., Sugawara, Y., Isobe, K., Hoshi, T., and Suzuki, T. (2004) Immobilization of Pd(OAc)2 in Ionic Liquid on Silica: Application to Sustainable MizorokiHeck Reaction. *Org. Lett. 6*, 2325–2328.

[71] Zhao, H., Xia, S., and Ma, P. (2005) Use of ionic liquids as lsquo-greenrsquo solvents for extractions. *Journal of Chemical Technology & Biotechnology 80*, 1089–1096.

[72] Heintz, A., and Wertz, C. (2006) Ionic liquids: A most promising research field in solution chemistry and thermodynamics. *Pure and Applied Chemistry 78*, 1587–1593.

[73] Plechkova, N. V., and Seddon, K. R. (2008) Applications of ionic liquids in the chemical industry. *Chem. Soc. Rev. 37*, 123–150.

[74] Freemantle, M. In *An Introduction to Ionic Liquids*; Freemantle, M., Ed.; Royal Society of Chemistry: Cambridge, 2009.

[75] Zhang, Y., and Chan, J. Y. (2010) Sustainable chemistry: imidazolium salts in biomass conversion and CO2 fixation. *Energy Environ. Sci. 3*, 408–417.

[76] Walden, P. (1914) Molecular Weights and Electrical Conductivity of Several Fused Salts. *Bulletin of the Russian Academy of Sciences* 405–422.

[77] für Chemische Industrie in Basel, G. *CH 153446*; 1932.

[78] Graenacher, C. *US Patent 1943176*; 1934.

[79] Swatloski, R. P., Spear, S. K., Holbrey, J. D., and Rogers, R. D. (2002) Dissolution of Cellose with Ionic Liquids. *Journal of the American Chemical Society 124*, 4974–4975.

[80] Swatloski, R. P., Rogers, R. D., and Holbrey, J. D. *WO Patent 03/029329*; 2003.

[81] Kosan, B., Michels, C., and Meister, F. (2008) Dissolution and forming of cellulose with ionic liquids. *Cellulose 15*, 59–66.

[82] Wendler, F., Kosan, B., Krieg, M., and Meister, F. (2009) Possibilities for the Physical Modification of Cellulose Shapes Using Ionic Liquids. *Macromol. Symp. 280*, 112–122.

[83] Cao, Y., Li, H., Zhang, Y., Zhang, J., and He, J. (2010) Structure and properties of novel regenerated cellulose films prepared from cornhusk cellulose in room temperature ionic liquids. *J. Appl. Polym. Sci. 116*, 547–554.

[84] Gericke, M., Trygg, J., and Fardim, P. (2013) Functional Cellulose Beads: Preparation, Characterization, and Applications. *Chem. Rev. 113*, 4812–4836.

[85] El Seoud, O. A., Koschella, A., Fidale, L. C., Dorn, S., and Heinze, T. (2007) Applications of ionic liquids in carbohydrate chemistry: a window of opportunities. *Biomacromolecules 8*, 2629–2647.

[86] Gericke, M., Fardim, P., and Heinze, T. (2012) Ionic Liquids Promising but Challenging Solvents for Homogeneous Derivatization of Cellulose. *Molecules 17*, 7458–7502.

[87] Heinze, T., Koschella, A., Liebert, T., Harabagiu, V., and Coseri, S. In *The European Polysaccharide Network of Excellence (EPNOE)*; Navard, P., Ed.; Springer Vienna, 2013; pp 283–327.

[88] Heinze, T., and Gericke, M. In *Production of Biofuels and Chemicals with Ionic Liquids*; Fang, Z., Smith, R. L., and Qi, X., Eds.; Springer Netherlands, 2014; Vol. 1; pp 107–144.

[89] Sun, N., Rodriguez, H., Rahman, M., and Rogers, R. D. (2011) Where are ionic liquid strategies most suited in the pursuit of chemicals and energy from lignocellulosic biomass? *Chem. Commun. 47*, 1405–1421.

[90] Stark, A. (2011) Ionic liquids in the biorefinery: a critical assessment of their potential. *Energy Environ. Sci. 4*, 19–32.

[91] Mora-Pale, M., Meli, L., Doherty, T. V., Linhardt, R. J., and Dordick, J. S. (2011) Room temperature ionic liquids as emerging solvents for the pretreatment of lignocellulosic biomass. *Biotechnol. Bioeng. 108*, 1229–1245.

[92] Wang, H., Gurau, G., and Rogers, R. D. (2012) Ionic liquid processing of cellulose. *Chemical Society reviews 41*, 1519–1537.

[93] Luo, J., Cai, M., and Gu, T. In *Green Biomass Pretreatment for Biofuels Production*; Gu, T., Ed.; Springer Netherlands, 2013; pp 127–153.

[94] Liu, L., and Chen, H. (2006) Enzymatic hydrolysis of cellulose materials treated with ionic liquid [BMIM] Cl. *Chinese Science Bulletin 51*, 2432–2436.

[95] Li, Q., He, Y.-C., Xian, M., Jun, G., Xu, X., Yang, J.-M., and Li, L.-Z. (2009) Improving enzymatic hydrolysis of wheat straw using ionic liquid 1-ethyl-3-methyl imidazolium diethyl phosphate pretreatment. *Bioresource Technology 100*, 3570–3575.

[96] Dadi, A. P., Varanasi, S., and Schall, C. A. (2006) Enhancement of cellulose saccharification kinetics using an ionic liquid pretreatment step. *Biotechnol. Bioeng. 95*, 904–910.

[97] Kim, S.-J., Dwiatmoko, A. A., Choi, J. W., Suh, Y.-W., Suh, D. J., and Oh, M. (2010) Cellulose pretreatment with 1-n-butyl-3-methylimidazolium chloride for solid acid-catalyzed hydrolysis. *Bioresource Technology 101*, 8273–8279.

[98] Dadi, A. P., Schall, C. A., and Varanasi, S. (2007) Mitigation of cellulose recalcitrance to enzymatic hydrolysis by ionic liquid pretreatment. *Applied Biochemistry and Biotechnology 137-140*, 407–421.

[99] Li, C., Knierim, B., Manisseri, C., Arora, R., Scheller, H. V., Auer, M., Vogel, K. P., Simmons, B. A., and Singh, S. (2010) Comparison of dilute acid and ionic liquid pretreatment of switchgrass: Biomass recalcitrance,

delignification and enzymatic saccharification. *Bioresource Technology 101*, 4900–4906.

[100] Cheng, G., Varanasi, P., Li, C., Liu, H., Melnichenko, Y. B., Simmons, B. A., Kent, M. S., and Singh, S. (2011) Transition of Cellulose Crystalline Structure and Surface Morphology of Biomass as a Function of Ionic Liquid Pretreatment and Its Relation to Enzymatic Hydrolysis. *Biomacromolecules 12*, 933–941.

[101] Samayam, I. P., Hanson, B. L., Langan, P., and Schall, C. A. (2011) Ionic-Liquid Induced Changes in Cellulose Structure Associated with Enhanced Biomass Hydrolysis. *Biomacromolecules 12*, 3091–3098.

[102] Chen, Y., Stipanovic, A. J., Winter, W. T., Wilson, D. B., and Kim, Y.-J. (2007) Effect of digestion by pure cellulases on crystallinity and average chain length for bacterial and microcrystalline celluloses. *Cellulose 14*, 283–293.

[103] Chundawat, S. P. S., Bellesia, G., Uppugundla, N., da Costa Sousa, L., Gao, D., Cheh, A. M., Agarwal, U. P., Bianchetti, C. M., Phillips, G. N., Langan, P., Balan, V., Gnanakaran, S., and Dale, B. E. (2011) Restructuring the Crystalline Cellulose Hydrogen Bond Network Enhances Its Depolymerization Rate. *J. Am. Chem. Soc. 133*, 11163–11174.

[104] Erdmenger, T., Haensch, C., Hoogenboom, R., and Schubert, U. S. (2007) Homogeneous Tritylation of Cellulose in 1-Butyl-3-methylimidazolium Chloride. *Macromol. Biosci. 7*, 440–445.

[105] Heinze, T., Dorn, S., Schöbitz, M., Liebert, T., Köhler, S., and Meis-

ter, F. (2008) Interactions of Ionic Liquids with Polysaccharides 2: Cellulose. *Macromol. Symp. 262*, 8–22.

[106] Vitz, J., Erdmenger, T., Haensch, C., and Schubert, U. S. (2009) Extended dissolution studies of cellulose in imidazolium based ionic liquids. *Green Chem. 11*, 417–424.

[107] Fukaya, Y., Hayashi, K., Kim, S. S., and Ohno, H. (2010) Design of Polar Ionic Liquids To Solubilize Cellulose without Heating. *1033*, 55–66.

[108] Rinaldi, R., Palkovits, R., and Schüth, F. (2008) Depolymerization of Cellulose Using Solid Catalysts in Ionic Liquids. *Angewandte Chemie International Edition 47*, 8047–8050.

[109] Zavrel, M., Bross, D., Funke, M., Büchs, J., and Spiess, A. C. (2009) High-throughput screening for ionic liquids dissolving (ligno-)cellulose. *Bioresource Technology 100*, 2580–2587.

[110] Zhang, J., Zhang, H., Wu, J., Zhang, J., He, J., and Xiang, J. (2010) NMR spectroscopic studies of cellobiose solvation in EmimAc aimed to understand the dissolution mechanism of cellulose in ionic liquids. *Phys. Chem. Chem. Phys. 12*, 1941–1947.

[111] Çetinkol, O. P., Dibble, D. C., Cheng, G., Kent, M. S., Knierim, B., Auer, M., Wemmer, D. E., Pelton, J. G., Melnichenko, Y. B., Ralph, J., Simmons, B. A., and Holmes, B. M. (2010) Understanding the impact of ionic liquid pretreatment on eucalyptus. *Biofuels 1*, 33–46.

[112] Lucas, M., Wagner, G. L., Nishiyama, Y., Hanson, L., Samayam, I. P., Schall, C. A., Langan, P., and Rector, K. D. (2011) Reversible swelling

of the cell wall of poplar biomass by ionic liquid at room temperature. *Bioresource Technology 102*, 4518–4523.

[113] Cheng, G., Varanasi, P., Arora, R., Stavila, V., Simmons, B. A., Kent, M. S., and Singh, S. (2012) Impact of Ionic Liquid Pretreatment Conditions on Cellulose Crystalline Structure Using 1-Ethyl-3-methylimidazolium Acetate. *J. Phys. Chem. B 116*, 10049–10054.

[114] Ohira, K., Abe, Y., Kawatsura, M., Suzuki, K., Mizuno, M., Amano, Y., and Itoh, T. (2012) Design of Cellulose Dissolving Ionic Liquids Inspired by Nature. *ChemSusChem 5*, 388–391.

[115] Remsing, R. C., Swatloski, R. P., Rogers, R. D., and Moyna, G. (2006) Mechanism of cellulose dissolution in the ionic liquid 1-n-butyl-3-methylimidazolium chloride: a 13C and 35/37Cl NMR relaxation study on model systems. *Chem. Commun.* 1271–1273.

[116] Remsing, R. C., Hernandez, G., Swatloski, R. P., Massefski, W. W., Rogers, R. D., and Moyna, G. (2008) Solvation of carbohydrates in n,n'-dialkylimidazolium ionic liquids: a multinuclear NMR spectroscopy study. *The journal of physical chemistry. B 112*, 11071–11078.

[117] Sellin, M., Ondruschka, B., and Stark, A. In *Cellulose Solvents: For Analysis, Shaping and Chemical Modification*; Liebert, T. F., Heinze, T. J., and Edgar, K. J., Eds.; American Chemical Society, 2010; Vol. 1033; pp 121–135.

[118] Zhang, H., Wu, J., Zhang, J., and He, J. (2005) 1-Allyl-3-methylimidazolium Chloride Room Temperature Ionic Liquid: A New

and Powerful Nonderivatizing Solvent for Cellulose. *Macromolecules 38*, 8272–8277.

[119] Zhao, Y., Liu, X., Wang, J., and Zhang, S. (2013) Effects of anionic structure on the dissolution of cellulose in ionic liquids revealed by molecular simulation. *Carbohydrate Polymers 94*, 723–730.

[120] Zhao, Y., Liu, X., Wang, J., and Zhang, S. (2012) Effects of cationic structure on cellulose dissolution in ionic liquids: a molecular dynamics study. *Chemphyschem : a European journal of chemical physics and physical chemistry 13*, 3126–3133.

[121] Liu, H., Sale, K. L., Holmes, B. M., Simmons, B. A., and Singh, S. (2010) Understanding the Interactions of Cellulose with Ionic Liquids: A Molecular Dynamics Study. *J. Phys. Chem. B 114*, 4293–4301.

[122] Liu, H., Sale, K. L., Simmons, B. A., and Singh, S. (2011) Molecular Dynamics Study of Polysaccharides in Binary Solvent Mixtures of an Ionic Liquid and Water. *J. Phys. Chem. B 115*, 10251–10258.

[123] Liu, H., Cheng, G., Kent, M., Stavila, V., Simmons, B. A., Sale, K. L., and Singh, S. (2012) Simulations Reveal Conformational Changes of Methylhydroxyl Groups during Dissolution of Cellulose I in Ionic Liquid 1-Ethyl-3-methylimidazolium Acetate. *J. Phys. Chem. B 116*, 8131–8138.

[124] Rabideau, B. D., Agarwal, A., and Ismail, A. E. (2013) Observed Mechanism for the Breakup of Small Bundles of Cellulose I and I in Ionic Liquids from Molecular Dynamics Simulations. *J. Phys. Chem. B 117*, 3469–3479.

[125] Cho, H. M., Gross, A. S., and Chu, J.-W. (2011) Dissecting Force Interactions in Cellulose Deconstruction Reveals the Required Solvent Versatility for Overcoming Biomass Recalcitrance. *J. Am. Chem. Soc. 133*, 14033–14041.

[126] Praprotnik, M., and Site, L. In *Biomolecular Simulations*; Monticelli, L., and Salonen, E., Eds.; Humana Press, 2013; Vol. 924; pp 567–583.

[127] Frenkel, D., and Smit, B. *Understanding Molecular Simulation, Second Edition: From Algorithms to Applications (Computational Science)*; Academic Press, 2001.

[128] Smith, J. C., and Roux, B. (2014) Eppur Si Muove! The 2013 Nobel Prize in Chemistry. *Structure 21*, 2102–2105.

[129] Shea, J. E., and Brooks, C. L. (2001) From Folding Theories to Folding Proteins: A Review and Assessment of Simulation Studies of Protein Folding and Unfolding. *Annual Review of Physical Chemistry 52*, 499–535.

[130] Mangoni, M., Roccatano, D., and Di Nola, A. (1999) Docking of flexible ligands to flexible receptors in solution by molecular dynamics simulation. *Proteins 35*, 153–162.

[131] Wang, H.-L., Cheng, X., Taylor, P., McCammon, J. A., and Sine, S. M. (2008) Control of Cation Permeation through the Nicotinic Receptor Channel. *PLoS Comput Biol 4*, e41+.

[132] Karplus, M., and McCammon, J. A. (2002) Molecular dynamics simulations of biomolecules. *Nature Structural & Molecular Biology 9*, 646–652.

[133] Merzel, F., and Smith, J. C. (2002) Is the first hydration shell of lysozyme of higher density than bulk water? *Proceedings of the National Academy of Sciences 99*, 5378–5383.

[134] Makarov, V. A., Feig, M., Andrews, B. K., and Pettitt, B. M. (1998) Diffusion of solvent around biomolecular solutes: a molecular dynamics simulation study. *Biophysical journal 75*, 150–158.

[135] Born, M., and Oppenheimer, R. (1927) Zur Quantentheorie der Molekeln. *Ann. Phys. 389*, 457–484.

[136] Friesner, R. A., and Guallar, V. (2005) Ab Initio Quantum Chemical and Mixed Quantum Mechanics/Molecular Mechanics (QM/MM) Methods for Studying Enzymatic Catalysis. *Annual Review of Physical Chemistry 56*, 389–427.

[137] Lamoureux, G., MacKerell, A. D., and Roux, B. (2003) A simple polarizable model of water based on classical Drude oscillators. *The Journal of Chemical Physics 119*, 5185–5197.

[138] Warshel, A., Kato, M., and Pisliakov, A. V. (2007) Polarizable Force Fields: History, Test Cases, and Prospects. *J. Chem. Theory Comput. 3*, 2034–2045.

[139] Mackerell, A. D. (2004) Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem. 25*, 1584–1604.

[140] Verlet, L. (1968) Computer "Experiments" on Classical Fluids. II. Equilibrium Correlation Functions. *Physical Review 165*, 201–204.

[141] Sagui, C., and Darden, T. A. (1999) Molecular Dynamics Simulations of Biomolecules: Long-Range Electrostatic Effects. *Annual Review of Biophysics and Biomolecular Structure 28*, 155–179.

[142] Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *The Journal of Chemical Physics 98*, 10089–10092.

[143] Liu, Z., Huang, S., and Wang, W. (2004) A Refined Force Field for Molecular Simulation of Imidazolium-Based Ionic Liquids. *J. Phys. Chem. B 108*, 12978–12989.

[144] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics 79*, 926–935.

[145] Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2008) GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *Journal of Computational Chemistry 29*, 622–655.

[146] Hockney, R. W. (1970) Potential calculation and some applications. *Methods Comput. Phys.*

[147] Beeman, D. (1976) Some multistep methods for use in molecular dynamics calculations. *Journal of Computational Physics 20*, 130–139.

[148] Swope, W. C., Andersen, H. C., Berens, P. H., and Wilson, K. R. (1982) A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys. 76*, 637–649.

[149] Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics 23*, 327–341.

[150] Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997) LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem. 18*, 1463–1472.

[151] Hansmann, U. H. E., and Okamoto, Y. (1993) Prediction of peptide conformation by multicanonical algorithm: New approach to the multiple-minima problem. *J. Comput. Chem. 14*, 1333–1338.

[152] Lyubartsev, A. P., Martsinovski, A. A., Shevkunov, S. V., and Velyaminov, P. N. V. (1992) New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. *The Journal of Chemical Physics 96*, 1776–1783.

[153] Hansmann, U. H. E. (1997) Parallel tempering algorithm for conformational studies of biological molecules. *Chemical Physics Letters 281*, 140–150.

[154] Sugita, Y., and Okamoto, Y. (1999) Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters 314*, 141–151.

[155] Okabe, T., Kawata, M., Okamoto, Y., and Mikami, M. (2001) Replica-exchange Monte Carlo method for the isobaricisothermal ensemble. *Chemical Physics Letters 335*, 435–439.

[156] Cheng, X., Cui, G., Hornak, V., and Simmerling, C. (2005) Modified

Replica Exchange Simulation Methods for Local Structure Refinement. *J. Phys. Chem. B 109*, 8220–8230.

[157] Karplus, M., and Kushick, J. N. (1981) Method for estimating the configurational entropy of macromolecules. *Macromolecules 14*, 325–332.

[158] Levy, R. M., Karplus, M., Kushick, J., and Perahia, D. (1984) Evaluation of the configurational entropy for proteins: application to molecular dynamics simulations of an -helix. *Macromolecules 17*, 1370–1374.

[159] Andricioaei, I., and Karplus, M. (2001) On the calculation of entropy from covariance matrices of the atomic fluctuations. *The Journal of Chemical Physics 115*, 6289–6292.

[160] Schlitter, J. (1993) Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chemical Physics Letters 215*, 617–621.

[161] Baron, R., de Vries, A. H., Hünenberger, P. H., and van Gunsteren, W. F. (2006) Comparison of Atomic-Level and Coarse-Grained Models for Liquid Hydrocarbons from Molecular Dynamics Configurational Entropy Estimates. *J. Phys. Chem. B 110*, 8464–8473.

[162] Dolenc, J., Baron, R., Oostenbrink, C., Koller, J., and van Gunsteren, W. F. (2006) Configurational Entropy Change of Netropsin and Distamycin upon DNA Minor-Groove Binding. *Biophysical Journal 91*, 1460–1470.

[163] Missimer, J. H., Steinmetz, M. O., Baron, R., Winkler, F. K., Kammerer, R. A., Daura, X., and van Gunsteren, W. F. (2007) Configu-

rational entropy elucidates the role of salt-bridge networks in protein thermostability. *Protein Science 16*, 1349–1359.

[164] Chang, C.-E., Chen, W., and Gilson, M. K. (2005) Evaluating the Accuracy of the Quasiharmonic Approximation. *J. Chem. Theory Comput. 1*, 1017–1028.

[165] Baron, R., van Gunsteren, W. F., and Hünenberger, P. H. (2006) Estimating the Configurational Entropy from Molecular Dynamics Simulations: Anharmonicity and Correlation Corrections to the Quasi-Harmonic Approximation. *Trends in Physical Chemistry 11*, 87–122.

[166] Rojas, O. L., Levy, R. M., and Szabo, A. (1986) Corrections to the quasiharmonic approximation for evaluating molecular entropies. *J. Chem. Phys. 85*, 1037–1043.

[167] Hnizdo, V., Darian, E., Fedorowicz, A., Demchuk, E., Li, S., and Singh, H. (2007) Nearest-neighbor nonparametric method for estimating the configurational entropy of complex molecules. *J. Comput. Chem. 28*, 655–668.

[168] Li, D.-W., Khanlarzadeh, M., Wang, J., Huo, S., and Brüschweiler, R. (2007) Evaluation of Configurational Entropy Methods from Peptide FoldingUnfolding Simulation. *J. Phys. Chem. B 111*, 13807–13813.

[169] Killian, B. J., Kravitz, J. Y., and Gilson, M. K. (2007) Extraction of configurational entropy from molecular simulations via an expansion approximation. *The Journal of chemical physics 127*, 024107+.

[170] Nguyen, P. H., and Derreumaux, P. (2012) Configurational entropy: an

improvement of the quasiharmonic approximation using configurational temperature. *Phys. Chem. Chem. Phys. 14*, 877–886.

[171] Numata, J., and Knapp, E.-W. (2012) Balanced and Bias-Corrected Computation of Conformational Entropy Differences for Molecular Trajectories. *J. Chem. Theory Comput. 8*, 1235–1245.

[172] Di Nola, A., Berendsen, H. J. C., and Edholm, O. (1984) Free energy determination of polypeptide conformations generated by molecular dynamics. *Macromolecules 17*, 2044–2050.

[173] Bergenstråhle, M., Berglund, L. A., and Mazeau, K. (2007) Thermal response in crystalline Ibeta cellulose: a molecular dynamics study. *The journal of physical chemistry. B 111*, 9138–9145.

[174] Matthews, J. F., Himmel, M. E., and Brady, J. W. (2010) Simulations of the Structure of Cellulose. *1052*, 17–53.

[175] Matthews, J. F., Bergenstråhle, M., Beckham, G. T., Himmel, M. E., Nimlos, M. R., Brady, J. W., and Crowley, M. F. (2011) High-temperature behavior of cellulose I. *The journal of physical chemistry. B 115*, 2155–2166.

[176] Mostofian, B., Smith, J. C., and Cheng, X. (2011) The solvation structures of cellulose microfibrils in ionic liquids. *Interdisciplinary Sciences: Computational Life Sciences 3*, 308–320.

[177] Yui, T., Nishimura, S., Akiba, S., and Hayashi, S. (2006) Swelling behavior of the cellulose Ibeta crystal models by molecular dynamics. *Carbohydrate research 341*, 2521–2530.

[178] Paavilainen, S., Rog, T., and Vattulainen, I. (2011) Analysis of Twisting of Cellulose Nanofibrils in Atomistic Molecular Dynamics Simulations. *J. Phys. Chem. B 115*, 3747–3755.

[179] Hadden, J. A., French, A. D., and Woods, R. J. (2013) Unraveling Cellulose Microfibrils: A Twisted Tale. *Biopolymers 99*, 746–756.

[180] Hadden, J., French, A., and Woods, R. (2014) Effect of microfibril twisting on theoretical powder diffraction patterns of cellulose I. *21*, 879–884.

[181] Gross, A. S., and Chu, J.-W. (2010) On the Molecular Origins of Biomass Recalcitrance: The Interaction Network and Solvation Structures of Cellulose Microfibrils. *J. Phys. Chem. B 114*, 13333–13341.

[182] Shrake, A., and Rupley, J. A. (1973) Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *Journal of Molecular Biology 79*, 351–371.

[183] Muldoon, M. J., Gordon, C. M., and Dunkin, I. R. (2001) Investigations of solvent-solute interactions in room temperature ionic liquids using solvatochromic dyes. *J. Chem. Soc., Perkin Trans. 2* 433–435.

[184] Reichardt, C. (2005) Polarity of ionic liquids determined empirically by means of solvatochromic pyridinium N-phenolate betaine dyes. *Green Chem. 7*, 339–351.

[185] Bhargava, B. L., and Balasubramanian, S. (2007) Refined potential model for atomistic simulations of ionic liquid [bmim][PF[sub 6]]. *The Journal of Chemical Physics 127*, 114510+.

[186] Gary-Weale, A. (2009) Correlations in the Structure and Dynamics of Ionic Liquids. *Australian Journal of Chemistry 62*, 288–297.

[187] Urahata, S. M., and Ribeiro, M. C. (2005) Single particle dynamics in ionic liquids of 1-alkyl-3-methylimidazolium cations. *The Journal of chemical physics 122*, 024511+.

[188] Klein, H. C. R., Cheng, X., Smith, J. C., and Shen, T. (2011) Transfer matrix approach to the hydrogen-bonding in cellulose I[sub alpha ] fibrils describes the recalcitrance to thermal deconstruction. *The Journal of Chemical Physics 135*, 085106+.

[189] Shen, T., and Gnanakaran, S. (2009) The stability of cellulose: a statistical perspective from a coarse-grained model of hydrogen-bond networks. *Biophysical journal 96*, 3032–3040.

[190] Kim, J., Keyes, T., and Straub, J. E. (2010) Generalized Replica Exchange Method. *The Journal of Chemical Physics 132*, 224107+.

[191] Dill, K. A. (1985) Theory for the folding and stability of globular proteins. *Biochemistry 24*, 1501–1509.

[192] Richmond, T. J. (1984) Solvent accessible surface area and excluded volume in proteins. *Journal of Molecular Biology 178*, 63–89.

[193] Connolly, M. L. (1985) Computation of molecular volume. *J. Am. Chem. Soc. 107*, 1118–1124.

[194] Tracy, M. A., and Pecora, R. (1992) Dynamics of Rigid and Semirigid Rodlike Polymers. *Annual Review of Physical Chemistry 43*, 525–557.

[195] Shen, T., Langan, P., French, A. D., Johnson, G. P., and Gnanakaran, S. (2009) Conformational Flexibility of Soluble Cellulose Oligomers: Chain Length and Temperature Dependence. *J. Am. Chem. Soc. 131*, 14786–14794.

[196] Kroon-Batenburg, L. M. J., Kruiskamp, P. H., Vliegenthart, J. F. G., and Kroon, J. (1997) Estimation of the Persistence Length of Polymers by MD Simulations on Small Fragments in Solution. Application to Cellulose. *J. Phys. Chem. B 101*, 8454–8459.

[197] Bianchi, E., Ciferri, A., Conio, G., Cosani, A., and Terbojevich, M. (1985) Mesophase formation and chain rigidity in cellulose and derivatives. 4. Cellulose in N,N-dimethylacetamide-lithium chloride. *Macromolecules 18*, 646–650.

[198] Burchard, W., Seger, B., Habermann, N., Klüfers, P., and Wilhelm, U. (1994) Cellulose in Schweizers Reagens: ein stabiler, polymerer Metallkomplex hoher Kettensteifheit. *Angewandte Chemie 106*, 936–939.

[199] Saalwächter, K., Burchard, W., Klüfers, P., Kettenbach, G., Mayer, P., Klemm, D., and Dugarmaa, S. (2000) Cellulose Solutions in Water Containing Metal Complexes†. *Macromolecules 33*, 4094–4107.

[200] Stoddart, F. *Stereochemistry of Carbohydrates*; John Wiley & Sons Inc, 1971.

[201] Ionescu, A. R., Bérces, A., Zgierski, M. Z., Whitfield, D. M., and Nukada, T. (2005) Conformational Pathways of Saturated Six-Membered Rings. A Static and Dynamical Density Functional Study. *J. Phys. Chem. A 109*, 8096–8105.

[202] Rao, V. S. R. *Conformation of Carbohydrates*; CRC Press, 1998.

[203] Gross, A. S., Bell, A. T., and Chu, J.-W. (2011) Thermodynamics of Cellulose Solvation in Water and the Ionic Liquid 1-Butyl-3-Methylimidazolim Chloride. *J. Phys. Chem. B 115*, 13433–13440.

[204] Harpole, K. W., and Sharp, K. A. (2011) Calculation of Configurational Entropy with a BoltzmannQuasiharmonic Model: The Origin of High-Affinity ProteinLigand Binding. *J. Phys. Chem. B 115*, 9461–9472.

[205] Gross, A. S., Bell, A. T., and Chu, J.-W. (2012) Entropy of cellulose dissolution in water and in the ionic liquid 1-butyl-3-methylimidazolim chloride. *Phys. Chem. Chem. Phys. 14*, 8425–8430.

[206] Bellesia, G., Chundawat, S. P. S., Langan, P., Dale, B. E., and Gnanakaran, S. (2011) Probing the Early Events Associated with Liquid Ammonia Pretreatment of Native Crystalline Cellulose. *J. Phys. Chem. B 115*, 9782–9788.

[207] Wada, M., Nishiyama, Y., Bellesia, G., Forsyth, T., Gnanakaran, S., and Langan, P. (2011) Neutron crystallographic and molecular dynamics studies of the structure of ammonia-cellulose I: rearrangement of hydrogen bonding during the treatment of cellulose with ammonia. *Cellulose 18*, 191–206.

[208] *GAUSSIAN 09*; 2009.

[209] Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput. 4*, 435–447.

[210] Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method. *The Journal of Chemical Physics 103*, 8577–8593.

[211] Bussi, G., Donadio, D., and Parrinello, M. (2007) Canonical sam-

pling through velocity rescaling. *The Journal of Chemical Physics 126*, 014101+.

[212] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics 81*, 3684–3690.

[213] *MATLAB 7.12.0*; 2011.

[214] Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *Journal of molecular graphics 14*.

# Appendix

# A) Fiber Simulations

A 36-chain I$\beta$-cellulose crystal fiber (Figure 1a) with 20 glucose units per chain was solvated in an equilibrated cubic box of BmimCl ion pairs or TIP3P[144] water molecules. After deleting overlapping molecules and energy minimization, the fiber was surrounded by $\sim$25 Å of solvent on both sides along the direction of its polymerization axis and by 30-40 Å of clearance around the other surfaces. These large values correspond to a few head-to-tail distances of BmimCl, thus allowing for a reliable analysis of solvent structures close to the fiber surfaces. The total number of atoms is approximately 180,000 in the BmimCl simulation and around 160,000 in the water simulation. GLYCAM06[145] molecular mechanics force field parameters were used for the cellulose. BmimCl parameters had been developed previously[143] and partial charges were assigned through fitting to electrostatic potentials computed at the Hartree-Fock level with 6-31G* basis set using the GAUSSIAN programm (see Table 2)[208]. Simulations were performed and analyzed with GROMACS[209]. Long-range electrostatic interactions were calculated using the particle-mesh Ewald summation[142, 210] with a real-space cut-off at 12 Å. Van der Waals interactions were cut off at 10 Å.

The systems were equilibrated to the target temperatures of 450 K (BmimCl) and 375 K (water) which correspond to the corresponding biomass pretreatment conditions. The temperature was maintained by velocity rescaling with a stochastic term to ensure proper sampling[211]. After equilibrating the density with a short NPT simulation applying the Berendsen barostat[212], the system was subjected to 5 ns of NVT equilibration. The production run was performed for over 100 ns with a time step of 2 fs in the NVT ensemble. Bonds

**Table 2** Partial charges for $Bmim^+$. Atom names correspond to atoms shown in Figure 3. Hydrogen $H_{Cx}$ is the hydrogen bonded to carbon Cx.

| atom name | q(e) | atom name | q(e) |
|---|---|---|---|
| C8 | -0.07130 | C1 | -0.00550 |
| $H_{C8}$ | 0.02940 | $H_{C1}$ | 0.22580 |
| C7 | 0.03090 | N2 | 0.05960 |
| $H_{C7}$ | 0.01570 | C4 | -0.08460 |
| C6 | 0.01070 | $H_{C4}$ | 0.10850 |
| $H_{C6}$ | 0.02040 | C2 | -0.14260 |
| C5 | -0.01530 | $H_{C2}$ | 0.23400 |
| $H_{C5}$ | 0.07960 | C3 | -0.21830 |
| N1 | 0.06820 | $H_{C3}$ | 0.26330 |

to hydrogen atoms were constrained using the LINCS algorithm[150].

To identify the arrangement of $Bmim^+$ cations around surface glucose monomers, a set of three-dimensional Euler angles ($\theta_x$, $\theta_y$, $\theta_z$) was derived for every cation with a maximum imidazole ring-center to pyranose ring-center distance of 7 Å. Euler angles were computed from rotational matrices and correspond to successive rotations of the cations around the z-, y-, and x-

axis of their coordinate system (Figure 3) in order to transform them to the reference coordinate system of the adjacent glucose unit (Figure 1b). Three-dimensional histograms were derived with a bin size of 7.2° (50 bins) in each direction of the orientation space. These orientation histograms were normalized and smoothed with a three-dimensional Gaussian function in order to remove statistical noise. Histograms were created with MATLAB[213] and orientations illustrated with VMD[214]. Cation orientations were evaluated with respect to the 16 central glucose units of any surface glucan chain. Terminal cellobiose units were neglected in this analysis and in the computation of all radial distribution functions in order to eliminate potential artifacts due to their greater flexibility.

## B) Replica-Exchange Simulations

A cellulose decamer was solvated in a cubic box with 1,330 BmimCl ion pairs and in a dodecahedral box with about 11,500 water molecules. Both systems comprised approximately 35,000 atoms. The solvated systems were first equilibrated at 350 K for 100 ps in the NVT ensemble and at the corresponding REMD target temperatures for 300 ps in the NPT ensemble. All force field and simulation parameters were identical to the fiber simulations.

REMD simulations were performed in the NPT ensemble with a pressure correction term in the exchange probability[155]. The lowest REMD temperature for both systems is 350 K. NPT-simulations of water solutions at elevated temperatures are known to induce volume fluctuations which are too large for accurate replica exchange. Hence, the maximum temperature used in this study for the cellulose-water system is 513 K, while that for the cellulose-BmimCl system is 550 K. Replicas are exchanged at every picosecond with

the probability $P_{ex}$ given in equation (3.6). In order to reach an average exchange probability between 10%-20% for replicas at any two neighboring temperatures, 68 replicas at temperatures increasing according to an exponential distribution were simulated for the cellulose-water system and 60 such replicas were simulated for the cellulose-BmimCl system. Both systems were run over 100 ns resulting in a total simulation time of around 13 $\mu$s for the REMD study.

The cellulose configurational entropy was calculated using the quasi-harmonic method with corrections derived by Di Nola[172] as given in equation (3.10). The entropy was evaluated in torsional space resulting in 120 degrees of freedom for the entire cellulose decamer. In these, and all other profiles, mean values were computed as block averages over the simulation time and error bars (if given) correspond to one standard deviation.

# Vita

Barmak Mostofian was born in Tehran, Iran, in 1982. At the age of five, he moved with his parents to Germany, where he spent his school years. In 2001, he graduated from high-school, winning a national prize for his academic achievements in physics. Barmak pursued studies in Molecular Biotechnology at the University of Heidelberg. During this time, he has been an Erasmus-fellow at the bioinformatics group of the University of Manchester and he visited the University of Toronto for collaborative work before receiving the M.Sc. degree in 2007. In 2008, he joined the Center for Molecular Biophysics at the Oak Ridge National Laboratory and the Genome Science & Technology program at the University of Tennessee to obtain his doctoral degree in life sciences.