UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Sciences des Polymères**

Arrêté ministériel : 7 août 2006

Présentée par

Pan CHEN

Thèse co-dirigée par **Jean-Luc PUTAUX** et **Karim MAZEAU**

préparée au sein du **Centre de Recherches sur les Macromolécules Végétales (CERMAV), UPR 5301 CNRS** dans **l'École Doctorale Chimie et Sciences du Vivant**

Simulation de cristaux de cellulose par dynamique moléculaire

Thèse soutenue publiquement le **13 Juin 2013** devant le jury composé de :

Madame Anne MILET
Professeur, Département de Chimie Moléculaire, Grenoble (présidente)
Monsieur Trevor FORSYTH
Professeur, Université de Keele, Angleterre (rapporteur)
Monsieur Lars BERGLUND
Professeur, KTH, Royal Institute of Technology, Stockholm, Suède (rapporteur)
Monsieur Jean-Luc PUTAUX
Directeur de Recherche, CERMAV-CNRS, Grenoble (co-directeur de thèse)
Monsieur Karim MAZEAU
Directeur de Recherche, CERMAV-CNRS, Grenoble(co-directeur de thèse)
Monsieur Yoshiharu NISHIYAMA
Chargé de Recherche, CERMAV-CNRS, Grenoble (co-encadrant)

Université Joseph Fourier

献给我的父母和兄弟及平凡

To my parents, my brother

and Fan Ping

Acknowledgements

The Chinese Scholarship Council (CSC) is acknowledged for the financial support of my 45-months stay in France. The French Agence Nationale de la Recherche is acknowledged for providing travelling expenses to attend five international conferences.

This work has been carried out under the supervision of Dr. Jean-Luc PUTAUX, Dr. Yoshiharu NISHIYAMA and Dr. Karim MAZEAU.

Four years ago, I was very lucky to be accepted to be a PhD student in CERMAV and Université Joseph Fourier by Dr. NISHIYAMA and Dr. PUTAUX.

Dr. NISHIYAMA introduced me to the cellulose community, explained me the scientific way to answer questions, taught me to solve many technical problems in modeling and programing, introduced the famous scientists that worked on cellulose to me, used his own experience in science to encourage me to overcome the difficulties and to develop my own subject. Discussions with him sometimes lasted for hours and problems always became easier to be solved. Not only in the PhD subject but also in the daily life, I benefited from his help. He provided the financial support to me for studying French courses in CUEF and helped me a lot to look for an apartment. What he has offered is largely over the responsibility as a co-supervisor. I deeply and sincerely appreciate his generous help through my whole PhD progress. In addition, thanks for his invitation during the weekends and I really enjoyed the time with his friendly family members.

I also would like to greatly thank Dr. Jean-Luc PUTAUX, my supervisor, who has the critical contributions on accepting me and on the frequent correction of all the chapters of my thesis and all of the abstracts for conferences. Moreover, I also appreciate him for handling almost all of the paperwork from the beginning to the end of the PhD period. Thanks for his kindness and patience. It’s a pity that I didn’t have additional time to perform any experiment that I desired with him.

Dr. Karim MAZEAU, my co-supervisor, who has rich experience in polysaccharide modeling, has given me advice and many scientific ideas and views on cellulose modeling. Discussion with him always inspired me and solution became clear. I am extremely grateful to him for his great help in correcting and writing publications and thesis. I also appreciate him as well

as Dr. NISHIYAMA and Dr. PUTAUX for kindly supporting me to present my work in many international conferences and local meetings. Without this training, I would not be able to fluently introduce my work with confidence.

Thanks for the members of my "comité de these", Dr. Hélène JAMET, Dr. Serge CROUZY and Dr. Trevor FORSYTH, who had all followed my PhD work and provided useful advice.

I would like to thank Dr. Henri CHANZY for the English correction on chapter I and on one of my publications, and for his advice and encouragement to my oral presentations.

I would like to thank Dr. Landry CHARLIER, who gave me technical help in modeling when I just started this PhD, and also for his endless and overstate jokes.

I was very happy to stay in the "Glycomaterials" team and very cheerful to chat with Dr. Laurent HEUX, Dr. Bruno JEAN, Dr. Sonia MOLINA-BOISSEAU, Mr. Pierre SALLIER as well as all the other members, on whichever topics. Many thanks go to Marie-France METRAL and Christine LANCELON-PIN for their warm greetings every morning during the tea break.

Last but not least, I would like to thank all the friends I met, without whom the life in France would have been less colorful. I may not write down all of name here one by one, but you must be able to feel for whom I am writing in this paragraph. All I want to say is: I owe your many thanks. We have been sharing many activities together (traveling, ski, BBQ, dinner and so on), so long that I have forgotten the first moment we met. In the future, we may not be able to see each other so frequently and conveniently as now, but what I am sure is that, no matter where I go, no matter what I do, these beautiful memories will accompany me all the time for the rest of my life. There will always be a place, right in my heart, for all of you guys. Cheers!! Friendship lasts forever!!

Table of contents

Introduction

Cellulose is a structural polysaccharide that widely occurs in lower and higher plants, some bacteria, fungi, amoebae, even in some sea animals. Cellulose–based products such as wood, cotton, hemp, etc., have been used by humans as important raw materials or tools for thousands years. Until now, the application of cellulose is still strongly connected with the daily life of human beings, although the modern society is simultaneously searching for other high performance materials.

The renewability and biodegradability of cellulose-based products meets the eco-friendly requirements for alternative biosourced materials. For this reason, extensive research is performed on cellulosic products, as denoted by the increasing number of articles published on the related topics.

Cellulose is a polymer consisting of β-(1,4)-linked glucosyl units. Since 1913, when its first X-ray diffraction analysis was performed, the crystal structure of cellulose has been studied using both theoretical and experimental approaches, with tools such as molecular modeling, X-ray, neutron and electron diffraction, solid-state nuclear magnetic resonance (CP/MAS ^{13}C NMR), infrared and Raman spectroscopy and microscopy imaging (atomic force, scanning and transmission electron microscopy). Several of these techniques dedicated to polymer were first applied to cellulose. However, no single technique could completely unravel the full structural details of cellulose, since each method has its own limitations. The structural complexity of cellulose, from the nano size of its crystals to the coexistence of crystalline and amorphous phases, is another hurdle that needs to be overcome. It is only a combination of several techniques that will be able to allow unraveling the mysteries of cellulose.

Despite the fact that today, more features have been understood about cellulose, there are still many controversies about its structure and ultrastructure. To solve these, one solution is to develop new techniques to overcome the limitation of current experimental approaches. In this context, molecular modeling appears to be most promising. This technique, which is based on physical chemistry concepts, utilizes the capability of computers. While molecular modeling of cellulose has been used for around for 50 years, it is only today that rapid improvement in computer hardware together with new efficient algorithms, allow the efficient handling of

systems containing millions of atom agitated within the dynamics of microseconds.

In the molecular modeling field, many methods (French 2012), such as quantum mechanics (QM), molecular mechanics (MM) and molecular dynamics (MD), can be selected, depending on the size and properties of the target system, the main constraints coming from computational speed. Considering the balance between accuracy and efficiency, MD simulation and analysis based on empirical force fields were carried out in this work. Thus, this thesis is a numerical study of the structure-properties relationships of crystalline cellulose by using force field-based MD simulation techniques.

Using MD simulation with the GROMOS force fields, we have studied the structural and physical properties of four types of crystalline cellulose (namely I_α, I_β, II and III_I), including the structural details of bulk and surface regions, thermal behavior, mechanical properties and allomorphic interconversion. In this work, the reasonability of the employed methods was validated by directly comparing simulation results and experimental observations. In particular, the discrepancy existing between our simulations and some experimental results will be discussed to better understand how the force field parameters and computational methods affect the simulation results, in order to fill the gap between experiment and theory.

This thesis manuscript is organized as follows. Chapter I briefly introduces the background knowledge on cellulose and modeling related to this work. Chapter II lists the computational models, the methods and procedures together with the data analysis tools that have been designed or used in this work. Chapter III presents our results on the crystal structures and thermal properties of crystalline cellulose by using GROMOS 53a6 and GROMOS $56A_{carbo}$ force fields. Chapter IV points out the force field parameters that are crucial to model the thermodynamic properties and crystal structures of cellulose. Three parameters were optimized against experimental observations. In Chapter V, we have 1) explained the physical origin of the temperature-induced phase transition of cellulose I, 2) presented the results on the stability of hydrogen bonding in cellulose I simulated using the infinite approach, 3) described the studies on the $I_\alpha \rightarrow I_\beta$ and $III_I \rightarrow I_\beta$ allomorphic conversions of cellulose by thermal treatment and the structure variation of cellulose II during annealing, 4) simulated the structure of different models of cellulose in water.

Chapter I – Current knowledge on crystalline cellulose and related molecular modeling studies

I.1. The crystal structures of cellulose

Crystalline cellulose occurs under many allomorphic forms, namely cellulose Iα, Iβ, II, III_I, III_{II}, IV_1, IV_{II}, etc. Numerous studies have been devised for the refining of the structure of these allomorphs. Recently, the combined use of synchrotron X-ray and neutron diffraction applied to samples of very high crystallinity has allowed proposing refined structures for cellulose Iα and cellulose Iβ from natural origin. The same type of approach was also applied to cellulose II and III_I, obtained by chemical transformation from cellulose I. Thanks to the selective deuteration of the crystalline cores of these samples, it has become possible for the first time to visualize the hydrogen-bonding network holding the structure. (Langan *et al.* 1999, 2001; Nishiyama *et al.* 2002, 2003; Wada *et al.* 2004). The study of the crystalline structure of cellulose and its modeling implies some knowledge on the geometry of the glucosyl units: the building blocks of the cellulose chain, the conformation angle linking the glucosyl residues together, their packing into unit cells, etc.

I.1.1. Geometry of the glucosyl residue: the ring puckering

The conformation of a six-member aliphatic ring, such as that of a glucopyranosyl unit, can in principle occur as a "boat", a "half chair", a "chair" or a "skew". These conformations can be described by the Cremer-Pucker puckering parameters (Cremer and Pople 1975)(**Figure I.1**): the amplitude Q and two angles Θ and Φ. With the parameter Θ close to 0°, the ring is in the classical $^4C_1$ conformation, which is the most stable for glucose. In crystalline cellulose, the glucoses adopt exclusively the $^4C_1$ conformation.

I.1.2. The glucosidic linkage torsion angles

The rotation around the glycosidic bonds in polysaccharides is their principal source of flexibility. In cellulose, it is described by the two dihedral angles: O5-C1-O1-C4 (ϕ) and C1-O1-C4-C5 (ψ), which have some torsional freedoms (**Figure I.2**). Torsional freedom is hindered by the formation of hydrogen bonds between two adjacent residues; one pair of ϕ and ψ is thus favored, yielding the linear motif for cellulose chains.

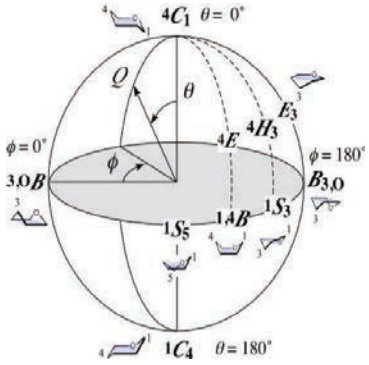

Figure I.1. Puckering parameters and the corresponding conformation of glucose ring (C: chair, B: boat, S: skew) *(from internet)*.

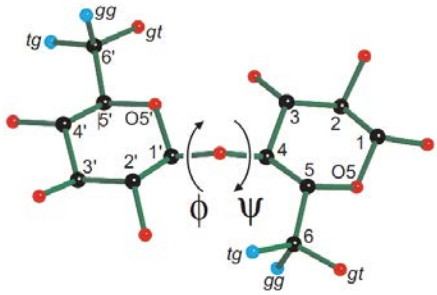

Figure I.2. Definition of the glucosidic linkage angles and the three staggered conformations of the hydroxymethyl group of cellulose (French and Johnson 2006)

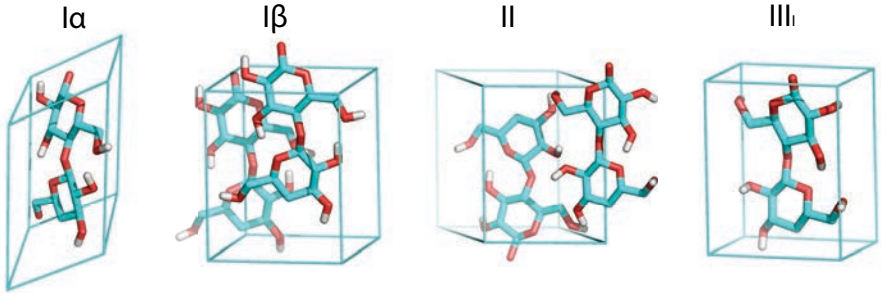

Figure I.3. Unit cells of crystalline cellulose with explicit position of the hydrogen atoms involved in the hydrogen bonding.

I.1.3. Conformation of the hydroxymethyl groups

The conformation of hydroxymethyl group of cellulose is usually described by the dihedral angle ω around the C5-C6 single bond (by the sequence of atoms O5-C5-C6-O6), as shown in **Figure I.2.** The three staggered conformations of the hydroxymethyl groups are named trans-gauche (*tg*: ω from 120 to 240°), gauche-gauche (*gg*: ω from 240 to 360°) and

gauche-trans (*gt*: ω from 0 to 120°). In cellulose Iα and Iβ, the hydroxymethyl groups are in *tg* conformation, whereas those in cellulose II and III_I are in *gt* conformation.

I.1.4. Unit cells and chain staggering

Cellulose Iα has a triclinic unit cell and a P1 space group, containing a unique chain with two different glucosyl residues (**Figure I.3**). Cellulose Iβ is crystallized in a monoclinic $P2_1$ space group with two independent chains, named center chain and origin chain (or corner chain), each located on one of the 2_1 screw axes of the unit cell, the conformation of the center chain being slightly different from that of the corner counterpart. Along the chain direction, the chains are staggered with a translation of c/4 in the triclinic unit cell of Iα and of ±c/4 in the monoclinic cell of Iβ. The chains in both crystals are arranged in a parallel-up fashion.

Cellulose II also has a monoclinic unit cell and a $P2_1$ space group containing two independent chains (corner and center), while cellulose III_I has a monoclinic unit cell in the space group of $P2_1$, containing a single chain. The chains in cellulose II are arranged in an anti-parallel manner in contrast with the parallel organization of cellulose I. In cellulose III_I, the chains are arranged in a parallel-up fashion. Along the chains direction (*c*-axis), the chains of cellulose II are staggered with ±c/4 difference, whereas the chains in III_I have no staggering difference.

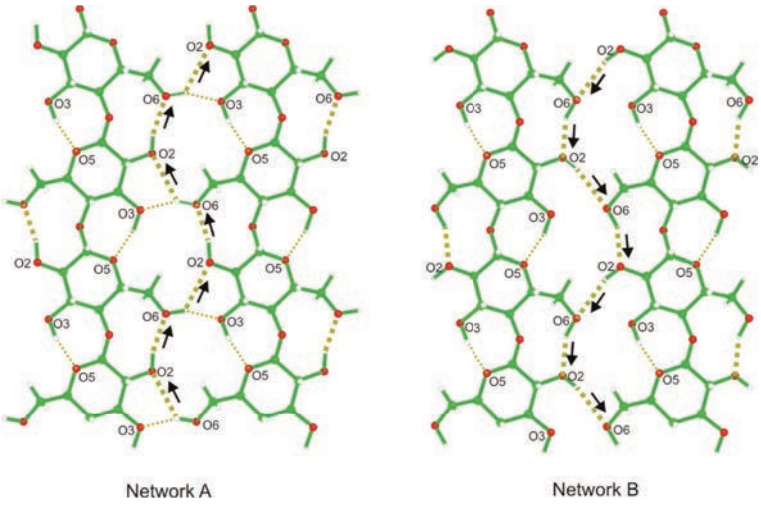

Figure I.4. Two types of hydrogen-bonding network, pattern A and pattern B, of native cellulose Iβ (Nishiyama *et al.* 2008).

I.1.5. Structure of the intra-chain hydrogen bonds

The hydrogen-bonding network in cellulose is highly correlated with the conformation of their hydroxymethyl groups. Two types of hydrogen bonds are distinguished: intra-chain and inter-chain hydrogen bonds. In cellulose I, there are two intra-chain hydrogen bonds between two adjacent residues: O3-HO3···O5 and O2-HO2···O6. In cellulose II and III_I, there are also two types of intra-chain hydrogen bond: O3-HO3···O5 and HO3-O3···O6.

I.1.6. Structure of the inter-chain hydrogen bonds

The two hydrogen atoms (HO2 and HO6) of cellulose I were found to be delocalized into two locations each. Thus, two hydrogen-bonding patterns occur, named pattern A and pattern B (**Figure I.4**), respectively. In pattern A, cellulose I contains one strong inter-chain hydrogen bond: O6-HO6···O3 and the other weak HBs O6-HO6···O2 that only occurs in the glucose plane. Every residue can form one of this hydrogen bond in average. In the direction normal to the glucose plane, there are a number of inter-chain hydrogen bonds formed between aliphatic hydrogen and oxygen atoms. However, they are considered to be very weak compared with the O-H···O hydrogen bonds located within the cellulose sheets. Thus, the mean glucosyl plane is also called the hydrogen-bonding plane in $I\alpha$ and $I\beta$.

By using the neutron diffraction technique applied for the refinement of cellulose I, the location of the hydrogen atoms involved in the hydrogen bonds in both II and III_I were found to have unique position without structural disorder. Different from cellulose I, the inter-chain hydrogen bonds in cellulose II are more complex. They can be separated into three types: origin-origin, center-center, and origin-center hydrogen bonds. The origin-origin hydrogen bond only includes O2-HO2···O6 and the center-center counterpart only includes O6-HO6···O2. The origin-center hydrogen bond includes $(O2\text{-}HO2)_c\cdots(O2)_o$ and $(O6\text{-}HO6)_o\cdots(O6)_c$. In addition, several relative weak inter-chain HB exists, such as $(HO6\text{-}O6)_o\cdots(O3)_c$ and $(O6\text{-}HO6)_o\cdots(O5)_c$. Note that the subscript o and c stand for the origin and center chain, respectively. In cellulose III_I, there are two types of inter-chain hydrogen bonds, namely: HO2-O2···O6 and HO6-O6···HO2.

I.2. Molecular modeling

Molecular modeling is a method based on the principles of theoretical physical chemistry, which uses the capacity of computers to study at the atomic scale the properties, aiming to complement the experimental measurement. It has been carried out to study structural features of cellulose since the 1980s, mainly helping to predict structural and mechanical properties and providing an understanding of the origin of these features.

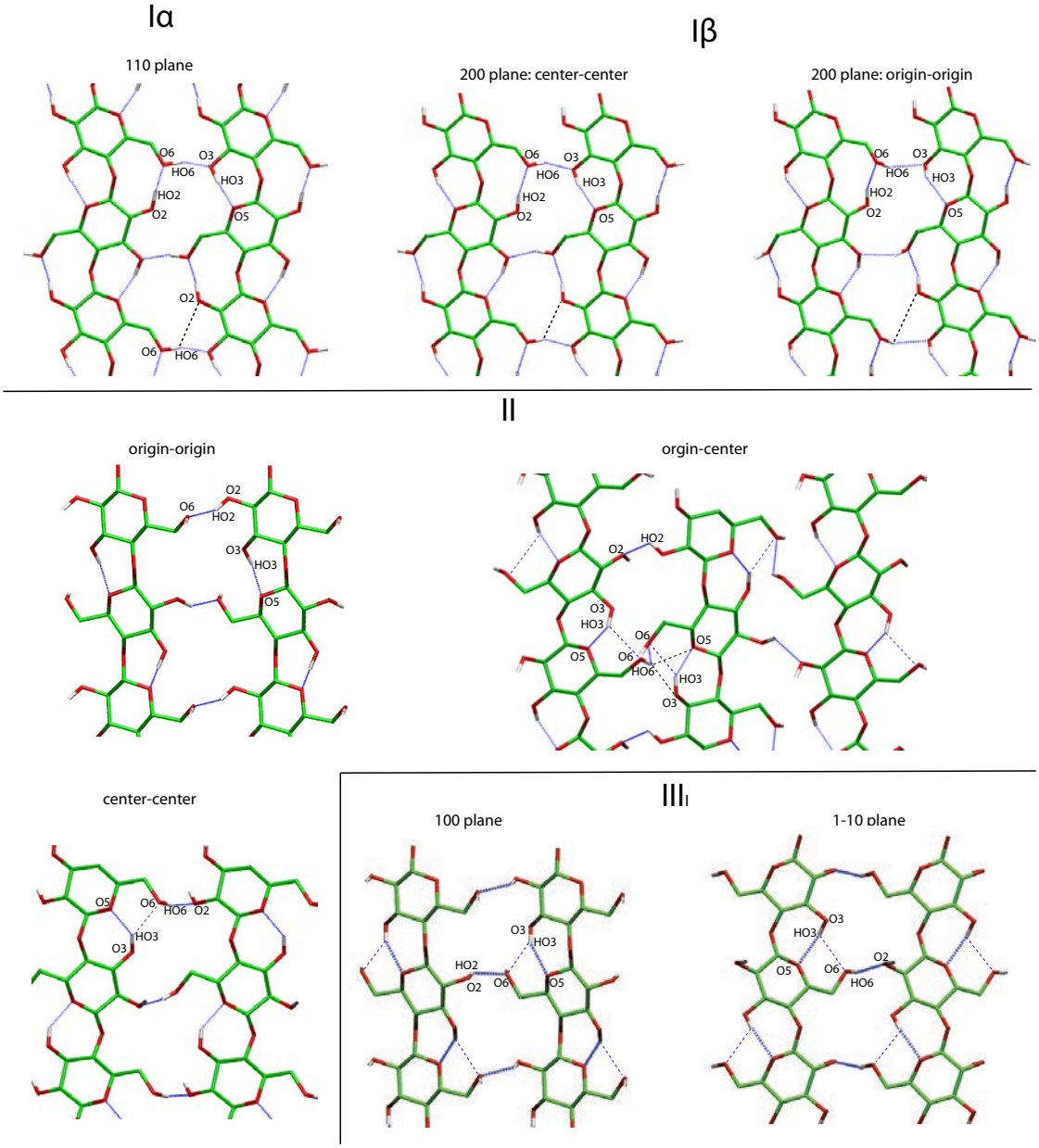

Figure I.5. The hydrogen-bonding networks in cellulose allomorphs. The main hydrogen bonds are indicated by blue dotted lines, while the weak HBs are indicated by black dotted lines.

Based on the given properties of a system, several computational techniques are available, such as quantum mechanic calculation, atomistic force-field method, coarse-grain models, etc. (**Figure I.6**). Preferentially, it is necessary and important to choose a proper technique or a combination between different techniques. Considering that our computational resources are limited and that our target systems are from 3 to 30 nm in dimension, atomistic force field methods were selected for this thesis.

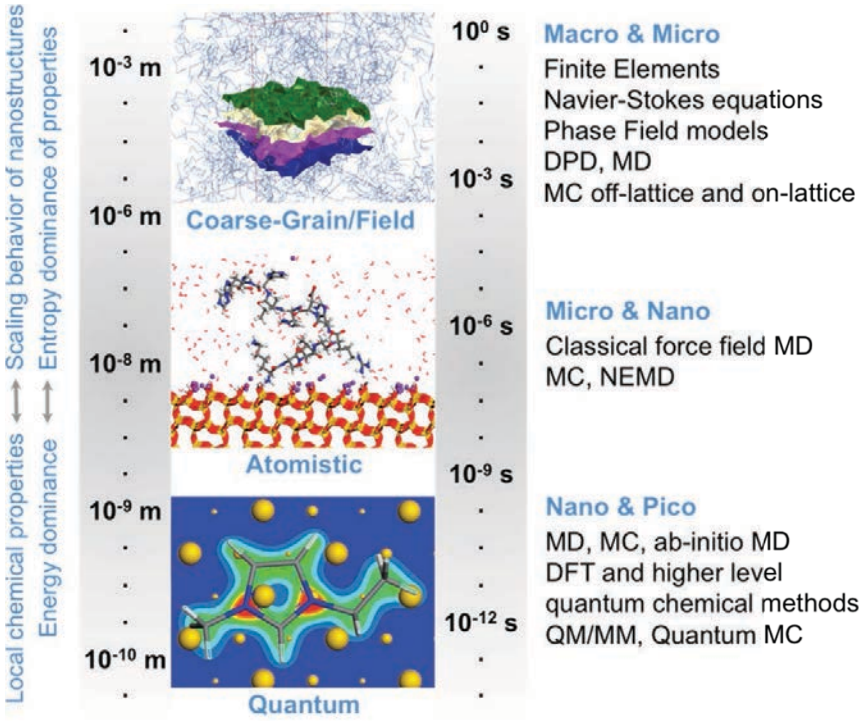

Figure I.6. Illustration of simulation techniques with associated system sizes and timescale. (Heinz *et al.* 2012)

I.2.1. Atomistic force field

An atomistic force field is used to describe the rules of interaction between atoms and to measure the energy of a system. The force field contains a series of functions of energy profiles and the parameters defined within those functions. The popular force fields used for cellulose include CHARMM (Gross *et al.* 2011; Gross and Chu 2010; Matthews *et al.* 2006, 2011, 2012), GLYCAM (Chundawat *et al.* 2011; Yui and Hayashi 2007; Yui and Hayashi 2009; Yui *et al.* 2006; Zhang *et al.* 2011) GROMOS (Bergenstråhle *et al.* 2007, 2008, 2009; Kroon-Batenburg *et al.* 1996), PCFF (Mazeau 2005; Mazeau and Rivet 2008) COMPASS

(Eichhorn and Davies 2006, Eichhorn and Young 2001), MM3 (Viëtor *et al.* 2000) and so on. In general, the atomistic force field can be divided into two types: the full-atoms force fields and the united force fields. The difference between them is that the formers treat explicitly the aliphatic hydrogen atoms but the latter do not. Two united force fields, GROMOS 53a6 (Oostenbrink *et al.* 2004, Lins and Hünenberger 2005) (Gr53a6FF) and GROMOS 56A$_{carbo}$ (Hansen and Hünenberger 2011) (Gr56A$_{carbo}$FF), in which the carbon and the linked aliphatic hydrogen atom are considered as a single particle, are used in this thesis.

I.2.2. Molecular mechanics (MM) and molecular dynamics (MD) simulations

Both MM and MD simulations have been used. MM aims at optimizing the total potential energy of a system, with the goal of providing static configuration that have the lowest energy. With MD, one looks at the modeling of the atomic motion over time. This technique yields structures that are more close to the actual molecular organization. The first MD simulation, which was carried out to study the phase transition of a hard-sphere system was achieved in 1957 (Alder and Wainwright 1957). It has since become a standard tool.

I.2.3. Advantage and limitation of MD simulation

MD simulation studies the properties of cellulose at atomic scale, with providing visualized microscopic information hardly detectable experimentally. In particular, MD should be able to (i) predict the unknown structural details of cellulose (*i.e.*, the surface structure), (ii) provide a fundamental understanding of the origin of the allomorphic conversion (*i.e.*, the cellulose Iα to Iβ in solid-state transition) and (iii) help for illustrating the molecular mechanism of mercerization of cellulose I, etc.

The main limitation of MD may come from two aspects. One is the employed force field parameter that may lack of accuracy. The other is the simulation time and the size of the modeled system that is respectively much shorter and smaller compared to what is observed in reality. More details can be found in the recent reviews papers (Foley *et al.* 2011; Mackerell 2004; van Gunsteren *et al.* 2006; Jorgensen and Tirado-Rives 2005).

I.3. Background

I.3.1. Controversies about the cellulose structure

I.3.1.1. Parallel-up versus parallel-down structures

The structure of crystalline cellulose is often studied using X-ray diffraction of polycrystalline samples. However, the number of observed intensities is, in most cases, too limited to determine the structure at atomic resolution from diffraction data only. For example, the parallel-up and parallel-down structures of cellulose (**Figure I.7**), which only differs in the value of the monoclinic angle, have been debated for years due to the limited diffraction data available.

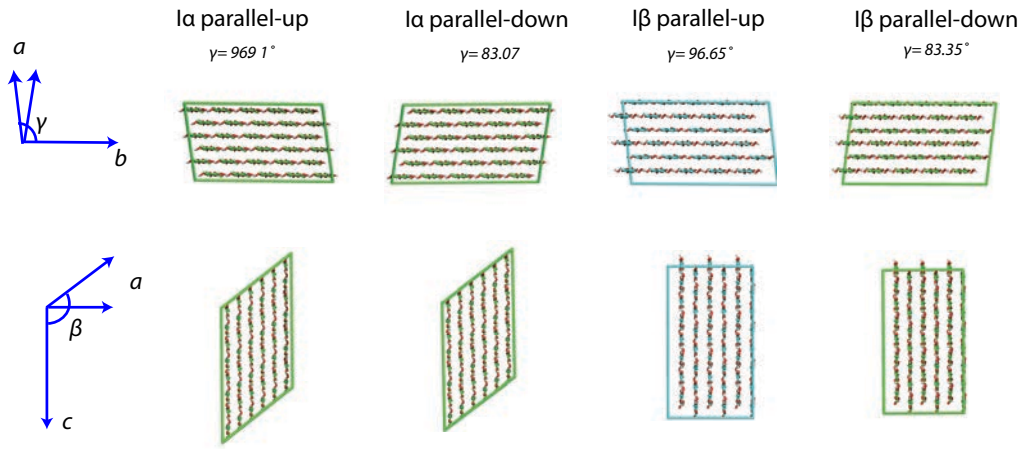

Figure I.7. Representation of parallel-up and parallel-down structures of cellulose Iα and Iβ. In the upper drawings, the chains are seen end-on while they are seen edge-on in the lower drawings.

I.3.1.2. Structures based on parallel or anti-parallel chains

The mechanism of solid-state allomorphic conversion from the parallel cellulose I to the antiparallel cellulose II by regeneration or mercerization has been a continuing controversial topic since the structural definition of these two allomorphs. By using MD simulation, the number of studies on cellulose II is much less than that on cellulose I.

I.3.1.3. Coexistence of cellulose Iα and Iβ

Cellulose Iα and Iβ always coexist in the same specimens, but the ratio of these two entities depends on the source from which they are extracted. Before the recognition of this duality, it

was indeed difficult to get any agreement between the various structural studies dealing with different types of celluloses. It was only in 1984 that ^{13}C solid-state NMR clearly identified the duality: cellulose from *Valonia* and bacterial origin were rich in cellulose Iα, while those of tunicin or higher plants were dominated by the Iβ phase.

I.3.1.4. Nanodomains in cellulose Iα and Iβ

The location of Iα and Iβ within the cellulose microfibrils may affect the Iα to Iβ solid-state conversion Iα→Iβ. Three types of models have been proposed for the coexistence of Iα and Iβ domains: (i) Iβ core surrounded by Iα (Wada and Okano 2001); (ii) alternating Iα and Iβ along the same microfibril (Horikawa and Sugiyama 2009); (iii) alternating Iα and Iβ in the crosses section (Imai and Sugiyama 1998).

I.3.1.5. Description of the surface of native cellulose crystals

^{13}C solid-state NMR spectra give different signals for the surface and interior structure for cellulose (Saito *et al.* 2009; Vietor *et al.* 2002). By using TEM (Brown and Wladkowski 1996) a square shape of cellulose microfibril was observed for large cellulose crystals of from the wall of *Valonia*. The shape was rectangular for the crystals of *Micrasterias* and lozenge-like for those from the mantle of tunicates. High resolution AFM measurement shows the highly organized pattern at the surface of *Valonia* microfibrils (Baker *et al.* 2000). Unfortunately, no detail is available for the shape and the surface of the microfibrils of smaller diameter, for which different models have been proposed, but not proven so far.

I.3.2. Anisotropic thermal behavior of crystalline cellulose

The thermal behaviors of four allomorphs of cellulose have been experimentally investigated by following the change in the *d*-spacings of the X-ray diffraction patterns as a function of temperature. In these studies (Wada 2002; Hori and Wada 2005, 2006; Wada *et al.* 2003, 2010,), the anisotropic thermal expansion properties were found for all the allomorphs. For example, the thermal expansion in the *a*-axis of Iβ is larger than that of the *b*-axis. This was explained by comparing the hydrogen-bonding network along the *b*-axis with the one along the

a-axis is determined by Van der Waals interactions. Several reports (Langan *et al.* 2005; Matthews *et al.* 2012) seem to agree with such an explanation. It was also found that the thermal expansion in the *b*-axis of cellulose Iβ depended on the source of cellulose. Both positive and negative thermal expansion coefficients (TECs) were observed (Hori and Wada 2005; Wada *et al.* 2010). This difference was explained by considering the nanoscale size of the crystal in the corresponding specimens.

I.3.3. Allomorphic conversions of cellulose

When cellulose Iα is heated above 260°C, still below the pyrolysis temperature, either by a simple thermal treatment in helium environment or by hydrothermal treatment in various media, it irreversibly converts to Iβ, as confirmed by solid-state NMR and X-ray diffraction measurements (Debzi *et al.* 1991; Wada *et al.* 2003). It was reported that such conversion was not complete, with a few percent of the Iα phase remaining in the annealed structure. Based on their structural similarity, two types of conversion mechanism, either chain translation or chain rotation, were proposed. MM was also performed to study this process. However, none of them could definitely describe the conversion mechanism. It was due to this irreversible conversion as well as the slightly higher energy of Iα, when compared to Iβ-calculated by molecular modeling (Heiner *et al.* 1995; Kroon-Batenburg and Kroon 1997; Kroon-Batenburg *et al.* 1996) that the Iβ phase is regarded to be more thermodynamically stable.

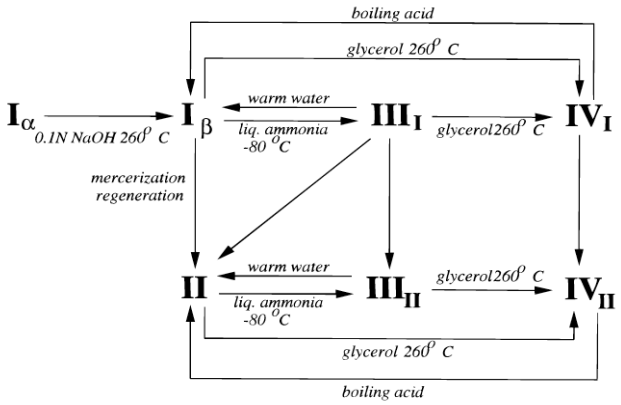

Figure I.8. Allomorphic conversions between the various cellulose allomorphs (Kroon-Batenburg *et al.* 1996). In this scheme, cellulose III_{II} remains questionable.

Cellulose II could be prepared from cellulose I either by regeneration or mercerization process. Due to the irreversible process, cellulose II is thought to be more stable than cellulose I (Langan *et al.* 2001). During this conversion, the formation of cellulose II required the chain to rotate by 180° since X-ray determinations have confirmed that the cellulose chains are initially parallel in cellulose I but antiparallel in cellulose II. This process seems unlikely since it should be energetically unfavorable, especially in the mercerization process where the intracrystalline structure was swollen but not dissolved.

The reversible transformation between cellulose I and cellulose III_I is another documented solid-state conversion of cellulose. When immersing native cellulose samples into liquid NH_3 at -80°C, the NH_3 molecules penetrate into the cellulose lattice to from the crystalline cellulose ammonia complex. By bringing this complex at room temperature, the NH_3 molecules evaporate and the cellulose is converted into the metastable cellulose III_I, which in turn can revert to cellulose I at high temperature. These successive transformations, which involve substantial translation and movements of the cellulose chains, have been studied, not only by X-ray, neutron and electron diffraction analysis, but also by MD, using the GLYCAM (Bellesia *et al.* 2011) force field, which has allowed describing the transformation at the atomistic level.

I.3.4. High temperature structures of cellulose I

Upon heating native cellulose to 500 K, a “high temperature phase” can be obtained. Synchrotron X-ray diffraction (Hidaka *et al.* 2010) has shown that this high temperature structure has a monoclinic unit cell containing two independent chains, with a $P2_1$ space group. MD simulations (Bergenstråhle *et al.* 2007; Matthews *et al.* 2011; Zhang *et al.* 2011) by using CHARMM C35, GLYCAM O6 and GROMOS 45a4 force fields were carried out to predict the high temperature structures. All of them agree that in such a phase, the conformation of hydroxymethyl group is a mixture of *gg* and *gt*.

I.3.5. Elastic modulus estimation

The high modulus of cellulose together with a relatively low density have attracted a number of studies dealing with the incorporation of cellulose as reinforcing phase in nanocomposites.

However, due to the small size of cellulose crystallites and the limitation of the techniques, the "true" crystal modulus of cellulose remains ambiguous.

The elastic modulus of native cellulose (Moon *et al.* 2011) has been evaluated by various techniques including tensile strength testing, X-ray diffraction, Raman spectroscopy, AFM and molecular modeling. From these techniques, modulus values ranging from 5 to 220 GPa have been obtained (Moon *et al.* 2011). Such differences are either due the difference in sample origin, the fiber dimensions or the employed technique. It is generally accepted that the modulus of an independent microfibril is around 130 GPa. It should be noted here that, a number of studies was achieved based on molecular modeling technique, indicating that there is an advantage of MM for studying the mechanical properties of cellulose.

The modulus of cellulose II was found to be smaller than that of cellulose I by MM using the COMPASS force field (Eichhorn and Davies 2006; Eichhorn and Young 2001; Eichhorn *et al.* 2005). A suggested explanation at molecular level is that each glucose residue in cellulose I contains 2 intra-chain hydrogen bonds but only contains one in cellulose II, thus the stiffness of cellulose I in the chain direction is stronger than that of cellulose II. However, most of the modeling studies on modulus of cellulose are based on MM approach except the ones from Bergenstråhle *et al.* (2007), Neyertz *et al*. (2000) and Wohlert *et al.* (2012) that employed MD simulation.

Chapter II - Methods

II.1. GROMACS (*GROningen MAchine for Chemical Simulations*)

During molecular dynamics simulation, most of the computation resource was spent on the calculation of non-bonded interaction. As an engine of MD simulation, GROMACS (Hess *et al.* 2008) is primarily designed for biochemical molecules and is extremely efficient in calculating non-bonded interaction. Due to several advantages – it is free, fast and flexible –, the GROMACS package was used for all the MD simulation in this thesis.

II.2. Model construction

The initial coordinates for model construction were obtained from the X-ray and neutron diffraction data and contain explicit positions of hydrogen atoms (Langan *et al.* 1999, 2001; Nishiyama *et al.* 2002, 2003; Wada *et al.* 2004)). Due to the fact that two types of hydrogen-bonding system occurred in cellulose Iβ and Iα, in most studies of this work only the hydrogen-bonding pattern A was considered except in special cases where pattern B will be also be considered (in **Chapter V**).

II.2.1. Infinite models

Three models were constructed for Iα, containing 64 (8×8), 36 (6×6) and 40 chains (5×8), respectively. Both 64-chain and 36-chain models have their periodic boundaries parallel to the ($1\bar{1}4$), (010) and (100) planes, whereas those in the 40-chain model are parallel to the ($11\bar{4}$), ($1\bar{1}0$) and (110) planes. Two models were constructed for Iβ, containing 36 (6×6) and 40 chains (5×8), respectively. The 36-chain model has its periodic boundaries parallel to the (110), ($\bar{1}10$) and (001) planes, while those in the 40-chain model are parallel to the (200), (010) and (001) planes. The model for cellulose II contains 40 chains (5×8), with the periodic boundaries parallel to the (020), ($1\bar{1}0$) and (001) planes, whereas the model of cellulose III_I contains 45

chains (5×9) with the boundaries parallel to the (100), (010) and (001) planes. All the models approximately have a $4 \times 4 \times 4$ nm^3 cubic shape, which is recommended by the GROMACS program. Selected models are given in **Figure II.1**.

Along the chain direction, the unit cell was repeated 4 times, the box size thus corresponding to 8 anhydroglucose units, and was covalently bonded to its own periodic image. Periodic boundary conditions were applied along the *x*, *y* and *z* directions, yielding models without boundaries in order to simulate infinite systems.

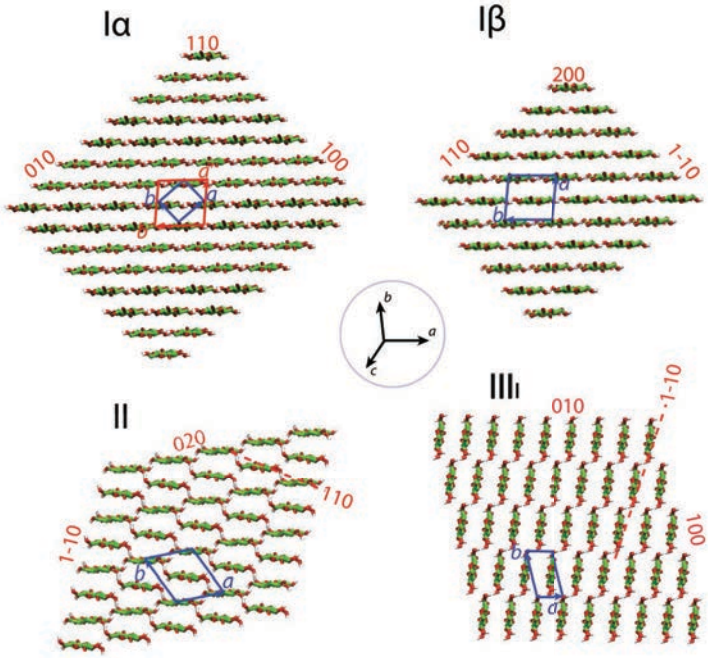

Figure II.1. Infinite models of the cellulose allomorphs.

II.2.2. Semi-finite models (infinite along the chain axis)

The semi-finite systems were constructed based on the infinite models (**§ II.2.1**). The models were placed in a new simulation box with the dimension of their cross-section ~2 nm larger than their original ones. Water molecules were substantially added into the box to fill the additional space, yielding the models immersed in water by the side plane. Periodic boundary conditions were applied along the *x*, *y* and *z* directions. Since the cross-section was still covalently linked to its periodic image and the sides of crystal were in contact with water molecules, the new models were only infinite along the chain direction, thus corresponding to the semi-infinite models.

II.2.3. Finite models

Finite models were only constructed for Iα and Iβ. Both models contain 36 chains (6×6) and have a cubic shape. Each chain has a degree of polymerization (DP) of 40 and is about 21 nm in length. The terminal groups HO1 and HO4, O4 were added in each chain to make it “finite”. The models were placed at the center of a box with dimensions of 7.4 × 7.4 × 30 nm^3 and were surrounded with water molecules.

II.2.4. Water model

The simple point charge (SPC) water model (Berendsen *et al.* 1981) was used in our simulations. In this model, water molecules are treated as rigid and thus the bond lengths and angle are fixed. A positive point charge was assigned to the hydrogen atom and a negative charge to the oxygen atom. Only the oxygen atom has Lennard-Jones parameters.

II.3. Molecular dynamics setups

II.3.1. Basic principle of MD simulation

During MD simulation, the system solves Newton’s equation. When a force field is selected to calculate the “force” and computational procedures defines how and what to calculate, it is necessary to build a starting structure, which originated from crystallography data in our case. The initial velocity is randomly generated by the system according to a Maxwell distribution. Energies are calculated at each time step (normally equal to femtoseconds) according to the force field. Force and acceleration of each atom can simply be derived from the calculated energy. The velocity at time t and its next step $t+\Delta t$ can be obtained from the integration algorithm (such as leaf-frog algorithm). Thus, the position of each atom at the next time step can be obtained. By repeating this process, the molecules undergo continual changes in conformation and shape. The coordinates and velocity of the system can be saved at every pre-defined time step, yielding a so-called "trajectory file".

II.3.2. Atomistic force field methods: potential energy profiles

The GROMOS force fields (Lins and Hünenberger 2005; Hansen and Hünenberger 2011), primarily developed for carbohydrates by the computer-aided chemistry group in ETH (Zurich, Switzerland), were used. The term "GROMOS" also refers to the software package that was developed by the same group. The potential energy function in GROMOS is the sum of different energy terms (Eq. 1). They describe the energy of the bond strength (Eq. 2), bond bending (Eq. 3), improper (Eq. 4) and proper (Eq. 5) dihedral (so-called "bonded interactions"), as well as electrostatic interaction (Eq. 6) and van der Waals (VdW) (Eq. 7) interactions (so-called "non-bonded interactions"):

$$E = E_{Strength} + E_{Bending} + E_{improper-dih} + E_{proper-dih} + E_{elec} + E_{LJ} \quad (1)$$

$$E_{Strength} = \sum (1/4) k_{b,m} [b_m^2 - b_{o,m}^2]^2 \quad (2)$$

$$E_{Bending} = \sum (1/2) k_{\theta,m} [\cos\theta_m - \cos\theta_{o,m}]^2 \quad (3)$$

$$E_{improper-dih} = \sum (1/2) k_{\xi,m} [\xi_m - \xi_{m,o}]^2 \quad (4)$$

$$E_{proper-dih} = \sum k_{\phi,m} [1 + \cos\delta_m \cos(n_m \phi_m)] \quad (5)$$

$$E_{elec} = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left[\frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right] \quad (6)$$

$$E_{LJ} = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left[\left(\frac{C12_{ij}}{r_{ij}^{12}} - \frac{C6_{ij}}{r_{ij}^{6}} \right) \right] \quad (7)$$

The energy of strength, bending and improper dihedral are described by the harmonic potential. The $b_{o,m}$, $\theta_{o,m}$ and $\xi_{m,o}$ in the formulas indicate equilibrium values of the bonds length, bond angle and "out of plane" angle, respectively, which can be derived from structural data and the corresponding $k_{b,m}$, $k_{\theta,m}$ and $k_{\xi,m}$ force constants can be obtained from spectroscopy data.

The electrostatic energy obeys the law of Coulomb interaction. The electron cloud is simply represented by the fixed-point charges q_i, which are not affected by the local electrostatic environment. The atomic partial charges are initially derived from theoretical methods such as quantum-chemical calculation and finally from fitting to thermodynamic data of target

molecules.

The VdW interactions are taken into account via Lennard-Jones (LJ) potential term. The $C12_{ij}$ and $C6_{ij}$ parameters are derived by fitting thermodynamic properties such as heat of vaporization, density, free energy of solvation and so on. The r_{ij} in both Eq. 6 and 7 is the distance between two atoms. The torsional angle parameters (in Eq. 5) were derived by fitting torsional energy profile to quantum calculation data of small molecules. It has been argued that the parameterization of torsional parameters is the one that lacks the most accurate standard (Woods *et al* 1995).

II.3.3. Temperature coupling

Temperature is a physical quantity that is deduced from the average kinetic energy of microscopic particles in a system and thus can be counted from velocity and mass of each atom. In this thesis, the velocity-scaling algorithm was used to control the temperature by rescaling the velocity of all the atoms contained in the system.

II.3.4. Pressure coupling

Berendsen's algorithm (Berendsen *et al.* 1984) was used to control the pressure by rescaling the coordinate and box size of the system.

II.3.5. Constraints

All the bonds were constrained by using the LINCS algorithm (Hess *et al.* 1997).

II.3.6. Treatment of calculation of the non-bonded interaction: cut-off and correction

Cut-off criteria are always used in order to speed up the calculation of the non-bonded terms. With the cut-off value, the non-bonded interaction of the atoms involved in the range of half cut-off radius, or the short-range non-bonded interaction, was explicitly calculated, whereas the interaction out of the cut-off radius, or the long-range non-bonded interaction was treated in the different way. In this work, the long-range electrostatic interaction is calculated by using the

particles mesh Ewald (PME) methods (Darden *et al.* 1993). The long-range Lennard-Jones interaction was corrected by energy and pressure.

Almost in every force field, the 1-4 atoms pairs and the 1-N (N>4) are treated with different non-bonded force field parameters, generally because that the parameters of the 1-N (N>4) atoms pairs yield too strong interaction for 1-4 pairs due to their short distance. In GROMOS force fields, the LJ parameter for 1-4 pairs is weaker than that of 1-N (N> 4) pairs to allow the atoms being in the distance without strong repulsion, whereas the atomic partial charges are the same between 1-4 and 1-N (N>4) pairs. Some force field, such as AMBER, has provided a so-called 1-4 electrostatic factor that could adjust the amount of the 1-4 electrostatic interaction.

II.3.7. Ensemble and integration time step

The NPT ensemble (constant number N of particles, temperature T and pressure P), which describes the state of real cellulose, was set in the simulation. The integration time step (or step size) is an important parameter for accurate MD simulation, since it refers to the precision and efficiency of the simulation. In principle, the value of the integration time step should be no less than one tenth of the fastest motion period in the system. In the case of cellulose, the fastest molecular motion comes from the bond strength of O-H, the frequency of which can be measured by IR spectroscopy and is about 10^{14} Hz. The corresponding strength period is 10^{-14} s. Thus, the maximum integration time step can be 10^{-15} s. When the bond constraints are applied, the integration time step can be increased for efficiency consideration.

II.3.8. Periodic boundary condition

Since the number of the particles in our simulation is significantly smaller than that in any piece of a macroscopic system, the proportion of surface atoms in the simulation system would be much larger than that in reality, and the surface of the system is surrounded by vacuum, causing a larger surface effect than what it really should be. Thus, this simulation is less reliable to reflect the reality. One solution to this problem is to use periodic boundary conditions so that the surface of the system is surrounded by translated copies of itself. Thus, there are no boundaries and the edge effects are avoided.

II.3.9. Stepwise heating and cooling procedure (with different rates)

This thesis contains MD simulations carried out both at room and high temperature. To reach a target temperature, a stepwise strategy of heating or cooling was performed, in which the temperature was increased or decreased by 10 K at each step and followed by an equilibration step (varying from 100 ps to 12 ns, depending on the purpose of simulation) at this temperature.

II.3.10. Energy minimization and position restrained molecular dynamics

Energy minimization was performed before any MD simulation, firstly by using the steepest descent method and later by conjugated gradient methods. In the simulation where water molecules were included, position restrained MD simulation was performed for 200 ps after energy minimization to remove the "bad" contacts between solvent and solute caused by the random solvent-adding process.

II.4. Trajectory visualization and analysis

II.4.1. Default GROMACS tools, VMD, PYMOL

The GROMACS program provides many tools for analyzing the conformation and properties of systems in the trajectory file. The visualization of the trajectory files used the VMD program (Humphrey *et al.* 1996). The graphs of molecules presented in the thesis were drawn using VMD and PYMOL software.

II.4.2. Unit cell calculation

The unit cell parameters of the infinite models were calculated from the total box size of the system, while those of the semi-finite and finite models were calculated from the core chains.

Chapter III - MD studies on the crystal structures and thermal behavior of cellulose using GROMOS 53a6 and 56A$_{carbo}$ force fields

III.1. Introduction

To characterize the structure and understand the thermal behavior of cellulose, I have carried out atomic level MD simulations on the four allomorphs using two united force fields, GROMOS 53a6 (Oostenbrink *et al.* 2004) and GROMOS 56A$_{carbo}$(Hansen and Hünenberger 2011). The GROMOS 53a6 force fields for hexopyranose-based carbohydrates is almost identical to GROMOS 45a4 (Lins and Hünenberger 2005), which treats a glucose ring in cellulose chain into four atom types that named CH1, CH2, OA and H. Whereas the latest version of GROMOS force field for carbohydrates named 56A$_{carbo}$, which is optimized to overcome some disadvantages of 53a6, treats a glucose ring into six atom types named CH1r, CH2, OA, Or, OE and H, as shown in **Figure III.1**. The reoptimization yielded slightly different atomic partial charges, new inter-ring Lennard-Jones parameters (1-4 pair and 1-5 pair) and systematically varied torsional parameters.

I have evaluated the performance of these force fields to model the crystal structures. This was achieved by comparing the structures modeled at equilibrium with the experimental ones. In this chapter, I will first describe our specific computational procedures. Then, a detailed comparison between the experimental structures and those modeled at room temperature will be given. Finally, I will present simulation results on the thermal behavior of cellulose.

III.2. Computational procedures

III.2.1. Model construction

This chapter only refers to the infinite models of cellulose Iα, Iβ, II and III$_I$ as described in **Chapter II**. Only the most populated location of the hydroxyl hydrogen atoms (pattern A) has been considered for Iα and Iβ. The models are shown in **Figure II.1**.

III.2.2. MD tools and force field methods

The GROMACS software package (Hess *et al.* 2008) version 4.5.3 was used for all of the minimizations, molecular dynamics and analysis. Two types of united atom force fields, namely GROMOS 53a6 (Gr53a6FF) (Oostenbrink *et al.* 2004) and GROMOS $56A_{carbo}$ ($Gr56A_{carbo}FF$) (Hansen and Hünenberger 2011), have been used. In these united atom force fields, the carbon atoms and their hydrogen atoms in CH and CH2 groups are treated as single particles.

III.2.3. System setup

The cutoff distance for the short-range Coulomb interactions and the Lennard-Jones nonbonded interactions was set to 0.9 nm. The long-range dispersion force was corrected for energy and pressure. PME summation (Darden *et al.* 1997) was used for long-range Coulomb interactions. The equations of motion were solved using the standard leap-frog algorithm (Hockney *et al.* 1974) using time steps of 1 fs (0.5 fs for cellulose III_I with Gr53a6FF). The pressure and temperature of the systems were regulated using the Berendsen (Berendsen *et al.* 1984) and the velocity-rescaling algorithms (Bussi and Parrinello 2007), respectively. All bond lengths were constrained using the LINCS algorithm (Hess *et al.* 1997). MD frames were saved every ps. The *xx*, *yy*, *zz*, *xy*, *yz*, *zx* components of the compressibility were set to 2.5×10^{-5}, 2.2×10^{-5}, 7.9×10^{-7}, 1.2×10^{-5}, 4.2×10^{-5}, 4.4×10^{-5} bar^{-1} for the model of cellulose I_β; 7.3×10^{-7}, 6.1×10^{-5}, 3.1×10^{-5}, 5.4×10^{-5}, 4.9×10^{-5}, 1.2×10^{-5} for I_α; 1×10^{-6}, 1×10^{-6}, 1×10^{-7}, 1×10^{-6}, 1×10^{-6}, 1×10^{-6} for II and III_I.

III.2.4. Equilibration at room temperature and simulated annealing

The structures were first optimized by the steepest-descent, followed by the conjugate gradient methods. The convergence criterion to stop minimization was a maximum force of 1 kJ $mol^{-1}Å^{-1}$. The structures were then heated stepwise from 0 to 300 K in 6 ns, with typical 10 K increments in 100 ps. Then, the system was equilibrated at 300 K for 10 ns in the NPT ensemble, at atmospheric pressure. The periodic boxes were allowed to deform in size and shape, involving six degrees of freedom as described in **§ III.1**. The structural parameters at equilibrium were estimated from the last 5 ns of the trajectories.

The structures from the last frames of the 10 ns trajectories at 300 K were subsequently heated to 600 K for cellulose I and 500 K for cellulose II and III_I. When the temperature reached 600 K, the systems were equilibrated for 20 ns and then cooled down to 300 K in 12 ns. Heating and cooling rates ranged from 1 to 50 K/ns.

Figure III.1. The definition of the atom types and names in a glucosyl residue in Gr53a6FF and Gr56A$_{carbo}$FF.

III.2.5. Structural characteristics and properties

The structure of the cellulose crystals can be analyzed both in terms of conformation of the individual chains of cellulose and their ultrastructural organization. The puckering parameters of the pyran rings, the angles of the glycosidic linkage and the orientation of the exocyclic groups describe the conformation of the chains, while the unit cell parameters and the hydrogen-bonding systems are used to describe the ultrastructure. Atom names of a glucose ring are indicated in **Figure III.1** (right).

The Cremer-Pople puckering parameters (Q, Θ, Φ) describe the shape of the D-glucopyranose ring (Cremer and Pople 1975). They are calculated from the coordinates of ring atoms O5, C1, C2, C3, C4 and C5. The angles (φ, ψ, τ) of the glycosidic linkage were defined by the sequence of atoms O5-C1-O1-C4, C1-O1-C4-C5 and C1-O1-C4, respectively.

Three hydroxyl groups and one hydroxymethyl group are attached on the glucose ring. Their orientations were described by the torsion angles defined by C1-C2-O2-HO2, C2-C3-O3-HO3, C5-C6-O6-HO6 and O5-C5-C5-O6. These angles are generally named by τ2, τ3, τ6, and ω. Unit cell parameters were calculated from the total box size at each temperature. For convenient comparison of Iβ and Iα, the original unit cell of Iα was converted using

$a'=a+b=0.9699$, $b'=b-a=0.8201$ nm, $c'=c=1.04$ nm, $\alpha=89.9°$, $\beta=125.4°$, $\gamma=96.89°$ (Bučko *et al.* 2011). This converted structure is shown in **Figures II.1.**

The thermal expansion coefficient (TEC) α was calculated from the expression:

$$\alpha = \Delta L/(L * \Delta T)$$

where L is a unit cell parameter or a d-spacing and T is the temperature.

The hydrogen bonds at 300 K were measured from the calculated average structure. The geometry criteria to detect a hydrogen bond was that the distance between the donor (O) and acceptor (O) was lower than 0.4 nm and the angle between hydrogen-donor-acceptor lower than 60° or lower than 30°.

III.3. Comparison between simulated and experimental crystal structures

Room temperature MD simulations of cellulose Iβ, Iα, II and III_I have been performed using Gr53a6FF and Gr56A_{carbo}FF. Among these 8 simulations, those of cellulose II and III_I with Gr53a6FF failed. The hydroxymethyl groups departed from their initial orientations and the unit cells changed, from a monoclinic to a triclinic type.

III.3.1. Puckering parameters

Table III.1 summarizes the simulated and experimental puckering parameters of the glucosyl rings in cellulose crystals. For cellulose Iβ and II, parameters were separately calculated for the origin and the center chains, while for Iα and III_I, parameters were separately estimated for the two neighbor residues in the chains, named residues 1 and 2, respectively, in **Table III.1**.

All the pyranose rings in the models adopt the classical $^4C_1$ chair conformation, which agrees with the experimental observation. Calculated parameters Θ agree well with the experimental values but that in the center chain of cellulose II. Comparing the two force fields, the puckering amplitude Q calculated by using Gr56$_{carbo}$FF is larger than that by using Gr53a6FF and also larger than the experimental values. This may be due to one or two terms of optimized force field parameters, such as new partial charges, Lennard-Jones interaction of intra-ring 1-5 pairs and systematically varied torsional parameters, or due to all of them.

Table III.1: Puckering parameters in cellulose Iβ, Iα, II and III_I from experiment and simulation.

Allomorph		origin chain or residue 1		center chain or residue 2	
		Q (nm)	Θ (°)	Q (nm)	Θ (°)
Iβ	*53a6*	0.056(4)	9.3(4.8)	0.056(4)	8.3(4.5)
	56A$_{carbo}$	0.061(4)	8.8(4.4)	0.063(4)	10.5(5.2)
	Exp.	0.0542(0)	10.4(3)	0.0558(0)	6.7(2)
Iα	*53a6*	0.055(4)	8.5(4.6)	0.055(4)	8.5(4.6)
	56A$_{carbo}$	0.060(4)	8.3(4.5)	0.060(4)	8.3(4.5)
	Exp.	0.0585	7.3	0.0572	9.9
II	*53a6*	0.058(4)	9.1(4.9)	0.057(5)	10.0(6.6)
	56A$_{carbo}$	0.063(4)	9.1(4.5)	0.060(4)	9.2(4.7)
	Exp.	0.0605	10.8	0.0507	3.9
III_I	*53a6*	0.056(5)	9.6(5.6)	0.056(5)	10.1(6.6)
	56A$_{carbo}$	0.068(4)	13.8(5.4)	0.068(4)	13.8(5.4)
	Exp.	0.0550	10.6	0.0551	10.6

X-ray crystallography revealed that the two chains differ in the unit cell of the Iβ allomorph (Nishiyama *et al.* 2002). Similarly, Gr53a6FF and Gr56A$_{carbo}$FF give different equilibrium puckering amplitudes for the origin and center chains. However, Gr53a6FF gives values closer to the experimental ones than Gr56A$_{carbo}$FF, which overestimated the amplitude by 11.1%.

The two independent residues in the unit cell of the Iα allomorph also differ in the experimentally derived structure. The two rings are similar in the models. Gr53a6FF underestimated the puckering amplitude whereas Gr56AcarboFF overestimated it.

Cellulose II has two independent chains in its monoclinic unit cell, yielding two sets of puckering parameters. From Gr56A$_{carbo}$FF, the simulated amplitudes Q are larger than the experiment values. Q of the rings in the center chain is 16.7% larger.

All the pyranose rings in cellulose III_I have the same conformation. The simulated amplitude Q from Gr56A$_{carbo}$FF is 17.7% smaller than the experimental values. In comparison, the values from Gr53a6FF are very close to experimental data.

III.3.2. Glycosidic linkage

The values explored by the glycosidic angles (φ, ψ, τ) are shown as distribution curves in **Figure III.2**. In simulation, I did not observe an obvious difference between the values adopted by the glycosidic angles neither in the center and origin chains in cellulose Iβ and II nor in the two neighbor residues in cellulose Iα and III_I. Consequently, unlike for the calculation of puckering parameters, the glycosidic linkage angles were averaged for all the residues in the structures without distinction of the center and origin chains or of the two consecutive residues.

As shown in **Figure III.2**, all simulated angles were narrowly distributed around the experimental values, indicating that the linear conformation of the cellulose chains in the four allomorphs was well reproduced by using Gr53a6FF and $Gr56A_{carbo}FF$. Gr53a6FF and $Gr56A_{carbo}FF$ gave very similar distributions for cellulose Iβ, Iα and II. In contrast, they differ in cellulose III_I (**Figure III.2d**). It has to be noted that the values from $Gr56A_{carbo}FF$ significantly departed from experiment.

III.3.3. Conformation of hydroxymethyl groups (exocyclic groups)

The values explored by the torsion angle ω are shown as distribution curves in **Figure III.3**. Experimentally, the orientation of the hydroxymethyl groups is *tg* (around 180° according to our definition) in Iβ and Iα and *gt* (around 60°) in II and III_I (shown in **Figure III.3** by the vertical lines). The conformation of the hydroxymethyl group in the models is in agreement with the experiments for Iβ (*tg*), Iα (*tg*) and III_I (*gt*). The model of cellulose II from $Gr56A_{carbo}$ FF possesses about 90% and 10% of *gt* and *gg* orientations, respectively, although from experiments, only the *gt* conformation is explored. Note that the simulation conditions involved heating of the structure from 0 to 300 K in 6 ns. Another simulation for cellulose II that directly started at 300 K showed only the *gt* conformation, stable over 10 ns.

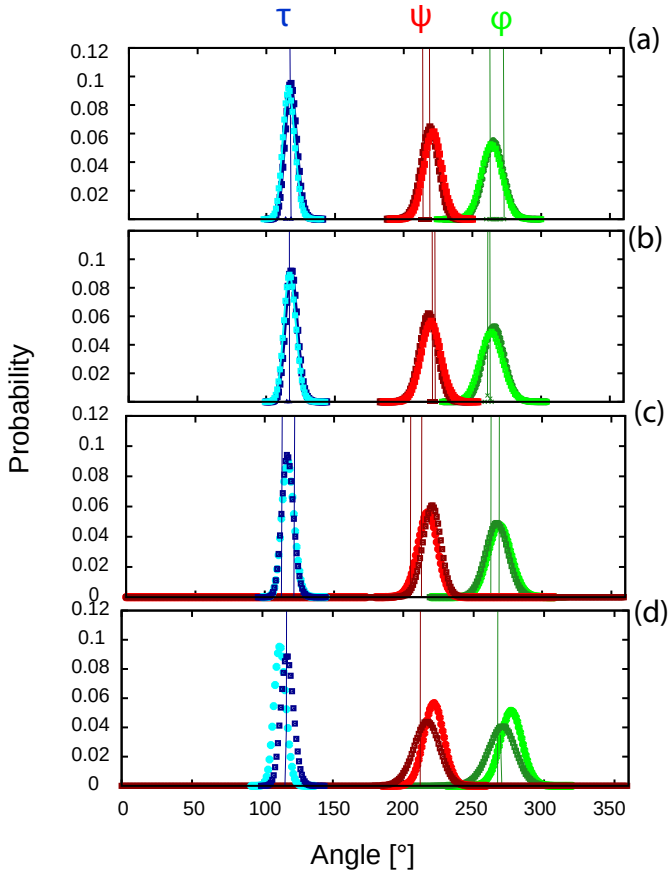

Figure III.2. Estimated glycosidic linkage angles (φ: green, ψ: red, τ: blue and cyan) of cellulose Iβ (a), Iα (b), II (c) and III_I (d) by using Gr53a6FF (dark blue, dark red, forest green) and Gr56A$_{carbo}$FF (cyan, red, green). The vertical lines indicate the experimental values. The two lines with the same color represent conformational difference either from the center chain and the origin chain or from the two neighbor residues in one chain.

III.3.4. Conformation of hydroxyl groups (exocyclic groups)

The values explored by the torsion angles τ2, τ3 and τ6 are shown as distribution curves in **Figure III.3**. The distributions of these angles in the equilibrated models of Iβ and Iα are very similar. Both Gr53a6FF and Gr56A$_{carbo}$ FF homogenize the orientation of the pendant groups in these crystal structures. Note that the distributions from Gr56A$_{carbo}$FF were wider than those from Gr53a6FF. For cellulose Iβ and Iα, the torsion angles were distributed close to their corresponding experimental values, showing the consistency between simulation and experiment. However, the orientation of the hydroxyl groups differs in the experimental crystal structures of Iβ and Iα allomorph, as seen by comparing the vertical lines between **Figures III.3a** and **III.3b**.

For cellulose II, the distributions of τ2, τ3 and τ6 were wide. Only τ2 (results from

Gr56A$_{carbo}$FF) and τ3 were located close to their experimental values. The τ6 showed three peaks between 0 and 200° and the τ2 (results from Gr53a6FF) showed two peaks none of them were close to the experimental value. This conformational multiplicity of τ2 and τ6, which disagrees with the experiments, originates from the two conformations adopted by the hydroxymethyl group (torsion angle ω).

For cellulose III$_I$, the distributions of the τ2 and τ6 torsion angles from Gr56A$_{carbo}$FF were close to the experimental values (**Figure III.3d**), whereas those from Gr53a6FF were totally different.

However, the center of the distribution of τ3 modeled with Gr56A$_{carbo}$FF differed from the experimental value by 101°, but that from Gr53a6FF was close to the experimental value. The poorly simulated distribution of τ3 suggests very different intra-chain hydrogen bonds between HO3 and O5. In summary, the simulated conformations of τ2, τ3 and τ6 in II and III$_I$ depart more from experimental observation than cellulose I.

III.3.5. Unit cell parameters

From the analysis of the data summarized in **Table II**, I found that the unit cell parameters related to the distance between the cellulose layers (*a* of Iβ, Iα and III$_I$, *b* of II) were overestimated for the four allomorphs independently of the force field. The parameters related to the distance between the cellulose chains within the sheets where the hydrogen bonds occur (for example, *b*-parameter for Iβ and Iα and *a*-parameter for II) were well reproduced except in cellulose III$_I$. The *c*-parameters were overestimated for Iβ, Iα and II, but underestimated for III$_I$ by using Gr56A$_{carbo}$FF. The monoclinic angles were underestimated for Iβ, Iα and II, but overestimated for III$_I$.

III.3.6. Intra-chain hydrogen bonds in cellulose I

The distribution of the distances and angles describing the intra-chain hydrogen bonds (HB) in crystalline cellulose are presented in **Figures III.4** and **III.5**.

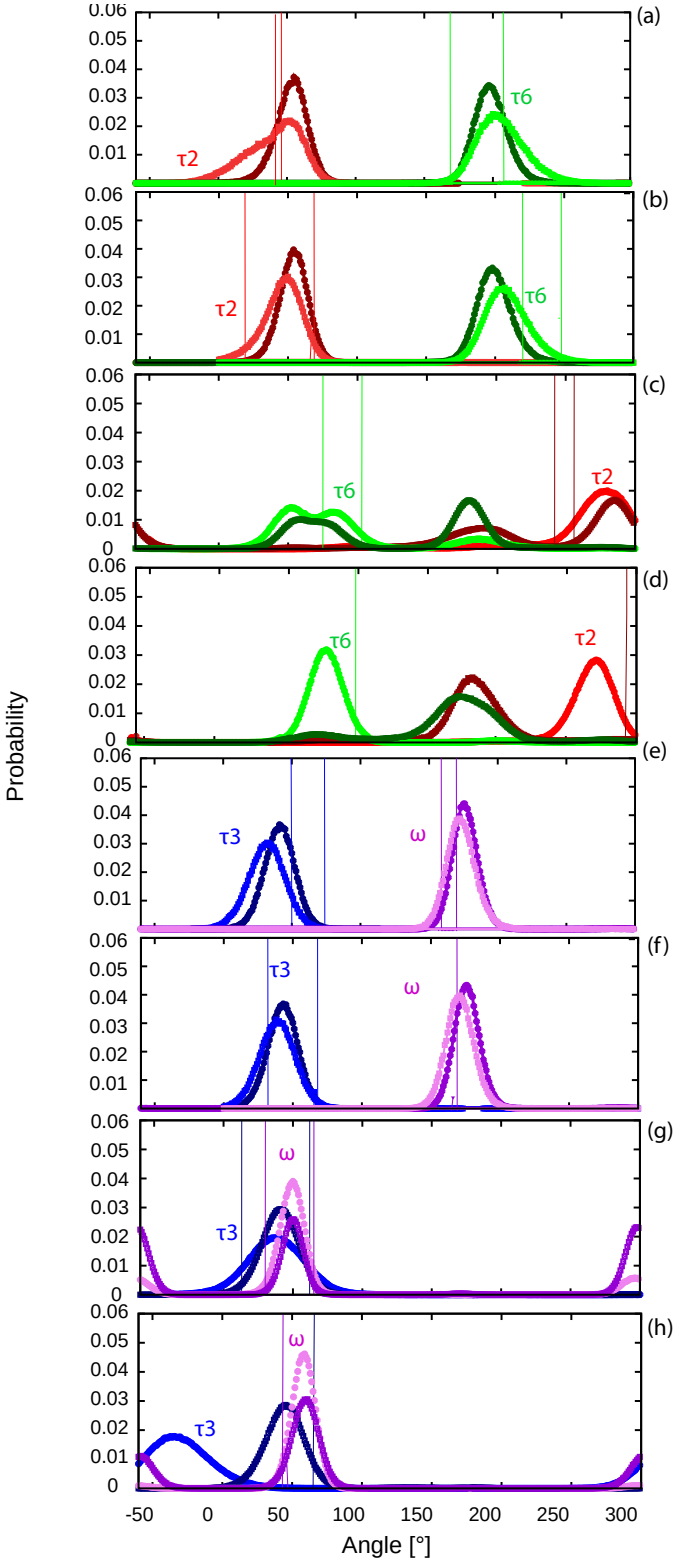

Figure III.3. Calculated dihedral angles of three hydroxyl groups (τ2, τ3 and τ6) and one hydroxymethyl group (ω) of cellulose Iβ (a,e), Iα (b,f), II (c,g) and III_I (d,h) by using Gr53a6FF (deep colors: dark red, dark blue, dark green and dark violet) and Gr56A_{carbo}FF (bright colors: red, blue, green, violet). The straight lines indicate the experimental values. The two lines with same color represent conformational differences either from the center chain and the origin chain, or from the two neighbor residues in one chain.

By using Gr53a6FF and Gr56A$_{carbo}$FF, this 2D HB network was well reproduced for both Iβ and Iα. However, in the simulations of Iβ, the O2---O6 distances were shorter and the H-O2-O6 angle smaller when compared to their experimental values, indicating that this HB in the models was stronger. In contrast, it was weaker than the experiments in the Iα models. The simulated O3-HO3---O5 intra-chain HB was found to be weaker than the experimental one for the Iα models but stronger in the case of the Iβ ones.

Table III.2. Unit cell parameters of the cellulose allomorphs from experiments (*Exp*) and MD simulations using two force fields. The parameters of III$_I$ from Gr53a6FF were measured from the data of the last frame.

Allomorph		***a*** **(nm)**	***b*** **(nm)**	***c*** **(nm)**	***α*** **(°)**	***β*** **(°)**	***γ*** **(°)**
Iβ	*53a6*	0.8384(6)	0.8159(2)	1.0515(1)	90.00(4)	90.00(4)	90.82(7)
	56A$_{carbo}$	0.8425(3)	0.8136(3)	1.043(0)	90.00(2)	90.01(3)	92.86(3)
	Exp.	0.7784(8)	0.8201(8)	1.038(1)	90	90	96.55
Iα	*53a6*	0.9918(4)	0.8164(1)	1.0542(1)	89.88(2)	122.30(3)	90.57(5)
	56A$_{carbo}$	0.9983(4)	0.8179(1)	1.0473(1)	89.91(2)	121.83(3)	93.77(3)
	Exp.	0.9699(7)	08201(6)	1.0400(1)	89.92(5)	125.44(5)	96.89(5)
II	53a6	0.7936(3)	0.9942(12)	1.046(0)	89.63(5)	89.61(4)	115.87(4)
	56A$_{carbo}$	0.8026(5)	0.9501(18)	1.040(0)	89.94(8)	89.98(7)	113.20(8)
	Exp.	0.801(3)	0.903(3)	1.031(5)	90	90	117.10(5)
III$_I$	53a6	0.5122	0.7535	1.0473	90.78	86.23	109.96
	56A$_{carbo}$	0.5009(1)	0.8182(2)	1.007(0)	89.96(4)	89.96(6)	111.80(4)
	Exp.	0.4450(4)	0.7850(8)	1.031(1)	90	90	105.10(5)

III.3.7. Inter-chain hydrogen bonds in cellulose I

Figures III.4 and **III.5** give the distribution of the distances and angles describing the inter-chain HBs. Gr53a6FF and Gr56A$_{carbo}$FF gave very similar HBs in Iα and Iβ. Compared to the experimental values, the simulated O2---HO6 HBs in Iβ have larger distances and angles while in Iα it have smaller angles, showing that this HBs were weaker in Iβ but stronger in Iα. In

contrast, the simulated O6---HO3 HBs in Iβ have shorter distances and smaller angles, indicating that it was simulated stronger than the experimental one. However, the opposite holds for Iα.

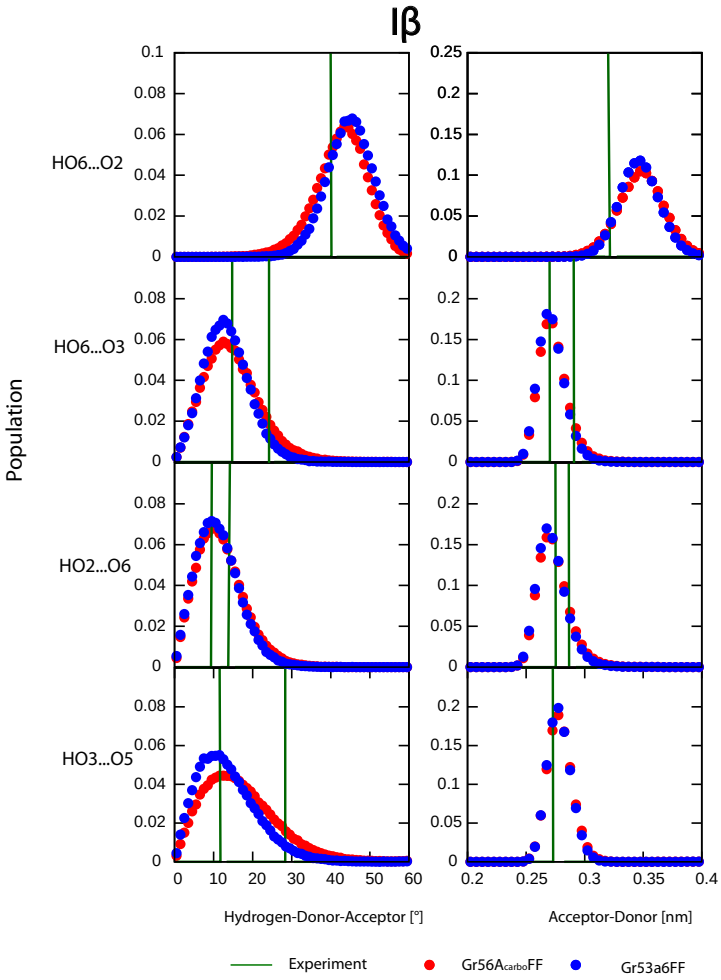

Figure III.4. Comparison of simulated and experimental HBs (O5-HO3, O2-HO6, O6-HO2 and O3-HO6) in native cellulose Iβ (Gr56A$_{carbo}$FF in red and Gr53a6FF in blue).

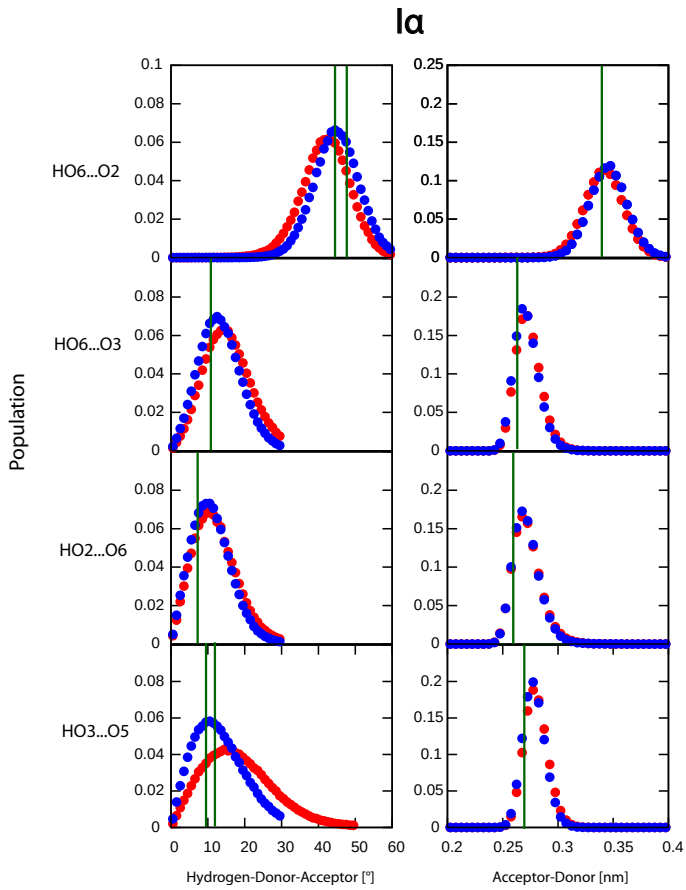

Figure III.5. Comparison of simulated and experimental HBs (O5-HO3, O2-HO6, O6-HO2 and O3-HO6) in native cellulose Iα (Gr56A$_{carbo}$FF in red and Gr53a6FF in blue).

III.3.8. Hydrogen-bonding system in cellulose II and III$_I$

Simulations also revealed the 3D HB network for II and III$_I$. However, the estimated HB distances and HB angles significantly changed from the experimental values during MD simulation. The results obtained with Gr53a6FF are not presented here because of the undesirable *gg* conformation for the hydroxymethyl group, which resulted in HBs completely different from experimental data.

Compared with the simulated HB network in cellulose I, the HB network in II and III$_I$ were poorly simulated, which could be expected from the simulation of the dihedral conformations of HO2, HO3 and HO6 in **Figures III.3c,d,f,g**.

For cellulose II, the donor to acceptor distances describing the O6---HO6, O6---HO2, O2---HO2 and O2---HO6 inter-chain HBs all decreased when compared to the experimental distances, showing a stronger inter-chain HB network (**Table III.4**).

The atoms involved in HBs have the tendency to be closer in the models than in the experiments, except O6---HO3 in cellulose III_I and O5---HO6 in cellulose II. All the results show that the simulated 3D HB networks are always different from those experimentally determined, which is also the case for cellulose III_I (**Table III.3**).

Table III.3. HBs in cellulose III_I simulated with Gr56A$_{carbo}$FF. The results with Gr53a6FF are not presented here. The experimental data are also shown. For MD simulations, the HBs were measured from the structure averaged from the last 5 ns of the trajectory files, which is different from the way it was measured for I_α and I_β. The cutoff distance was between 0.12 and 0.3 nm and the cutoff angle between 110 and 180°. The same strategy was used for cellulose II in **Table III.4**.

Donor-acceptor	MD Simulation		Experiment	
	Distance (nm)	Angle (°)	Distance (nm)	Angle (°)
O6---HO2-O2	0.166-0.186	163-171	0.199	123
O2---HO6-O6	0.16- 0.22	165-180	0.167	156
O3---HO6-O6	0.23- 0.28	90-110	0.299	122
O1---HO3-O3	0.245- 0.275	90-110	0.261	107
O5---HO3-O3	0.18-0.205	140-165	0.198	149
O6---HO3-O3	0.274- 0.296	115-145	0.243-0.244	117-118

Table III.4. HBs in cellulose II simulated with Gr56A$_{carbo}$FF after MD simulation. The experimental data is also given.

Donor-acceptor	MD Simulation		Experiment	
	Distance (nm)	Angle (°)	Distance (nm)	Angle (°)
O6---HO6-O6	0.169-0.171	163-180	0.200-0.204	117-118
O6---HO3-O3*	0.241-0.30	110-130	0.277, 0.239, 0.258	115, 142, 101
O6---HO2-O2	0.16-0.3	60-180	0.182	149
O5---HO3-O3*	0.17-0.20	140-170	0.184-0.188	135, 145
O5---HO6-O6	0.27-0.3	65-120	0.246	124
O3---HO6-O6	0.16-0.3	90-180	0.221	144
O2---HO6-O6	0.16-0.3	60-180	0.172	168
O2---HO2-O2	0.16-0.3	60-180	0.208-0.212	124

*Two types of HB can be distinguished, one located in the origin chain and the other located in center chain. No distinction was made in this table.

III.3.9. Structural stability of cellulose II and III_I by using Gr53aFF and Gr56A_{carbo}FF

As mentioned at the beginning of **§ III.3**, cellulose III_I showed structural deformation during MD simulation using Gr53a6FF. The starting monoclinic unit cell, in which the chains are non-staggered along the chain direction, slowly changed to a triclinic type or to another monoclinic unit cell. The chains in these unit cells were staggered. Furthermore, the initial *gt* conformation of the hydroxymethyl groups changed to a mixture of *gg* and *gt*.

While the chains are parallel in cellulose Iβ, Iα and III_I, they are antiparallel in cellulose II with a ±0.5 nm staggering along the chain direction. The monoclinic unit cell was maintained during MD but the hydroxymethyl groups rotated. During the 30 ns MD simulation, almost 40% of them explored the *gg* conformation and 60% the *gt* one (**Figure III.3g**).

Since the conformation of the hydroxymethyl groups largely determines the HB network, the structural deformation of II and III_I may due to the undesired *gg* conformation. Other treatments of the non-bonded force field terms may avoid the occurrence of *gg* conformation so that the monoclinic unit cell of III_I could be maintained during MD simulation. For example, neglecting the electrostatic interaction of the 1-4 pair between C4 and O6 will favor the *gt* conformation, similarly, an electrostatic factor of 0.5 decreases the Coulomb interaction strength for all the 1-4 pairs. Thus, the deformation of II and III_I must result from the non-bonded force field parameters.

On another hand, the crystal structure of cellulose II and III_I remained stable using Gr56A_{carbo}FF, which was probably a consequence of the well reproduced *gt* conformation of the hydroxymethyl groups (100% *gt* in III_I and almost 90% in II). However, many parameters, such as the HBs, dihedral of hydroxyl groups, unit cell parameters and puckering parameters were very different from experimental data. The inconsistency of the inter-chain HB network may also be related to the non-bond force field parameters in Gr56A_{carbo}FF while that between experiment and simulation strongly suggests that the current force field needs to be optimized for better reproducing experimental observations.

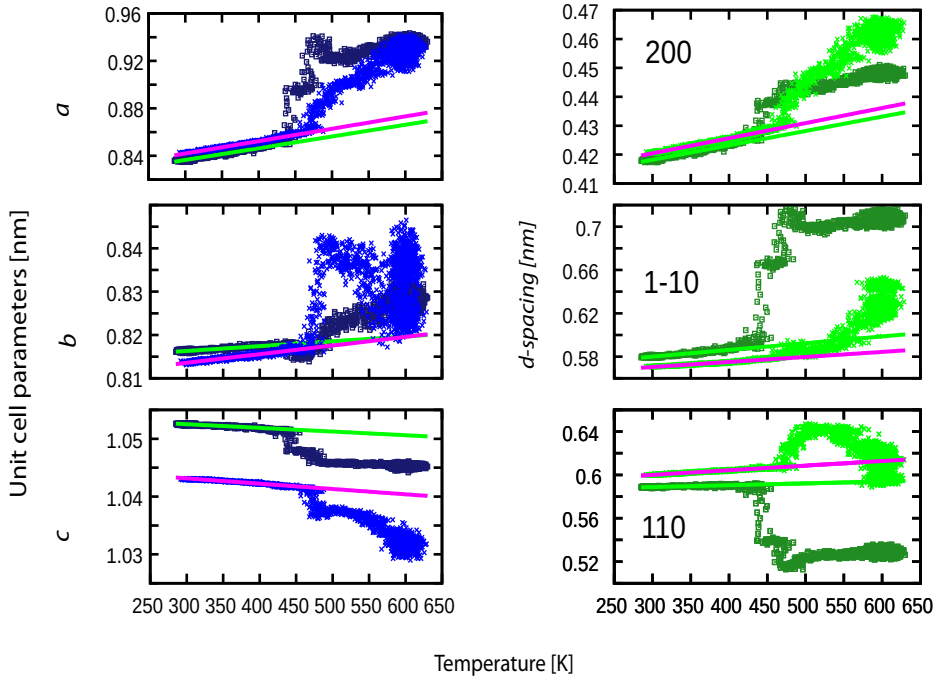

Figure III.6. Variation of unit cell parameters (*a*, *b*, *c*) and *d*-spacings (d_{110}, $d_{1\text{-}10}$, d_{200}) of Iβ as a function of temperature by using Gr53a6FF (deep blue and deep green) and Gr56A$_{carbo}$ FF (blue and green). The fitted lines were used to calculate the TEC.

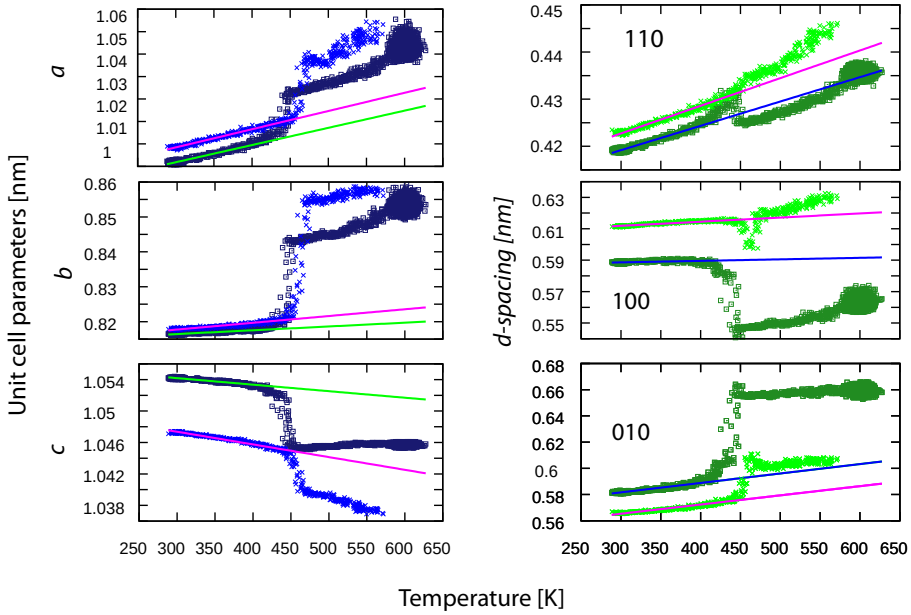

Figure III.7. Variation of the converted unit cell parameters (*a'*, *b'*, *c'*) and *d*-spacing (d_{110}, d_{100}, d_{010}) of Iα as a function of temperature by using Gr53a6FF (deep blue and deep green) and Gr56A$_{carbo}$ FF (blue and green). The fitted lines were used to calculate the TEC.

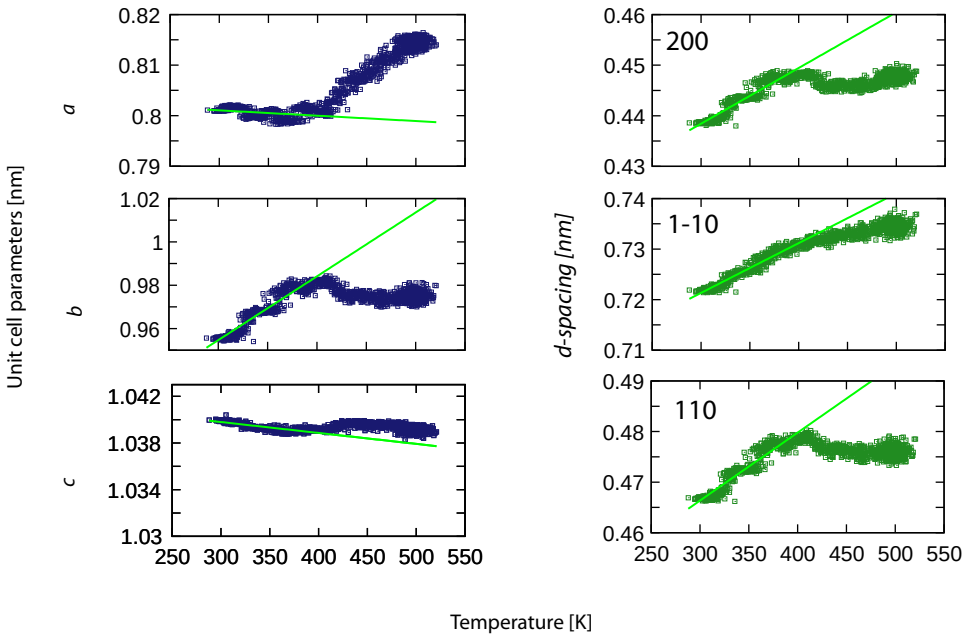

Figure III.8. Variation of unit cell parameters (*a*, *b*, *c*) and *d*-spacings ($d_{1\text{-}10}$, d_{010}, d_{110}) of II as a function of temperature by using Gr56A$_{carbo}$ FF (blue and green). The fitted lines were used to calculate the TEC.

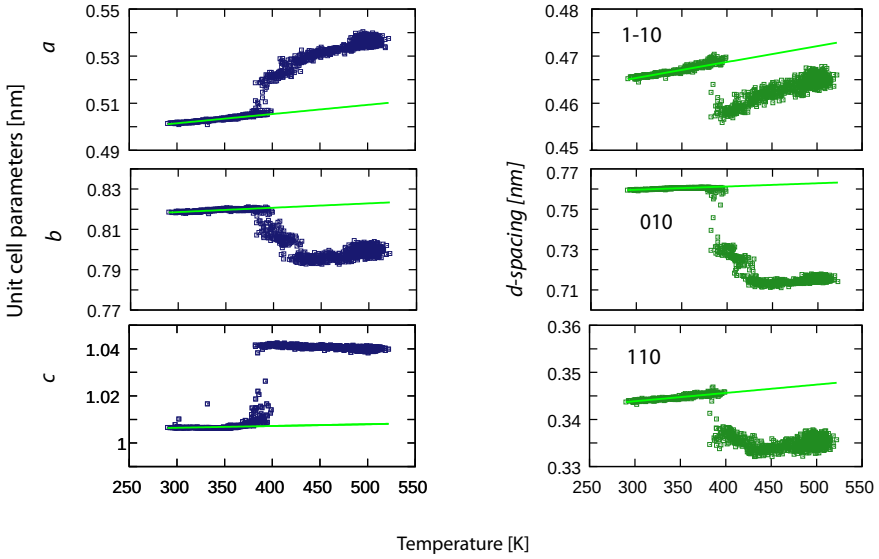

Figure III.9. Variation of unit cell parameters (*a*, *b*, *c*) and *d*-spacings ($d_{1\text{-}10}$, d_{010}, d_{110}) of III$_I$ as a function of temperature by using Gr56A$_{carbo}$FF (blue and green). The fitted lines were used to calculate the TEC.

III.4. Simulation of the thermal behavior

Figures III.6-9 show the variation of the unit cell parameters and *d*-spacings for cellulose Iβ, Iα, II and III_I as a function of temperature by using Gr53a6FF and Gr56A$_{carbo}$FF.

Upon heating, a transition can clearly be recognized by the sudden variation of the monitored dimension.

The thermal expansion coefficients (TEC) were calculated from the initial linear section, between 300 K and the transition temperature. TECs were estimated from 300 to 430 K for I, from 300 to 350 K for II and from 300 to 370 K for III_I. **Table III.5** lists the simulated TECs of the unit cell parameters and the main *d*-spacings of cellulose Iβ, Iα, II and III_I. The four cellulose allomorphs had positive TECs for the *a* and *b* unit cell parameters and *d*-spacings (except parameter *a* of cellulose II), and negative TECs, indicating a contraction, for parameter *c*.

Table III.5. Thermal expansion coefficients (TECs) of cellulose allomorphs along the indicated directions in the crystals calculated with Gr53a6FF and Gr56A$_{carbo}$FF. The calculation for Iα used the converted unit cell.

Allomorph	Force field	*a* $\times10^{-5}$	*b* $\times10^{-5}$	*c* $\times10^{-5}$	d_{200} $\times10^{-5}$	d_{110} $\times10^{-5}$	$d_{1\text{-}10}$ $\times10^{-5}$	d_{100} $\times10^{-5}$	d_{010} $\times10^{-5}$
Iβ	53a6	11.8	1.3	-0.6	11.9	2.9	10.7	--	--
	56A$_{carbo}$	12.3	2.5	-8.8	12.3	7.2	8.1	--	--
Iα	53a6	11.9	1.7	-0.8	--	12.4	--	1.6	12.4
	56A$_{carbo}$	12.2	3.0	-1.5	--	13.9	--	4.2	12.4
II	56A$_{carbo}$	-1.2	30.8	-0.8	24.4	28.1	13.4	--	--
III_I	56A$_{carbo}$	7.8	2.3	-0.9	--	5.5	7.5	7.0	2.1

III.4.1. Thermal behavior of cellulose Iβ

As seen in **Table III.5**, the anisotropic thermal expansion of Iβ was correctly reproduced with both Gr53a6FF and Gr56A$_{carbo}$FF. The TEC along the d_{200} direction is the highest. In good agreement with the experiments, the TEC along the *a*-axis (where the crystal is mostly stabilized by VdW interaction) is 6-9 times larger than that along the *b*-axis (where the 2D HB network occurs).

The anisotropic thermal expansion behavior is generally ascribed to the different types of interaction along the *a* and *b* axes, namely VdW interaction and hydrogen-bonding interaction, respectively (Wada 2002, Matthew *et al.* 2012). To verify this assumption, I have carried out MD simulation by setting all the atomic partial charges to zero, so that no hydrogen bonding interactions were taken into account. The same anisotropic thermal expansion was observed (**Figure III.10**), which suggests that this particular behavior may not be due to the hydrogen bonds. While HBs are important to stabilize the structure, they do not seem to be the origin of the anisotropic thermal expansion.

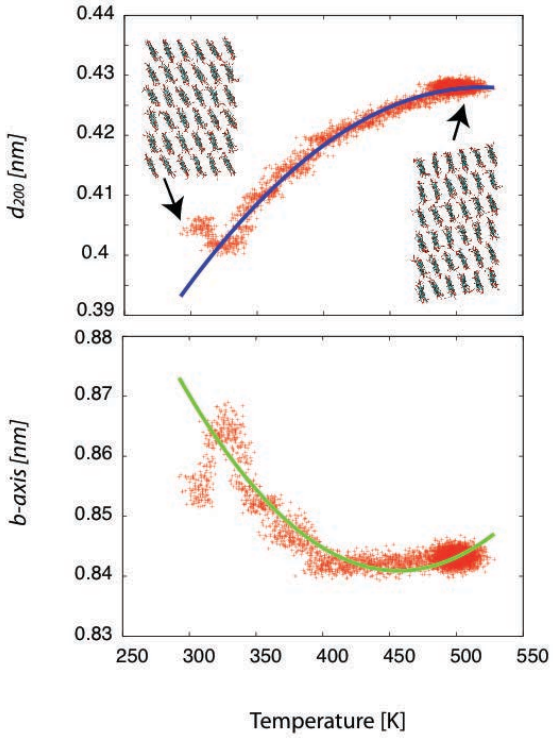

Figure III.10. Variation of d-spacing (d_{200}) and unit cell parameter *b* in Iβ as a function of temperature when all the atomic partial charges are set to zero. The inserted models correspond to the snapshots at room temperature (left) and high temperature (right).

It was reported that the crystal size influences the measured TEC (Nishiyama 2009). However, the simulation performed here used an infinite approach that represents the core structure of large crystals and thus ignores the size effect. TECs of d_{110} and $d_{1\text{-}10}$ in the large tunicate microcrystals differ by a factor of 3 (Wada 2002) which is in agreement with the model obtained using Gr53a6FF. In contrast, similar TECs have been measured in the smaller

microcrystals from algal cellulose (Wada *et al.* 2010), which agrees with those calculated with Gr56A$_{carbo}$FF. Note that the TEC along the *c*-axis from Gr56A$_{carbo}$FF is about 15 times larger than that from Gr53a6FF, and is almost of the same magnitude than the TEC along the *a*-axis, in contrast with the experiments.

III.4.1.1. Experimental reversible temperature-induced phase transition of cellulose Iβ

Figure III.11 shows the variation of the experimental *a*-axis parameter determined from the peak positions of the $1\bar{1}0$, 110 and 200 reflections measured on the diffraction pattern of an oriented tunicin fiber as a function of temperature, during heating and cooling. There is a sharp transition at 500 K upon heating with a sudden increase of the *a*-parameter, as previously observed (Hori and Wada 2005). Upon cooling, a similar transition, although broader, could be observed as well, at a slightly lower temperature. The transition was thus essentially reversible. The temperature difference at which the transition occurred between cooling and heating was probably due to the loss of crystallinity suggested by the slightly wider peak were after the annealing in helium. Indeed, the phase transition upon heating of a cellulose sample containing thinner crystallites occurs at a lower temperature and is less abrupt (Wada 2002).

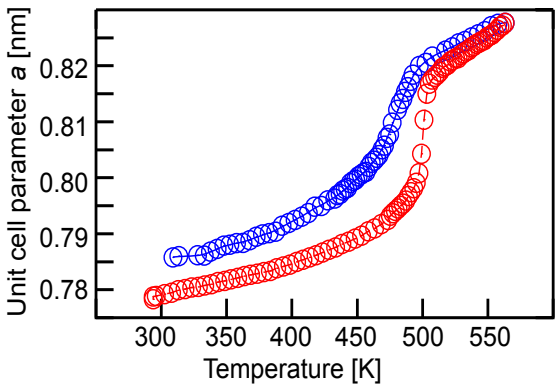

Figure III.11. Experimental variation of the crystal cell parameter *a* upon heating (red) and cooling (blue).

III.4.1.2. Simulated phase transition

Figures III.12, III.13, III.14, III.15 shows the histograms of the ω torsion angle, the puckering

parameters, the glycosidic bond angles and the τ2, τ3 and τ6 angles, respectively, as a function of temperature during heating by using Gr53a6FF.

I have seen that as the temperature increased, a phase transition occurred which is characterized by a departure from linearity of the thermal expansion along the *a*-axis (as described in **Figure III.6**). The phase transition occurred at 450 K, which is reasonably close to the experimental measurement of about 500 K.

When the temperature increased, the distribution of the conformational parameters became wider and no transition could be detected from the puckering parameters (**Figure III.13**) or the angles across the glycosidic bonds (**Figure III.14**). In contrast, at the transition temperature the hydroxymethyl and hydroxyl groups rotate (**Figures III.12 and III.15**). Rotation of the hydroxymethyl groups during heating agrees with the results of other groups (Bergenstråhle *et al.* 2007; Matthews *et al.* 2011; Zhang *et al.* 2011). Note that the *gg* orientation appeared prior the *gt* counterpart.

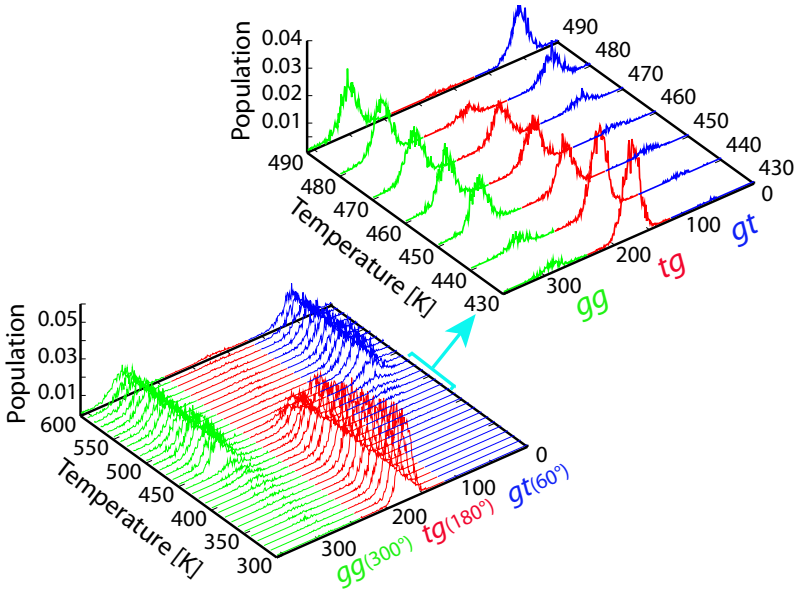

Figure III.12. Histogram of the torsion angle ω as a function of temperature. The top right figure is an enlargement around the phase transition in the 430-490 K temperature range (Gr53a6FF).

The monoclinic angle also drastically changed to an acute angle (72°) above 500 K. In other words, according to the standard definition (French and Howley 1989), the parallel-up structure

changed to a parallel-down structure. This behavior was independent of the box size or the choice of box boundary. On the other hand, the earlier results (Bergenstråhle *et al.* 2007) were reproduced by using the same simulation box together with the shear-compressibility set to 0. Zero shear compressibility is normally applied for simulation of protein in water solution, as the protein would anyway have the freedom to deform due to the water molecules. However, in the case of our infinite crystal, the shear compressibility must be given a finite value to allow adoption of the most stable intermolecular organization. The result from Bergenstråhle *et al.* (2007) was biased by the fixed angle of the periodic box due to the shear compressibility of zero.

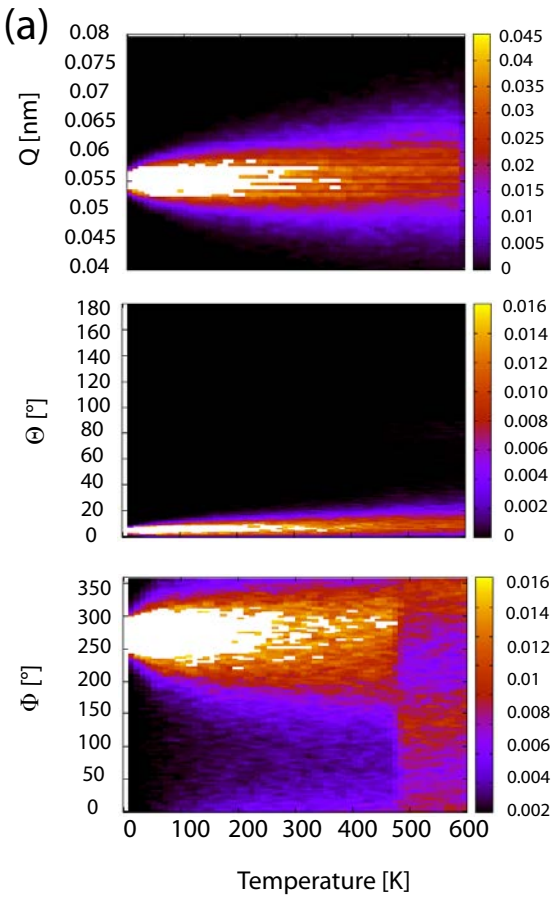

Figure III.13. Histogram of puckering parameters as a function of temperature during heating by using Gr53a6FF

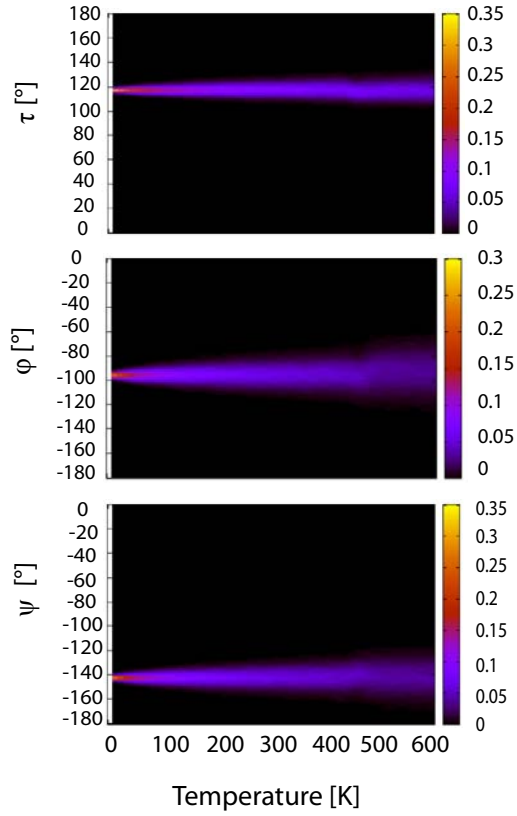

Figure III.14. Histogram of the glycosidic angles as a function of temperature upon heating by using Gr53a6FF.

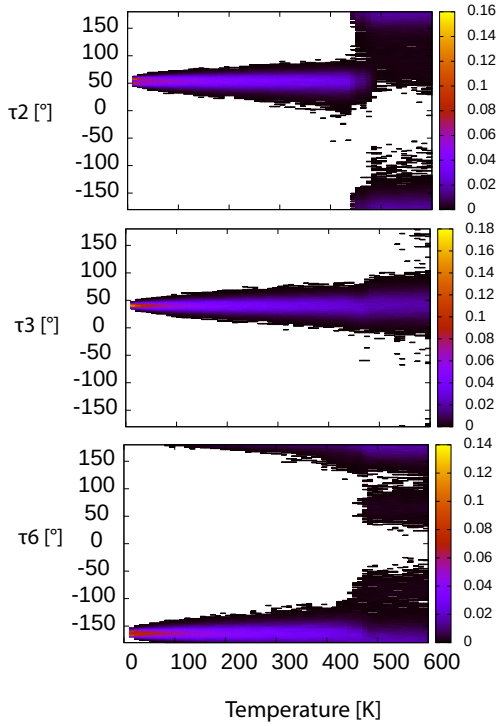

Figure III.15. Histogram of the τ2, τ3 and τ6 angles as a function of temperature during heating by using Gr53a6FF.

When the system was cooled to room temperature, the ω torsion angle mainly explored the *gt* and *gg* conformation without any trace of *tg*. (**Figure III.16)**. In addition, no discontinuous event was observed for the unit cell parameters (blue curve in **Figure III.17**). This is obviously a discrepancy with respect to the experiments (**Figure III.11)**. An irreversible transition was also obtained when using Gr56A$_{carbo}$FF.

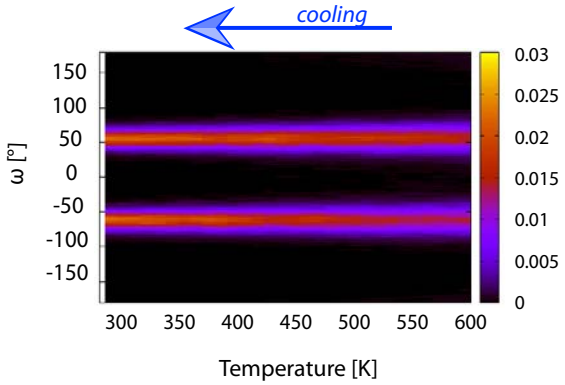

Figure III.16. Modeled variation of the hydroxymethyl group ω upon cooling (Gr53a6FF)

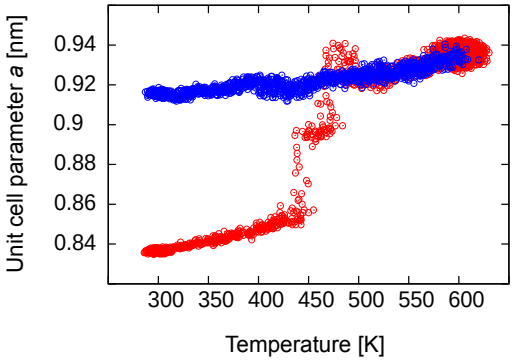

Figure III.17. Modeled variation of the crystal cell parameter *a* upon heating (red) and cooling (blue) (Gr53a6FF).

Comparing the potential energies at 300 K before and after annealing, it was found that the total enthalpy was lowest for the annealed structure with *gt* and *gg* conformers (**Table III.6**). Thus I can conclude that the irreversible behavior is not a consequence of the limited cooling rate, but is due to the force field parameters that describe the energy profile for rotation of the hydroxymethyl group. The optimization of the force field will be described in the next chapter.

Table III.6. Energy per glucose of cellulose Iβ at 300 K at the beginning of the simulation (quoted heating) and after annealing (quoted cooling).

K_1 (kJ/mol)	K_3 (kJ/mol)	E (kJ/mol)	
		Heating	Cooling
9.35	9.50	352.69	350.85

III.4.2. Thermal behavior of cellulose Iα

The TECs of Iα simulated by Gr53a6FF were similar to that of Iβ, with the greater TEC were for *a*-axis, d_{110} and d_{010} (d_{200} and $d_{1\text{-}10}$ for Iβ). The TECs were small along the *b* and *c* axes. The Gr56A$_{carbo}$FF gave almost identical results, particularly the small contraction of the *c*-axis, which differs from the overestimated one for Iβ. Comparing the two force fields, the TECs of Iα from Gr56A$_{carbo}$FF were slightly larger in all the direction than that from Gr53a6FF.

III.4.3. Thermal behavior of cellulose II and III$_I$

The order of the simulated TECs for cellulose II was TEC-d_{110} > TEC-d_{200} > TEC-d_{110} > 0 > TEC-*c*-axis, whereas it is measured TEC-d_{200} > TEC-d_{110} > 0 > TEC-*c*-axis > TEC-$d_{1\text{-}10}$ (Wada 2002). Besides, the simulated TEC along the *b*-axis was the largest below 350 K, which contrasts with the negative TEC observed experimentally. Only the contraction along the *c*-axis is in agreement with experiment. From the variation of unit cell parameters and *d*-spacings, the transition in cellulose II occurred between 350 and 400 K, which is also different from experiments where no transition occurred. These results indicate that the thermal properties of cellulose II were poorly reproduced.

The simulated TEC of cellulose III$_I$ along *a* and *b* axes, $d_{1\text{-}10}$ and d_{110} were within a reasonable range compared to the experimental values. However, that of d_{010}, which is measured negative, was estimated positive. A transition around 380 K can be seen in **Figure III.10**, which is 100 K lower than the experimental transition temperature.

III.5. Conclusions

In this chapter, I have modeled the structures and simulated the thermal behavior of four

crystalline celluloses and systematically compared the simulation results with the experiments. A good agreement has been found in the case of:

- the staggered structures of I and II and the non-staggered ones of III_I,
- the $^4C_1$ ring conformation of pyranose rings,
- the linear conformation of the chains,
- the conformation of the hydroxymethyl groups by using $Gr56A_{carbo}FF$
- the 2D HB network of allomorph I and the 3D ones of II and III_I,
- the anisotropic thermal expansion coefficients,
- the phase transition of cellulose Iβ during heating

However, several discrepancies were also found regarding:

- the slightly overestimated puckering amplitude by using $Gr56A_{carbo}FF$,
- the occurrence of *gg* conformation for allomorphs II and III_I by using Gr53a6FF,
- the lack of distinction between the simulated HB networks in allomorphs Iβ and Iα,
- the poorly reproduced 3D HB network in allomorphs II and III_I,
- the poorly precision in the unit cell parameters,
- the irreversible transition of cellulose Iβ

Considering that the force field method is only a simplification of reality, the agreement with experimental data was not expected to reach the precision of the first principle calculation. In summary, the GROMOS force fields could reproduce reasonable structure and thermal properties for cellulose I during heating, but not for cellulose II and III_I and the behaviors of them during cooling phase. These discrepancy found here suggested that the currently used force fields (Gr53a6FF and $Gr56A_{carbo}FF$) need necessary modification. The detailed optimization of these two force fields will be presented in the next chapter.

Chapter IV - Optimization of key parameters in GROMOS force fields

IV.1. Introduction

Upon heating at about 500 K, crystalline cellulose undergoes a phase transition, yielding the so-called "high temperature phase" (Wada 2002, Hori and Wada 2005, Wada *et al.* 2010), identified by an expansion of the unit cell that increases discontinuously by a few percent when this temperature is reached. The modeling of this transition by molecular dynamics (MD) has been performed using different force fields (Bergenstråhle *et al.* 2007; Matthews *et al.* 2011; Zhang *et al.* 2011; Chen *et al.* 2012). These MD simulations were able to reproduce the unit cell expansion of cellulose during the phase transition and showed that upon heating the hydroxymethyl groups shifted from the *tg* conformation to the *gt* and *gg* ones.

Experimentally, the high-temperature phase reverts to its initial low-temperature counterpart upon cooling. This was deduced by comparing the X-ray diagrams of cellulose before and after annealing, which shows that initial unit cell is restored (Nishiyama *et al.* 2002), ^{13}C NMR data (Horii *et al.* 1987) together with crystal structure refinement (Nishiyama *et al.* 2002) also indicate that the *tg* conformation of the hydroxymethyl groups is recovered in the cooled specimens. In the MD calculations shown in **Chapter III**, however, the sample recovery was not observed upon cooling, as briefly mentioned in the recent paper of Zhang *et al.* (2011) based on the GLYCAM force field. When using the Gr53a6FF, I also noticed that the hydroxymethyl group of the cooled cellulose structure did not go back to the experimentally determined *tg* conformation. For cellulose, the tendency to favor the calculated *gt* and *gg* conformations instead of the experimental *tg* is also seen in several room temperature MD studies (Matthew *et al.* 2006) as well as in structure predictions studies (Viëtor *et al.* 2000).

In **Chapter III**, I have seen that the GROMOS force fields overestimated the distance between the cellulose layers and underestimated the monoclinic angles at both room temperature and high temperature phase. The structure with poorly simulated unit cell parameters did not represent the experimental structure. Thus, the origin of the discrepancies should be identified for improving the parameters of the GROMOS force fields, and also to allow more realistic

predictions of other properties of cellulose. With this purpose in mind, I have pointed out three sets of key parameters in the GROMOS force field that directly relate to the reversibility of the temperature-induced phase transition of cellulose Iβ and to the precision of the unit cell parameters of four cellulose allomorphs.

IV.2. Simulation procedure

The model construction of cellulose and simulated annealing was described in **§ III.2**.

IV.3. Parameter optimization strategy

Sets of Lennard-Jones (LJ) parameters that minimize the deviation between simulated and experimental unit cell parameters of cellulose Iα, Iβ, II and III_I were searched. The simplex method of Nelder and Mead was used (Nelder and Mead 1965). The target function *F* can be written as:

$$F = \sum_{i=1}^{i=n} \left(100 \times \frac{X_i^{sim} - X_i^{exp}}{X_i^{exp}} \right)^2$$

where $X_i^{\exp}$ is an experimental unit cell parameter and X_i^{sim} is the corresponding simulated one. The starting LJ parameters were those of Gr53a6FF and the initial step sizes were set to 40 % of initial values. The maximum number of parameters *n* for 4 allomorphs is 24, but the subsets of target parameters were also used for different purposes.

IV.4. Optimization of the torsional parameter of Gr53a6FF

IV.4.1. Torsional bias energy profile of the hydroxymethyl group in Gr53a6FF

The distribution of the ω torsion angle is directly related to the "proper dihedral" term of the potential energy function in the GROMACS program (Hess *et al.* 2008, Lins and Hünenberger 2005). The dihedral energy (E_{dih}) of exocyclic hydroxymethyl group takes the following analytical form:

$$E_{dih} = K_1(1 - \cos\omega) + K_3(1 + \cos 3\omega)$$

where ω is the torsion angle and K_1 and K_3 are force constants of 9.35 and 9.5 kJ/mol,

respectively, in the Gr53a6FF. K_1 adds an energy penalty to the *tg* conformation and K_3 defines the height of the energy barriers between the three staggered positions. These values are based on quantum mechanic (QM) calculation of isolated methyl-glucoside. The resulting E_{dih} profile around the dihedral angle ω is shown in **Figure IV.1**.

A close inspection of the published energy profiles as a function of the torsion angle ω of glucose reveals a strong dependence on the orientation of the secondary hydroxyl groups: clockwise and counterclockwise. Also, noticeable differences can be seen between the reference energy curves, calculated using QM, and the predicted one using the 45a4 parameters (Lins and Hünenberger 2005). The torsion energy estimated with the molecular mechanics method was 10 kJ/mol higher than the one using the QM method for some values of ω. Furthermore, in native cellulose, the orientation of the secondary hydroxyl groups is neither clockwise nor counterclockwise, which are the only arrangements considered in the QM calculation.

In Lins's paper, the 20 ns simulation of β-D-glucopyranose in explicit water system showed 0% *tg* conformers (Lins and Hünenberger 2005). The analysis of the NMR coupling constants using Karplus-type equations predicts a population of *tg* ranging from negative value (Bock and Duus 1994) to 9% (Thibaudeau *et al.* 2004). Thus, the dihedral parameters are still far from being established.

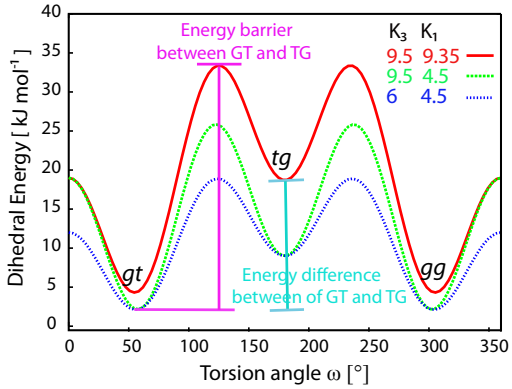

Figure IV.1. Dihedral energy profile of the rotation of the hydroxymethyl group. Continuous red curve: native parameters K_3 = 9.5, K_1 = 9.35; green dotted line: modified parameters K_3 = 9.5, K_1 = 4.5; blue dotted line: K_3 = 6, K_1 = 4.5.

IV.4.2. Simulated annealing by using torsional-parameter-revised Gr53a6FF

The two force constants K_1 and K_3 of the dihedral energy term were systematically modified to study their effect on the thermal behavior of the crystal. **Table IV.1** summarizes the reversibility of the transition for different combinations of the two parameters. When the barrier parameter K_3 was kept to its original value, reversibility was observed when K_1 was lower than or equal to 4.5 kJ/mol. By lowering K_3, a higher K_1 also became acceptable.

Table IV.1. Reversibility of the transition for selected values of K_1 and K_3. The × symbol was used when the transition was irreversible, ○ when the transition was reversible and blank when it was not tested. The slowest cooling rate was 1 K/ns.

K_3 (kJ/mol)	K_1 (kJ/mol)							
	9.35	8	7	6	5	4.5	4	3
9.5	×	×	×	×	×	○	○	○
8					×	○	○	○
7					×	○	○	○
6	×			×	○	○	○	○
5	×			×	○	○	○	
4	×				○			
3	×							

IV.4.2.1. Effect of simulation rate on reversibility by using Gr53a6FF

Figure IV.2 shows the total energy per residue before heating and after cooling as a function of K_1 when K_3 is fixed to 9.5 kJ/mol. All the annealing simulation was done under the cooling rate of 0.04 ns/K. Identical energies for the structures before heating and after cooling, could be obtained when K_1 is not larger than 4 kJ/mol. When K_1 is larger than 4 kJ/mol, the transition is irreversible and the energy difference between the heating structure and the cooling structure at 300 K gradually decrease from 4.3 to -1.8 kJ/mol with the increase of K_1 from 4.5 to 9.35 kJ/mol. It was until I used a slower cooling rate no less than 0.1 ns/K, that a reversible transition could also be obtained when K_1 equaled 4.5 kJ/mol, as shown in **Table IV.2**. It seems the smaller the energy difference, the slower the cooling rate that required. However, a cooling rate of 1 ns/K still could not allow obtaining the reversible transition when K_1 equals to 5 kJ/mol. Due to the expensive cost of computational time, the test was not performed further. From the perspective

of energy, theoretically, the reversible transition could be obtained when K_1 is less than 8 kJ/mol if the annealing rate is slow enough as the cooling rate in reality.

Table IV.2. Energy per glucose of cellulose Iβ at 300 K at the beginning of the simulation (quoted heating) and after annealing (quoted cooling) calculated with the original and modified values of the K_1 and K_3 parameters.

K_1 (kJ/mol)	K_3 (kJ/mol)	E (kJ/mol)	
		Heating	Cooling
9.35	9.50	352.69	350.85
4.50	9.50	342.89	342.92
4.00	9.50	341.90	341.91
3.00	9.50	339.88	339.90

IV.4.2.2. Transition temperature dependence

Figure VI.3 shows the population of the *tg* conformer as a function of temperature with different values K_1. Upon heating, the lower the value of K_1, the higher the transition temperature. In both the modeling and the experiments, the transition temperature showed a significant hysteresis (**Figures III.13** and **IV.3**). The transition temperature was systematically higher in the heating phase than in the cooling one. The hysteresis is however predicted larger in the modeling than the experiment. This is probably due to the very fast heating and cooling rates compared to the standard experimental timeframe. However, lowering the heating and cooling rates down to 1 K/ns does not significantly reduce the hysteresis.

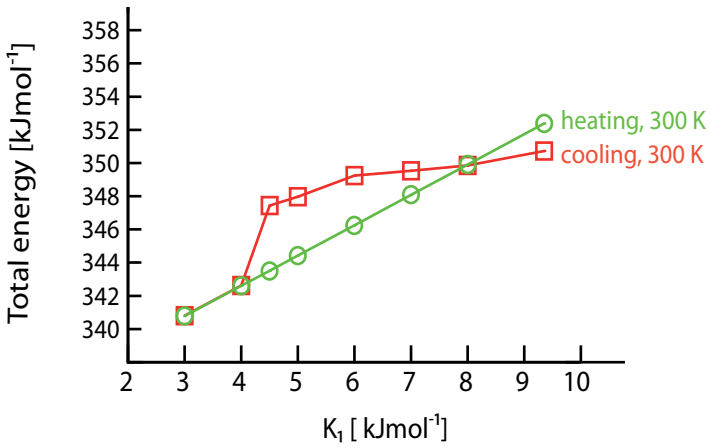

Figure IV.2. Energy per residue of cellulose Iβ at 300 K as a function of K_1 before heating (green) and after cooling (red), with a cooling rate of 0.04 ns/K.

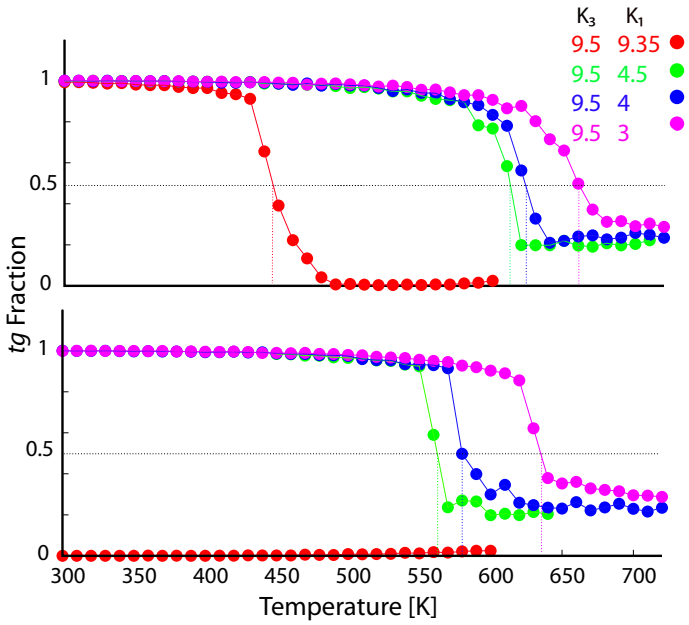

Figure IV.3. Amount of *tg* conformation during the heating (top) and cooling (bottom) processes for different K_1 values.

IV.4.2.3. Unit cell parameter a, hydroxymethyl group ω and hydroxyl groups

The simulated variation of the *a* parameter as a function of temperature during heating (red traces) and cooling (green traces) is shown in **Figure IV.4**. Whereas **Figure IV.4-top** is obtained with the standard values of $K_1 = 9.35$ kJ/mol and $K_3 = 9.50$ kJ/mol, **Figure IV.4-bottom** results from the lowering of K_1 to 4.5 kJ/mol. When shifting to this last K_1 value, one sees that during the cooling, the unit cell parameter *a*, which remained constant in

the standard situation, decreased discontinuously to reach its starting pre-heating value, indicating a reversibility of the process. On an other hand, the ω from modified force filed (**Figure IV.5**) also changed back to experimental *tg* conformation from the high temperature of *tg*, *gg* and *gt* situation, at the temperature where the *a*-axis suddenly decreased to its initial value. The corresponding torsional bias profile, E_{dih}, is given in **Figure IV.2**. Lowering K_1 to 4.5 kJ/mol decreases the relative energy of the *tg* conformer and lowers the energy barrier for the *gg* or *gt* to *tg* transition.

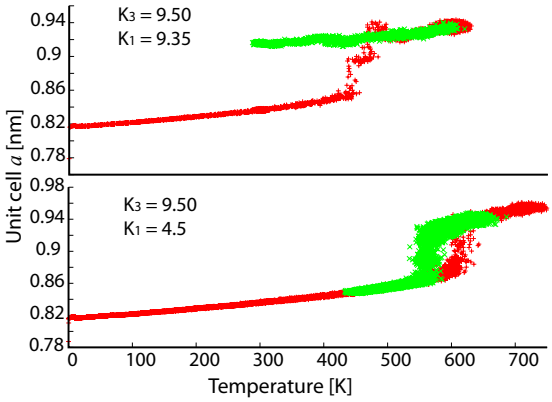

Figure IV.4. Unit cell parameter *a* as a function of temperature before (top) and after (bottom) modifying the torsional parameter. Heating in red and cooling in green. The top figure has been presented in **Chapter III** but in the plot here, the temperature varies between 0 and 600 K, whereas in **Chapter III**, it was between 300 and 600 K.

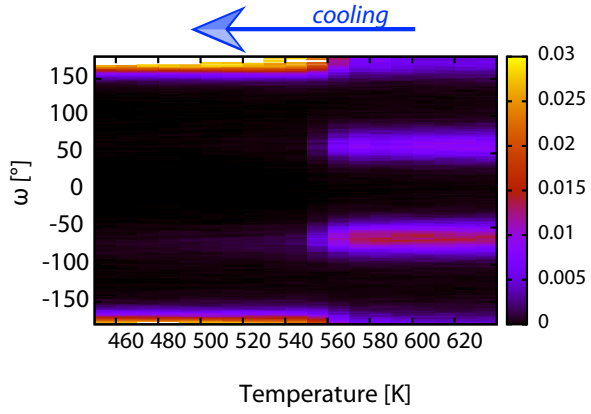

Figure IV.5. Histogram of the conformations of the hydroxymethyl group (torsion angle ω) as a function of temperature during cooling simulated by using the modified force field (K_1 = 4.5 kJ/mol).

IV.4.3. Optimization of the torsional parameters in Gr56A$_{carbo}$FF

MD simulation on the phase transtion of Iβ was also perform with the lasted optimzed GROMOS force field, Gr56A$_{carbo}$FF, in which the torsional parameters were systematically varied. The dihedral energy (E_{dih}^{Gr56}) of exocyclic hydroxymethyl group takes the following analytical form:

$$E_{dih}^{Gr56} = K_1[1-\cos(\omega)] + K_2[1+\cos(\omega+60°)] + K_3[1+\cos(3\omega)]$$

where ω is the torsion angle and K_1, K_2 and K_3 are force constants which respectively have the values of 4.5, 2 and 5.9 kJ/mol in the Gr56A$_{carbo}$FF. K_1 adds an energy penalty to the *tg* conformation, K_3 defines the height of the energy barriers between the three staggered positions and K_2 adjusts the energy difference between *gg* and *gt*.

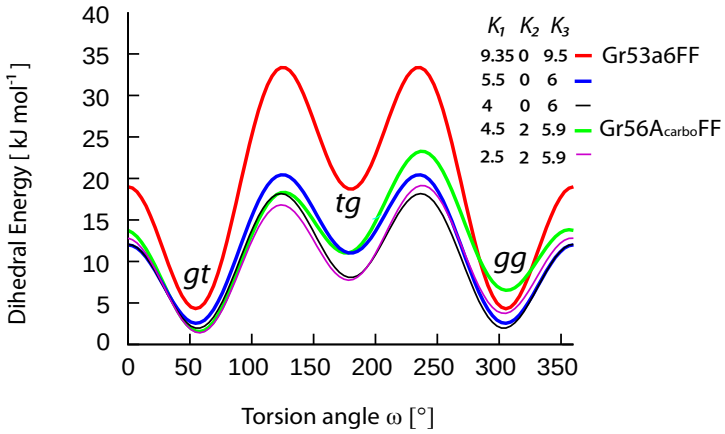

Figure IV.6. The dihedral energy profile of the rotation of the hydroxymethyl group, red curve : Gr53a6FF K_3 = 9.5, K_2 = 0, K_1 = 9.35, blue line: modified Gr53a6FF K_3 = 6, K_2 = 0, K_1 = 5.5, black: modified Gr53a6FF K_3 = 6, K_2 = 0, K_1 = 4, green: Gr56A$_{carbo}$FF K_3 = 5.9, K_2 = 2, K_1 = 4.5, purple: modifed Gr56A$_{carbo}$FF K_3 = 5.9, K_2 = 2, K_1 = 2.5.

Both of these two sets of torsional parameters were optimized to be able to reproduce the experimental data concerning about the hydroxymethyl rotamer populations of glucose in water. The Gr53a6FF are against the experimental data (Nishida *et al.* 2008) that suggested a higher population of *gg* than *gt*, whereas the Gr56A$_{carbo}$FF selected the experimental data (Brochier-Salon and Morion 2000, Tibaudeau *et al.* 2004) that suggested a higher population of *gt* than *gg*.

Compared with Gr53a6FF, the energy profile of ω in Gr56A$_{carbo}$FF has smaller energy barrier due to the smaller K_3 and K_1 and has a higher energy of *gg* than *gt* due to the additive K_2. By decreasing K_3 to 6 and K_1 to 5.5 in Gr53a6FF, the energy profile became close to the one in Gr56A$_{carbo}$FF, although the energy barriers are slightly different.

In the annealing process by using Gr56A$_{carbo}$FF with the cooling rate of 1 ns/K, I found an irreversible phase transition of cellulose Iβ. However, the energy of the cooling structure at 300 K is slightly higher than the heating structure at 300 K, which means the transition could be reversible if the cooling rate could be slow enough according to tendency shown in **Figure IV.2.** For being able to observing the reversibility at hundred nanoseconds scale, the torsional parameter K_1 in Gr56A$_{carbo}$FF was also modified. By decreasing K_1 to 2.5 kJ/mol, which yields a similar energy profile compared with K_3 equals to 6 kJ/mol and K_1 equals to 4 kJ/mol in Gr53a6FF (**Figure IV.6**), a reversible transition could also be obtained from MD simulation. This result indicates that the force field parameters in Gr56A$_{carbo}$FF should be further optimized to be able to reproduce the experimental reversible phase transition of cellulose Iβ.

IV.5. Optimization of the non-bonded force field parameters

IV.5.1. Simulated unit cell structure of cellulose Iβ by molecular modeling

Despite the well-reproduced reversible expansion in *a*-axis during annealing process by revising selected torsional parameters, the *a*-parameters estimated by using the native and revised Gr53a6FF at 300 K are still both significantly different from the experimental value. Actually, none of the carbohydrate force fields could accurately reproduce the unit cell parameters of crystalline cellulose.

Table IV.3 lists the reported values of predicted unit cell parameters of the Iβ allomorph. This allomorph is the most abundant form in the terrestrial plants and consequently the most studied by molecular modeling, using various force fields and DFT methods. As the optimized molecular shapes are roughly similar in the different force fields, the unit cell parameters are expected to be most sensitive to the non-bonded energy terms, *i.e.,* the electrostatic and the dispersion terms. The unit cell parameters that vary the most as a consequence of using

different force fields are *a* and *γ*. Cellulose Iβ can be described as a stack of sheets mainly stabilized by HBs in the plane of pyranose rings. The *a* and *γ* parameters are directly related to the distance between the sheets and to the staggering from one layer to another.

One observation from **Table IV.3** is that the old version of GROMOS (GROMOS 87) better reproduces the experimental data than the more recent versions, Gr53a6 and Gr56A$_{carbo}$, the latter being specially optimized for carbohydrates. Among the different aliphatic carbons, the atom type CH1 carrying one hydrogen atom is the most important for cellulose, corresponding to five out of the six carbons in a glucose residue.

Table IV.3. Survey from the literature of the crystal cell parameters of cellulose Iβ simulated by using force field and DFT methods. The experimental ones are also reported.

Methods		a (Å)	b (Å)	c (Å)	γ (°)	Reference
FF	Gr87	7.60	8.10	10.40	96.0	Kroon 1997
	Gr45a4/53a6	8.34	8.17	10.50	90.9	Chen *et al.* 2012 Bergenstrahle 2007
	Gr56A$_{carbo}$	8.42	8.14	10.43	92.9	*this work*
	MM3	7.50	8.70	10.36	94.1	Vietor 2000
	CHARMM C27	8.47	8.11	10.50	90.0	Matthews 2006
	CHARMM C35	7.96	8.35	10.44	98.3	Gross 2011
	PCFF	7.90	8.54	10.70	94.4	Mazeau 2005
	GLYCAM 06	7.63	8.23	10.80	97.2	Zhang 2011 Matthews 2012
DFT	PBE	8.70	8.23	10.46	95.5	Bucko 2010
	PBE-D2	7.65	8.14	10.39	96.5	Bucko 2011
X-ray	Experimental	7.784	8.201	10.38	96.5	Nishiyama 2002

IV.5.2. Lennard-Jones parameters in the generation of GROMOS force fields

The LJ parameters of this CH group have significantly changed in the different GROMOS force fields. The LJ parameters of CH1 along with the distance (σ) and energy (ε) at the minimum between two interacting CH1 groups are given in **Table IV.4** and their energy profile are plotted in **Figure IV.7**. The first version, GROMOS 87, gave the minimal-energy distance of 4.232 Å. This was modified to 3.800 Å in the subsequent release in 43a1, and finally to

5.019 Å for version 45a3 and the more recent ones. This corroborates with the tendency of unit cell parameter *a* reported using GROMOS 87 and the GROMOS versions after 45a3, 7.6 Å and 8.34 or 8.40 Å, respectively.

The LJ potential energy is described by:

$$E_{LJ} = \frac{C12_{ij}}{r^{12}} - \frac{C6_{ij}}{r^{6}}$$

where *r* is the distance between the two atoms. The distance at which the energy is zero can be calculated by:

$$\sigma = \sqrt[6]{\frac{C6_{ij}}{C12_{ij}}}$$

and the depth of the potential energy by:

$$\varepsilon = \frac{C6_{ij}^{2}}{4C12_{ij}}$$

These values were calculated from reference papers (Koehler *et al.* 1987; Schuler *et al.* 2001; Lins and Hünenberger 2005; Oostenbrink *et al.* 2004; Hansen and Hünenberger 2011).

Table IV.4. The original Lennard-Jones parameters ($C6^{1/2}$ and $C12^{1/2}$), distance (σ) and energy (ε) at the minimum of the CH1 atom type in the different generations of the GROMOS force field. The parameters optimized in this work are also included.

Force field	σ (nm)	ε (kJ mol^{-1})	$C6_{ij}^{1/2}$ (kJmol^{-1}nm^{6})$^{1/2}$	$C12_{ij}^{1/2}$ (kJmol^{-1}nm^{12})$^{1/2}$
GROMOS 87	0.4232	0.5442	0.11180	8.472×10^{-3}
43a1	0.3800	0.3139	0.06148	3.373×10^{-3}
43a2	0.3800	0.3139	0.06148	3.373×10^{-3}
45a3	0.5019	0.09489	0.07790	9.850×10^{-3}
45a4	0.5019	0.09489	0.07790	9.850×10^{-3}
53a5, 53a6	0.5019	0.09489	0.07790	9.850×10^{-3}
56A$_{carbo}$	0.5019	0.09489	0.07790	9.850×10^{-3}
Optimization I	0.4473	0.1894	0.07790	6.972×10^{-3}
Optimization II	0.4462	0.1841	0.07623	6.772×10^{-3}

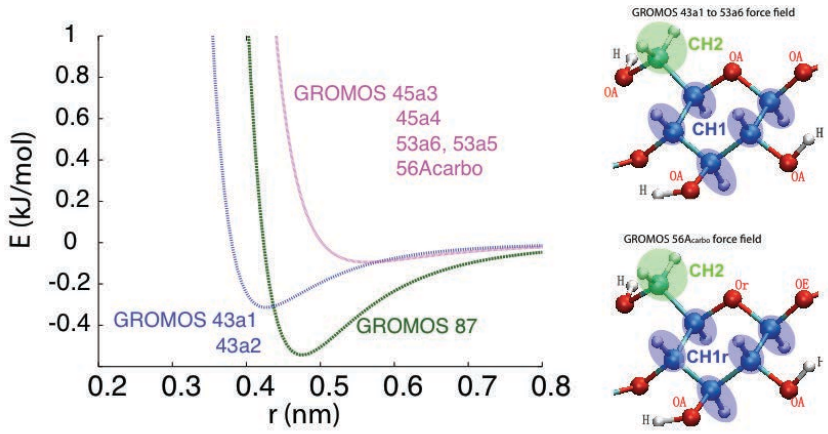

Figure IV.7. Energy profiles of the CH1–CH1 Lennard-Jones interaction defined in GROMOS force fields and the atom types of glucose ring defined in GROMOS force fields.

The difference in the parameters most likely comes from the target properties on which parameter optimization was carried out: the old parameter set of GROMOS 87 was based on the unit cell parameters of a group of alkanes and adamantanes (Dunfield *et al.* 1978) whereas the more recent versions were based on heat of vaporization, vapor pressure or liquid density of aliphatic molecules (Schuler *et al.* 2001). The attractive term, $C6^{1/2}$, has seen less variation over the history than the repulsive $C12^{1/2}$ one, swinging from 0.1118 to 0.0615 (kJ mol^{-1} $nm^{6})^{1/2}$.

IV.5.3. Atomic partial charge assignment in GROMOS, CHARMM and GLYCAM force fields

Table IV.5 lists the atomic charges for glucose in GROMOS, CHARMM and GLYCAM force fields. The partial charges on some atoms differ considerably, which is not surprising considering that they have been determined according to different methods. GROMOS 87 partial charges on the atoms were derived from fitting to the thermodynamic properties of pure liquids and aqueous solutions (Lins and Hünenberger 2005), whereas those of GROMOS 45a4/53a6 (Lins and Hünenberger 2005; Oostenbrink *et al.* 2004) and GROMOS $56A_{carbo}$ (Hansen and Hünenberger 2011) were based on RESP methods and the bond-increment approach (Gasteiger and Marsili 1980) respectively. GLYCAM06 (Kirschner *et al.* 2008) uses RESP methods with a restraint weight of 0.01 whereas those of CHARMM C35 (Guvench *et al.*

2009; Foley *et al.* 2011) are initially based on QM calculations and subsequently modified in order to correct some unrealistic solvation enthalpies of pure substances or simple mixtures, and further refined to reproduce bulk properties of neat liquids.

Table IV.5. The partial atomic charges of glucose in GROMOS, CHARMM, and GLYCAM force fields (Koehler *et al.* 1987; Schuler *et al.* 2001; Lins and Hünenberger 2005; Oostenbrink *et al.* 2004; Hansen and Hünenberger 2011).

	GROMOS 87	**GROMOS** 53A6	**GROMOS** 56A$_{carbo}$	**CHARMM** C35	**GLYCAM** 06
C6	0.15	0.232	0.232	0.23	0.282
O6	-0.548	-0.642	-0.642	-0.65	-0.688
HO6	0.398	0.41	0.41	0.42	0.424
C2	0.15	0.232	0.232	0.23	0.31
O2	-0.548	-0.642	-0.642	-0.65	-0.718
HO2	0.398	0.41	0.41	0.42	0.437
C3	0.15	0.232	0.232	0.23	0.284
O3	-0.548	-0.642	-0.642	-0.65	-0.709
HO3	0.398	0.41	0.41	0.42	0.432
O5	-0.360	-0.48	-0.464	-0.40	-0.471
C5	0.160	0.376	0.232	0.200	0.225
C1	0.400	0.232	0.464	0.38	0.384
O1	-0.360	-0.360	-0.464	-0.36	-0.468
C4	0.160	0.232	0.232	0.18	0.276

The comparison of the non-bonded parameters of the 53a6 and 56A$_{carbo}$ force fields shows that the CH1 dispersion parameters are identical, as are most of the atomic partial charges. The charges on the atoms C1, O1, C5, O5 however differ in these two force fields. The simulated cellulose crystals using these force fields give almost identical *a* parameters and a strong discrepancy for the γ parameter. This comparison reveals the high sensitivity of the monoclinic angle of the crystal cell on the atomic partial charges.

I have compared the parameters in the different generations of the GROMOS force field as well as most popular force-fields applied to cellulose crystals and performed MD simulations of four allomorphs of cellulose: Iα, Iβ, II and III$_I$ to estimate their ability to reproduce the experimental

unit cells. By modifying some of the parameters, I have investigated the influence of the non-bonded parameters on the ultrastructural organization of the cellulose chains in the crystal structures and proposed a general strategy for improving force fields to better simulate crystalline polysaccharides.

IV.5.4. Optimization of Lennard-Jones parameters of CH1, CH2, OA

Initially, I looked for a set of LJ parameters that could predict the dimensions of the unit cells of the four allomorphs with good accuracy. In this section, the partial charges on the atoms being considered were initially those of the native Gr53a6FF. Each optimization of the LJ parameters involved four independent simulations: in each simulation, three allomorphs of cellulose out of the four were selected for the refinement and the remaining one was used to test the optimized parameters.

In the Gr53a6FF, the glucose residue is defined by four types of LJ atoms: CH1, CH2, OA and H (**Figure IV.7-right**). However, the distribution of the non-bonded CH1 to CH1 distances observed in crystal structures from the Cambridge database strongly depends on the type of other atoms attached to the carbon atom (**Figure IV.8-top**). Carbon atoms having more oxygen as neighbors have a clear tendency to be closer to each other. In cellulose, the anomeric carbon bears two oxygen atoms whereas all the other CH1 atoms are bound to two carbon atoms and one oxygen atom. I thus defined a supplementary carbon atom type, C1, specific to the anomeric carbon. Similarly, the alcohol and ether oxygen atoms would have different steric effect, so I introduced the atom type Or for the ring oxygen and the glycosidic oxygen. The parameters of C1 and Or where those of CH1 and OA atom types as defined in the original force field. Neglecting the hydrogen, this makes a total of 5 atom types requiring 10 LJ parameters to be optimized. Sets of these 10 LJ parameters that significantly improve the prediction of unit cell parameters could be found, as indicated by the low values of the target function at the end of the optimization (between 3 and 11, see **Table IV.S1** of the Appendix). However, the optimized LJ parameters of the Or atom type varied considerably, leading to values of the energy minimum ε ranging from 2.19 to 60.61 kJ mol^{-1} which were obviously unrealistic. Hence, the ether oxygen atoms, O1 and O5, were no longer treated separately from

the hydroxyl oxygen atoms.

When all oxygen atoms in the glucose residue were set to OA atom type, only 8 LJ parameters of 4 different atom types had to be optimized. Under such conditions, I could also find sets of the 8 LJ parameters that give reasonable agreement for unit cell parameters of three allomorphs, with a minimal value of the target function of 9.5 (**Table IV.S2** of the Appendix). However, some of these sets of parameters, for instance, the one excluding Iβ and the one excluding II, were not able to improve the agreement with the unit cell of the fourth allomorph, which was left out from the optimization process. Furthermore, the optimized set of LJ parameters for which the target function is minimal gives unrealistic optimal separation for the C1 and CH2 atom types, when compared to the original parameter. Thus I concluded that the optimization of the LJ parameters of Gr53a6FF, without modifying the partial charges on the atoms, could not give a robust parameter set for modeling crystalline cellulose.

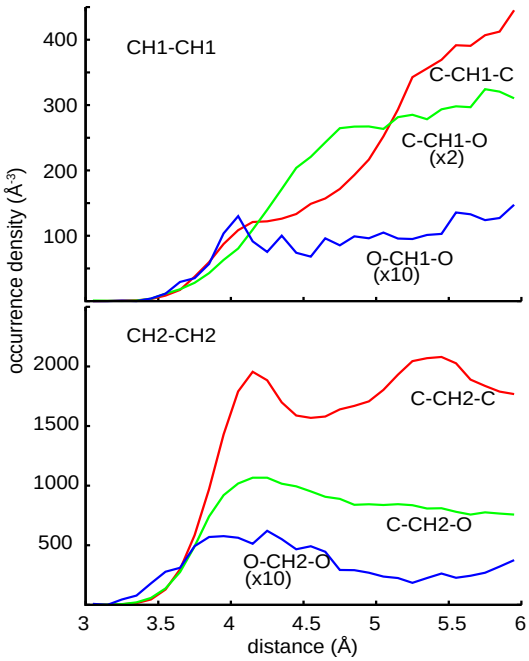

Figure IV.8. Occurrence density (occurrence divided by distance^2) of distance between non-bonded carbon atoms linked to different neighbor atom types in crystal structures compiled in Cambridge Structural Database.

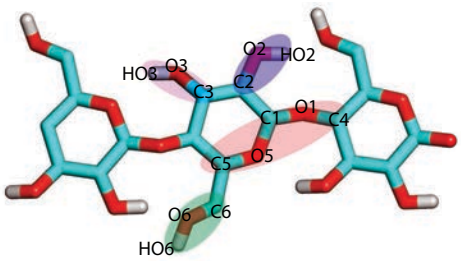

Figure IV.9. Molecular representation of a segment of a cellulose chain. Definition of the different groups in a glucose unit.

IV.5.5. Influence of the atomic partial charges on the monoclinic angle of cellulose I

The GROMOS force fields use the charge-groups concept in which the sum of total partial atomic charges within a group is zero. Four charge groups are defined fro a glucose residue, as shown in **Figure IV.9**. Based on the CHARMM C35 charges, I introduced perturbation by transferring a partial charge of ±0.1 e from one atom to another in the same group. In short, when 0.1 e is added to the charge on atom A, 0.1 e is simultaneously subtracted from the charge of atom B. MD simulation of the Iβ crystal was then performed considering the new set of charges, and the unit cell parameters at equilibrium were recorded. The cell parameters varied linearly with the atomic charges within this perturbation range. **Figure IV.10** shows the deviation between experimental and simulated unit cell parameters of Iβ caused by the charge transfer. The main effect from modifying the atomic charges of the atoms C1, C5, O1, C4 and O5 was to significantly vary the unit cell parameter γ. In contrast, the charges on C2, O2, C3, O3, C6 and O6 atoms had much less influence.

The analysis of the interatomic distances between atoms belonging to the two consecutive sheets in cellulose Iβ reveals that the following pairs are in close proximity (below 3.3 Å): O1—C2, C4—C6, C1—O6, O3—C1, C5—O1, O5—C6, C6—O2 and O2—C3. These short distances are shown in a model in **Figure IV.11**. Changing the charges on these atoms has a profound consequence on the equilibrium organization of two consecutive layers of cellulose chains. Increasing the charges on the O1, C1, C2, O2, O3, C5, C6 and O6 atoms increases the electrostatic attraction between the following pairs that are oppositely charged (O1—C2, C1—O6, O3—C1, C5—O1 and C6—O2) and, consequently, causes a left shift of the layer in front with respect to the one behind. Similarly, increasing the charges on the O2, C3, C4, O5

and C6 increases the electrostatic attraction between the oppositely charged pairs C6—O2, C6—O5 and O2—C3, and also increases the electrostatic repulsion between the C4—C6 pair. Both causes the layer in front to shift to the right. This shift of a cellulose layer with respect to the consecutive layer has a direct impact on the unit cell parameter γ. This explains why the partial charges have a significant impact on the monoclinic angle of the simulated cellulose Iβ.

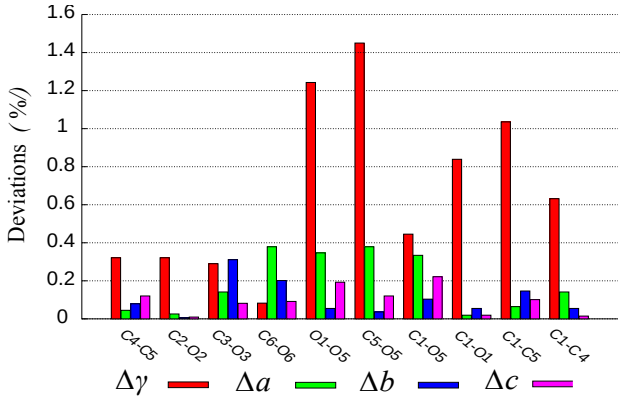

Figure IV.10. The variation of the unit cell parameters *a*, *b*, *c* and the γ angle of the Iβ crystal model of cellulose as a function of the modification of the atomic partial charges of atom pairs.

To compare the performance of different sets of partial charges proposed so far, I combined the atomic partial charges from CHARMM C35, GLYCAM 06, Gr56A$_{carbo}$FF and GROMOS 87 with all other parameters of Gr53a6FF, respectively, and performed 1 ns molecular dynamics simulations for each allomorph. For the partial atomic charges of the all atom force fields, CHARMM and GLYCAM, the charges of aliphatic hydrogen atoms were added to the corresponding united atom in GROMOS. In addition, the charges of GLYCAM have been slightly modified to satisfy the requirement that the sum of the charges within a group should be zero. The resulting equilibrium simulated unit cell parameters were summarized in **Table IV.S3** of the Appendix. The deviation of the unit cell γ with respect to the experimental value is given in **Figure IV.12**. The simulated monoclinic angles were smaller than the experimental value for cellulose Iβ and II with any set of partial charges, but the deviation was most important with Gr53a6 charge. CHARMM C35, GLYCAM 06 and Gr56A$_{carbo}$ gave the similar tendency of improving the agreement for cellulose Iβ without deteriorating other allomorphs.

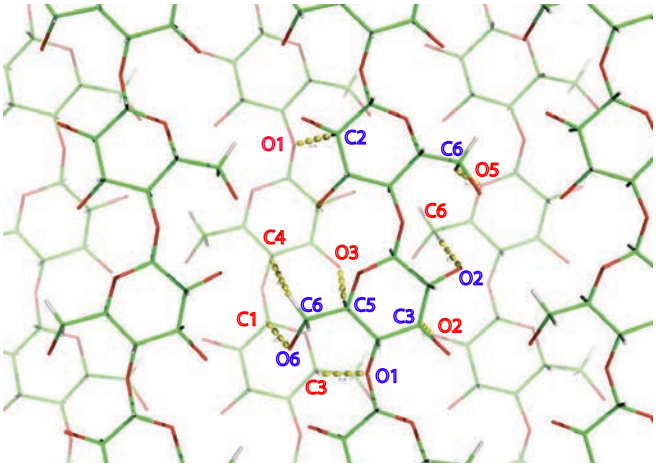

Figure IV.11. Graphical representation of two neighbor layers of cellulose Iβ viewed perpendicular to the chain axis. The three chains displayed with bright colors and blue atom names represent the layer in front; the back layer also contains three chains displayed with pale colors and red atom names. The dotted yellow lines indicate pairs of atoms distant by less than 3.3 Å.

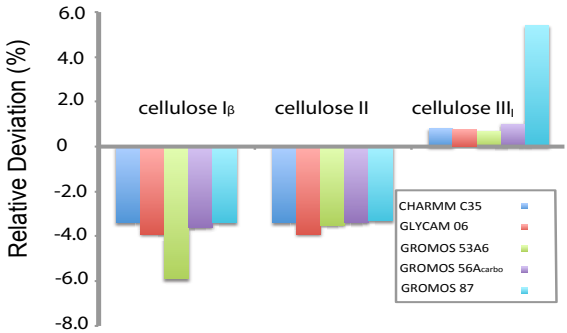

Figure IV.12. Deviations in percentage between the predicted (using different sets of charges) and the experimental monoclinic angles.

I looked further into the detail of the origins of the three sets of partial charges. Compared to the other force fields, the general characteristic of the atomic partial charges for glucose in GROMOS 87 is their smaller magnitudes (**Table IV.5**). Unfortunately, I could not find in the literature how those charges were determined. All other sets of charges are essentially based on the same type of QM calculation in vacuum and charge fitting. There is thus a consensus on the charge distribution on the C, O and H atoms making the alcohol groups for the other force fields. The difference between CHARMM and GLYCAM lies in the fact that the former has introduced an empirical correction for the condensed phase and thus has smaller charge amplitude. In addition, the magnitude of the partial charges may be varied with GLYCAM by

using a so-called electrostatic scale factor when calculating the potential energy of 1-4 pairs. Gr56A$_{carbo}$FF, which is based on the bond increment method, has a very strong polarity at C1-O1, with the carbon atom C1 bearing a larger positive charge than the hydroxyl hydrogen atoms. Although it reproduces the monoclinic angle of cellulose Iβ relatively well, the high atomic charge on aliphatic carbon seemed to be unrealistic. Based on the above consideration and since I could not find a better approach for the assignment of partial charges in condensed phase, the set of atomic charges of CHARMM C35 seemed the most adequate to model crystalline cellulose.

IV.5.6. Optimization of Lennard-Jones parameters of CH1

So far, I have studied separately the influence of the two non-bonded contributions (LJ and charges) on the crystal cell dimensions; the goal in this section was to consider them simultaneously. I used the charges on the atoms of CHARMM C35. No attempt was made to improve them. To reduce the number of variables, I decided to concentrate our efforts only on the optimization of the LJ parameters of the atom type CH1. As mentioned in the introduction, it is the most represented in glucose and its parameters strongly vary among the different generations of GROMOS.

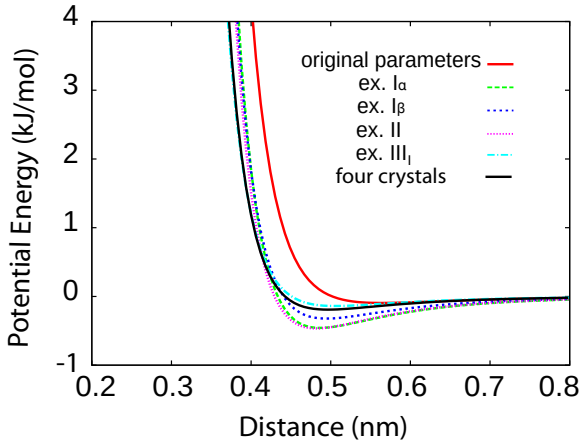

Figure IV.13. Energy profiles of the CH1 - CH1 interaction obtained by optimization of the two LJ parameters of CH1 against different set of targets. Black dotted line: optimization against four allomorphs, Green: optimization excluding Iα structure, Blue: optimization excluding Iβ structure; pink: excluding cellulose II structure; cyan: excluding cellulose III$_I$.

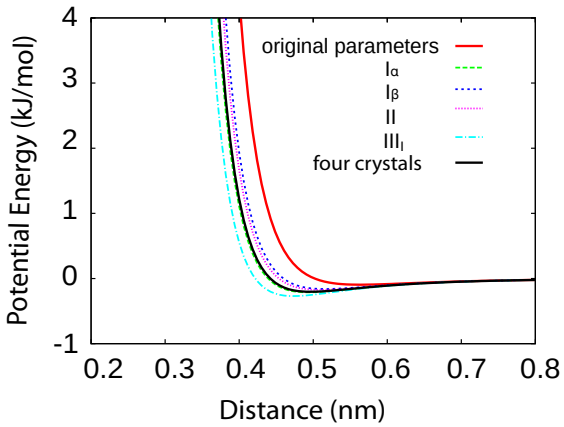

Figure IV.14. Energy profiles of the CH1-CH1 interaction obtained by optimization of the repulsive LJ parameter of CH1 against different set of targets. Each cellulose crystals have been considered independently. Red: original parameters; green: parameters obtained considering only the Iβ cellulose crystal; blue: cellulose Iα; pink: cellulose II; light blue: cellulose III_I. The curve obtained considering all the crystals is also included in black.

Two series of calculations were performed. In the first, the two LJ parameters for CH1 were optimized following the procedure already described. However, considering that the attractive term is generally derived from atomic polarizability (Oostenbrink *et al.* 2004), and that the repulsive term has no theoretical basis, in the second set of calculations I only optimized the repulsive parameter and the attractive term was kept constant at its original value of 0.0779 (kJ mol^{-1} $nm^6)^{1/2}$. This last series of calculations involved five optimizations: four of them considered a single allomorph and the remaining one used all the four crystals.

Until this point, I had used all six unit-cell parameters of each allomorph in the calculation of the target function. This potentially makes 24 target values, ten times more than the number of parameters to be optimized. However, I realized that the angles α and β of the monoclinic allomorphs rarely departed from 90˚ (the maximal change observed was 0.1%) due to symmetry. Furthermore, I expected that the *c*-axis was not very sensitive to the LJ parameters as the structure in *c* direction is mostly maintained by covalent bonds. Since I was taking the sum of the square of the difference from experimental values expressed in percentages, the departure of the *c*-parameter can over-influence optimization procedure. Thus I excluded the right angles and *c* parameters from the calculation of the target function during the optimization of the LJ parameters for CH1.

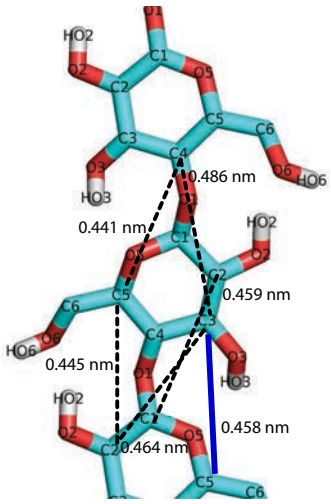

Figure IV.15. Representation of some short distances of 1-5 atom pairs and one 1-6 pair between two adjacent glucose rings in a cellulose chain. Atom names are labeled in black fonts. The 1-5 atoms pairs are shown in black dotted lines and the corresponding distances are given. The distance of 1-6 atom pairs are presented by blue solid line.

The results using a subset of 14 unit cell parameters (excluding *c* and right angles) are summarized in **Tables IV.S6** and **IV.S7** of the Appendix. **Figures IV.13** and **IV.14** give the predicted energy-distance profiles of the CH1 to CH1 LJ interaction using the optimized parameters obtained by different strategies and target structures. **Figure IV.13** exhibits the energy profiles by using both optimized $C12_{ij}$ and $C6_{ij}$, whereas **Figure IV.14** used the optimized $C12_{ij}$ and fixed $C6_{ij}$. With all sets of target allomorphs, the target function reached similar values; the optimized parameters were all in a reasonable range and the corresponding LJ energy profiles showed similar tendencies of shorter distances and deeper minima. Curiously, all the unit cell parameters, including the *c*-parameter, were better reproduced than when incorporated in the target function.

The constraints imposed by the *c* parameters, which are not directly related to the LJ parameters, probably make the target function landscape rougher. The simplex algorithm could probably not reach the global minimum of this rough function.

Finally, the values I retained, given in **Table IV.4**, were those obtained by optimizations using the four crystals simultaneously.

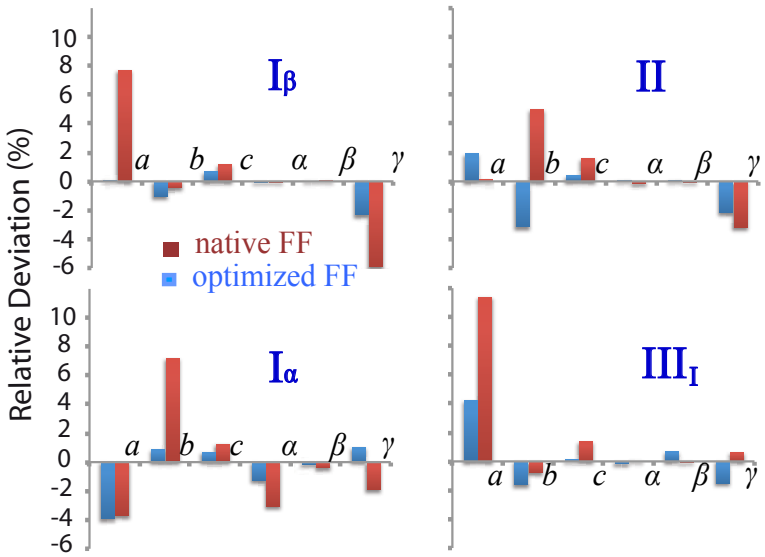

Figure IV.16. Deviation between the predicted unit cell parameters and the experimental ones, using the original Gr53a6FF force field and the optimized force field for the four cellulose allomorphs.

IV.5.7. Performance of the proposed non-bonded parameters

Final MD of 10 ns of the four crystal structures was performed using the GROMOS 53a6 force field framework with atomic charges imported from CHARMM C35 and the optimized repulsive LJ parameter for CH1 (line labeled "optimization I" in **Table IV.4**). The comparison of the crystal cell parameters with respect to experimental values is shown in **Figure IV.16**. **Table IV.S7e** shows the simulated unit cell parameters of all crystals at equilibrium together with the experimental ones and those obtained using the native force field. Clearly, the simulated data were in better agreement with experimental data just by one optimized LJ parameter and by replacement of the set of charges.

The latest GROMOS force field, Gr56A$_{carbo}$FF, which also failed at predicting the expected value of the unit cell parameter *a* for cellulose Iβ (**Table IV.3**), was thus applied with the strategy previously described. The resulting optimized repulsive parameter of CH1r was 0.006772 (kJ mol^{-1} nm^{12})$^{1/2}$, which is very close to that obtained for Gr53a6FF, 0.006972 (kJ mol^{-1} nm^{12})$^{1/2}$. The simulated data is also in better agreement with the experimental one (**Table IV.S8c**).

IV.6. Combination and validation of the optimized force field parameters

Generally, the optimization of the force field is not the simple combination of parameters individually optimized by fitting to different experimental properties. Since the parameters can be highly correlated, a given property of a system that is well reproduced with a set of force field parameters can be poorly simulated when another set of optimized parameters is used. For this reason, a set of new force field parameters (atomic charges and LJ parameters) of oxygen atoms optimized for Gr53a6FF by fitting some neat liquid properties, named Gr53a6$_{oxy}$FF, has recently been proposed (Horta *et al.* 2011). However, it has not been combined with Gr56A$_{carbo}$FF yet.

I have optimized the LJ parameters of CH1 against the experimental unit cell parameters for both Gr53a6FF and Gr56AFF, and suggested to exchange the atomic partial charges from CHARMM C35, while the torsional parameter of the hydroxymethyl group was independently optimized to reproduce the reversibility of the phase transition. As described in the following, I combined these three sets of force field parameters to obtain both precise unit cell dimensions and reversibility of phase transition of Iβ.

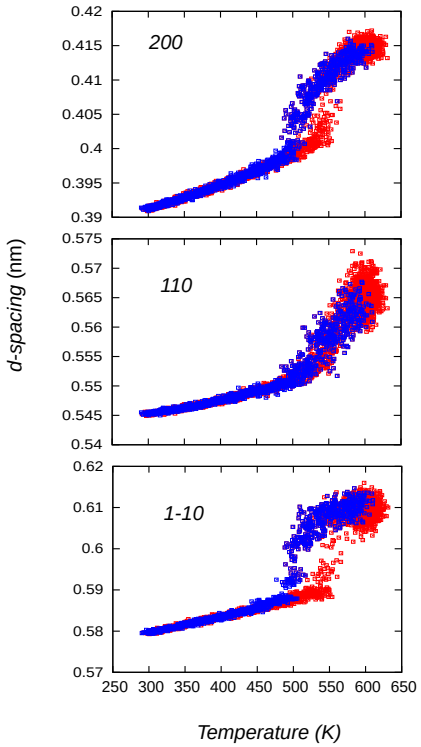

Figure IV.17. Variation of d_{110}, $d_{1\text{-}10}$, d_{200} spacings in cellulose Iβ with temperature upon heating (red) and cooling (blue) using revised Gr56A$_{carbo}$FF.

When the atomic charges of CHARMM, optimized LJ parameters of CH1 and the optimized torsional parameter of K_1 = 4.5 kJ/mol were implemented into Gr53a6FF, I could still obtain the improved unit cell parameters of cellulose Iβ at room temperature, indicating that the revision of K_1 affected the simulated unit cell parameters very little. However, the structure at high temperature still remained parallel-up (a monoclinic angle of 100°), which differs from that modeled by using native the force field that gives a monoclinic angle of 72°. I did not observe the reversible phase transition of cellulose Iβ with a cooling rate of 1.0 ns/K. The cooled structure has a higher energy than the initial structure, which suggests that the transition would be reversible with a slower cooling rate.

When the optimized torsional parameters K_1 = 1 kJ/mol and K_3 = 6 kJ/mol were used, the reversible transition was observed with a cooling rate of 1.0 ns/K. However, the transition occurred at the temperature over 600 K, which is 100 K higher than the experimental one. Further optimization on the non-bonded interaction, such as atomic partial charges and the LJ parameters, should be carried out when combining these optimized parameters. This result revealed that in MD, the reversibility of the phase transition is not only dominated by the energy profile of the hydroxymethyl group but also associated with the non-bonded force field parameters.

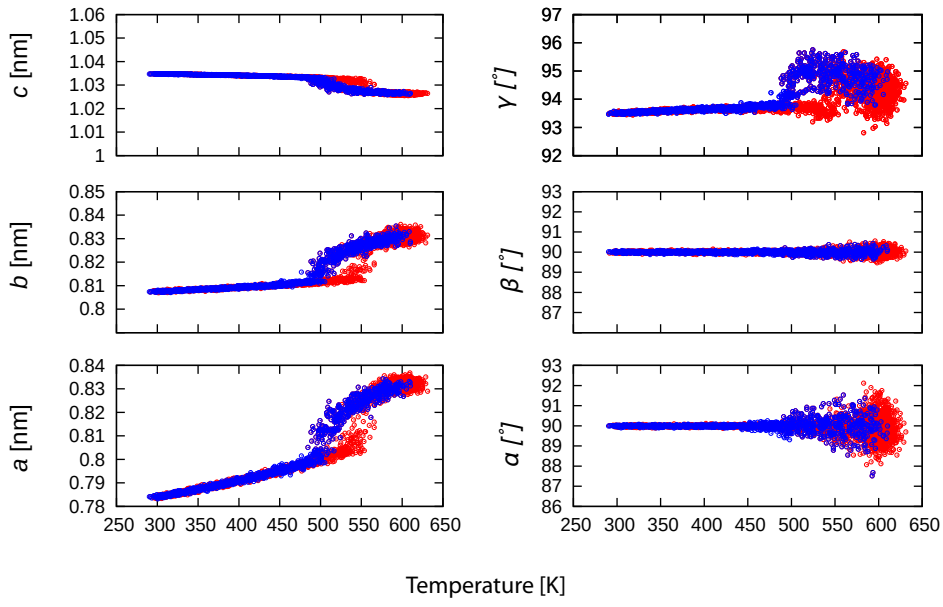

Figure IV.18. Variation of the unit cell parameters of cellulose Iβ with temperature upon heating (red) and cooling (blue) using revised Gr56A$_{carbo}$FF.

The optimized LJ parameter and CHARMM charges have also been imported into Gr56$_{carbo}$FF and the improved unit cell parameters was obtained at 300 K. However, the unit cell changed to a triclinic type at high temperature and we did not observe a reversible phase transition. Since the CHARMM charges did not significantly improve the simulated monoclinic angle, we decided to use the atomic charges of the original Gr56A$_{carbo}$FF, which means that only the LJ parameter of CH1 was revised in the optimized Gr56A$_{carbo}$FF compared to the native Gr56A$_{carbo}$FF. Surprisingly, the high temperature structure of cellulose Iβ has a monoclinic unit cell similar to the experimental ones at both room and high temperatures. Furthermore, a reversible transition was observed, with a transition temperature between 500 and 550 K, which is close to the experimental value, as shown in **Figure IV.18.**

Meanwhile, the simulated variations of *d*-spacing with temperature are very close to the experimental values, the deviation from experiment being less than 2% (**Figure IV.17**). Our results better reproduce the thermal response of cellulose than any other different carbohydrate force field (Matthew *et al.* 2012; Zhang *et al.* 2011; Bergenstrahle *et al.* 2007). Since it correctly reproduced room and high-temperature structures, phase transition reversibility and transition temperature of cellulose Iβ, the revised Gr56A$_{carbo}$FF has been mainly used for the simulations presented in the following chapter.

I have tried to combine the recently released Gr53A6$_{oxy}$FF parameters with those of revised Gr56A$_{carbo}$FF. The simulated room temperature structure of Iα and Iβ did not change when the new oxygen parameters were imported. However, the ability to reproduce the reversibility of phase transition was lost, indicating that the non-bonded force field parameters are highly correlated and the accuracy of them plays an important role to reproduce the thermodynamic properties of carbohydrates.

IV.7. Conclusion

This study has demonstrated that the absence of reversibility of the phase transition of native cellulose in simulations was directly related to one key torsional parameter in Gr53a6FF. The GROMOS force field could roughly reproduce the experimentally observed reversible phase

transition of native cellulose only when a set of alternative parameters was chosen for the torsion energy of the exocyclic hydroxymethyl groups. In addition, this study has demonstrated that the GROMOS force field could significantly be improved to model crystalline cellulose by modifying the Lennard-Jones parameter for atom type CH1 and increasing the partial charge distribution around the glycosidic linkage. The charge distribution has a high correlation with the monoclinic angles and the repulsion term of Lennard-Jones parameters has a direct influence on the spacing between glycopyranose planes in the simulated crystals. The optimized parameters (torsional parameters, atomic partial charges and Lennard-Jones parameters) were both combined into Gr53a6FF and Gr56A$_{carbo}$FF, yielding two optimized force fields, which have been used in the studies whose results will be presented in **Chapter V**.

Chapter V – Exploratory studies

V.1. Thermodynamics and molecular explanation of reversible temperature-induced phase transition of cellulose I

V.1.1. Introduction

The purpose of this section is to propose a thermodynamic explanation and molecular insights for the phase transition of native cellulose occurring at high temperature. In the previous chapters, we analyzed the variation of puckering amplitude and torsional parameters upon heating. At the transition temperature, we did not observe any discontinuity of the puckering amplitude (**Figure III.13**) and the distributions of the ϕ, ψ and τ conformational parameters (**Figure III.14**) by using original or revised parameters (**Figures S1, S2** and **S3** in Appendix). Thus we excluded these parameters from the analysis. In contrast, the ω, τ2, τ3 and τ6 torsion angles considerably varied (**Figures III.12, III.15** and **S4**).

Since the temperature-induced allomorphic conversion is reversible, it can be considered as a thermodynamic transition. First, we will propose a thermodynamic explanation from the simulation results by varying a force field parameter to modify the enthalpy of the high temperature phase. Then, we will give a molecular description of the free energy based on the analysis of the molecular dynamics trajectory.

V.1.2. The molecular origin of phase transition

A system exploring the accessible conformational space at constant pressure should minimize its Gibbs free energy:

$$G = H - TS$$

where H is the enthalpy of the system corresponding to the potential energy and S is the entropy. The phase transition should occur when

$$G_{tg} = G_{gg/gt}$$

$$T = \frac{H_{tg} - H_{gg/gt}}{S_{tg} - S_{gg/tg}}$$

K_1 directly adds to $H_{tg} - H_{gg/gt}$ decreasing the phase transition temperature. **Figure IV.1** show that the transition temperature varies almost linearly with K_1, which suggests that $S_{tg} - S_{gg/gt}$ is constant and independent from K_1. $H_{tg} - H_{gg/gt}$ is negative and thus $S_{tg} - S_{gg/gt}$ is also negative, which means that there should be an entropy gain by adopting the *gg/gt* conformations instead of the *tg* conformation. This entropy gain does not depend much on the local force-field parameters we modified since a linear tendency of phase transition temperature was observed (**Figure V.1.1**). From the slope in **Figure V.1.1**, the entropy of transition to high-temperature phase can be roughly estimated as 26 $JK^{-1}mol^{-1}$. The slope was almost the same with $K_3 = 6$ or 9.5, but the hysteresis was smaller with $K_3 = 6$.

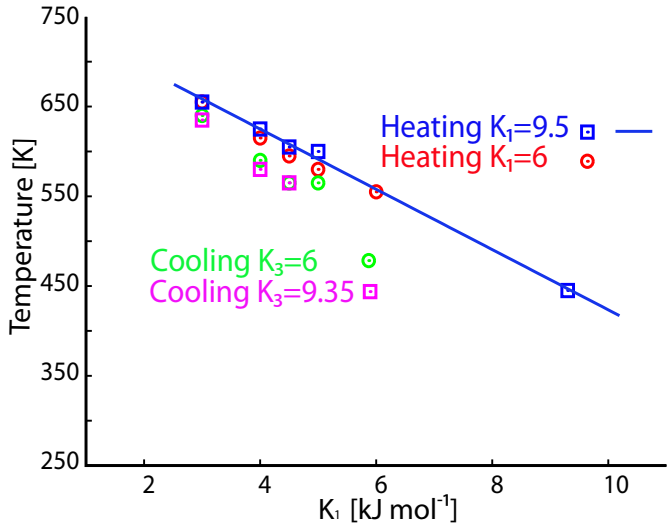

Figure V.1.1. Variations of the transition temperature as a function of K_1.

The torsional entropies *S* can be defined as follows (Li and Brüschweiler 2009):

$$S = -k_B \int_0^{2\pi} P(\chi) \ln\left(P(\chi)\right) d\chi$$

where k_B is the Boltzman constant, $P(\chi)$ is the probability to find the torsional angle at χ. In practice this was done using the histogram with 1 degree step extracted from the trajectory by

the following summation:

$$S = \sum_{0}^{360} P(\chi) \ln\left(P(\chi)\right)$$

The entropy of τ2, τ3, τ6, φ, ψ and ω as a function of temperature is shown in **Figure V.1.2**. The entropy gains of τ2, τ3, τ6 and ω during the transition were 8, 4, 8 and 6 $JK^{-1}mol^{-1}$, respectively. The sum of these four components corresponded to the expected entropy gain from the slope of **Figure V.1.2**.

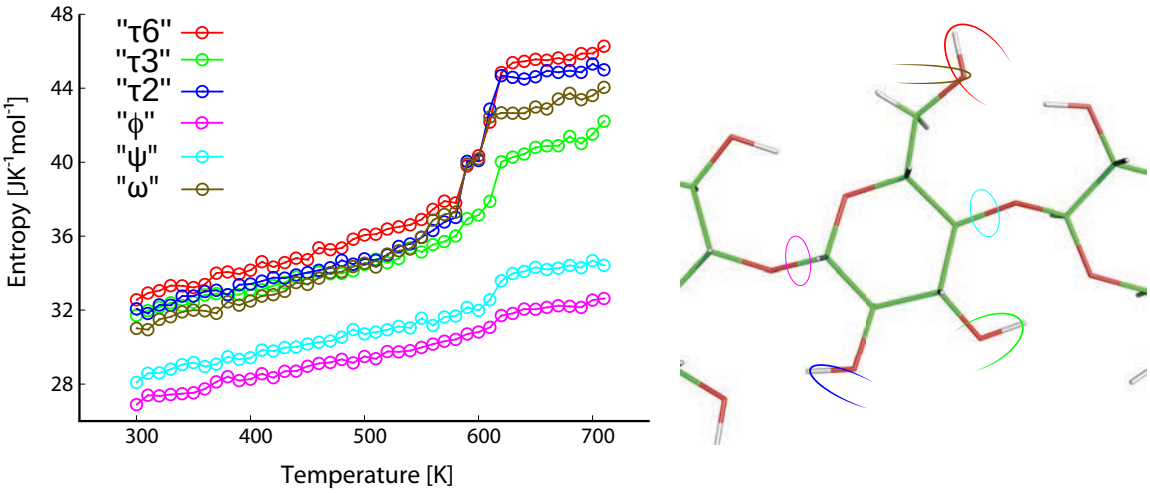

Figure V.1.2. Entropies of dihedral angles as a function of temperature with K_1 = 4.5 kJ/mol.

V.1.3. Transition frequency of the hydroxymethyl group

The transition frequency among different staggered positions of the hydroxymethyl groups increased almost exponentially with temperature in the 300–600 K temperature range (**Figure V.1.3**). Even at room temperature, the frequency of the order of one transition every 1–10 ns is still faster than the NMR timescale. The conformation is predominantly *tg* up to the transition temperature so the residence time at conformation different from *tg* is even shorter. This dynamic picture is different from the frozen micro-domain structure proposed earlier based on shorter simulation time of 1 ns using the PCFF force field (Mazeau 2005).

The reasonable number of transitions during the simulation is another indication that the simulation is reflecting thermodynamic aspects of the system, which governs the reversible phase transition. In the real system, the cellulose chains are closer and the energy barrier

between different staggered positions might be different depending on the force field we used. However, we observed a range of frequency over three orders of magnitudes and temperature of 300 K, and a few kJ difference in energy profile would not modify this general behavior.

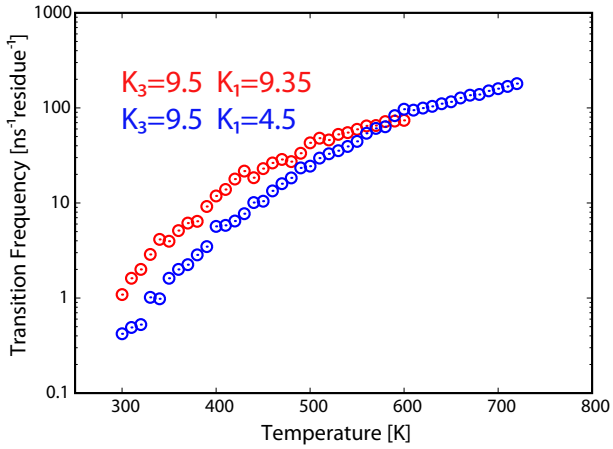

Figure V.1.3. Transition frequency of ω as a function of temperature.

V.1.4. Conclusions

By chosing a proper range of values for parameters describing the torsion energy of the exocyclic hydroxyl groups in Gr53a6FF and Gr56A$_{carbo}$FF, we could reproduce the reversible phase transition of native cellulose experimentally observed. This means that the thermal transition phenomenon can be partly decoupled from the packing of cellulose chain. The ring conformation and glycosidic torsion angle globally did not change as a function of temperature. Thus the thermal behavior of cellulose crystals can be understood as a phase transition phenomena of a rigid backbone with mobile side groups that induce the transition. A systematic variation of the torsion energy profile allowed us to understand the phase transition as the gain in entropy stabilization at high temperature. This means that the cellulose behavior at timescale longer than picoseconds can be probably described in a fairly large temperature range using rigid backbone with one flexible group, drastically reducing the number of effective parameters. The torsion energy profile of the exocyclic hydroxymethyl group is the key aspect determining the phase behavior of the cellulose crystals.

V.2. The effect of hydrogen bonding on the stability of cellulose I

V.2.1. Introduction

The hydrogen atoms at O2 and O6 in cellulose, namely HO2 and HO6, were found to be disordered, having at least two possible positions in Iα and Iβ allomorphs (Nishiyama *et al.* 2002, 2003, 2008). Two hydrogen-bonding (HB) network patterns (A and B) were proposed based on neutron fiber diffraction refinement. The refined occupancy of pattern A in the Iβ allomorph was 70%-80%, whereas it was about 55% in Iα. It was not sure whether these two HB patterns could interconvert dynamically or if they statistically occurred in different regions of a crystalline microfibril (Nishiyama *et al.* 2008). QM calculation and MD simulation based on empirical force field methods (Nishiyama *et al.* 2008) suggested that pattern A was energetically favorable and probably occurred in the crystalline core region, while pattern B was not stable and either was an spatial average of many different hydrogen bond patterns (Mazeau 2005) or occurred in the surface and in the defects of the crystalline regions (Nishiyama *et al.* 2008).

Most of the MD simulation studies based on force field methods were performed with allomorph Iβ using HB pattern A. Only a few studies were carried out on the structure having pattern B (Mazeau 2005; Matthew *et al.* 2012). In addition, MD studies on Iα with HB pattern B, accounting for about 45% of the total hydrogen bonds, have not been reported, to the best of our knowledge. In this chapter, we have carried out MD simulations to study the structural stability of allomorphs Iα and Iβ with two HB patterns by using the infinite approach and the revised Gr56A$_{carbo}$FF.

V2.2. Computational details

V.2.2.1. Model construction

The models are identical to those used in **Chapter III**. They contain 64 chains for Iα and 36 chains for Iβ. Each chain contains 8 glucose residues. For both Iα and Iβ, three supercells were constructed: one is with pattern A (called model 1), the second with pattern B (called model 2),

and the third is a mix of patterns A and B (called model 3), as shown in **Figures V.2.1**, **V.2.2** (top) and **V.3.3** (top). Since only one position for HO2 in the origin chain was observed experimentally, pattern A was attributed to the origin chain and pattern B to the center ones in models 3.

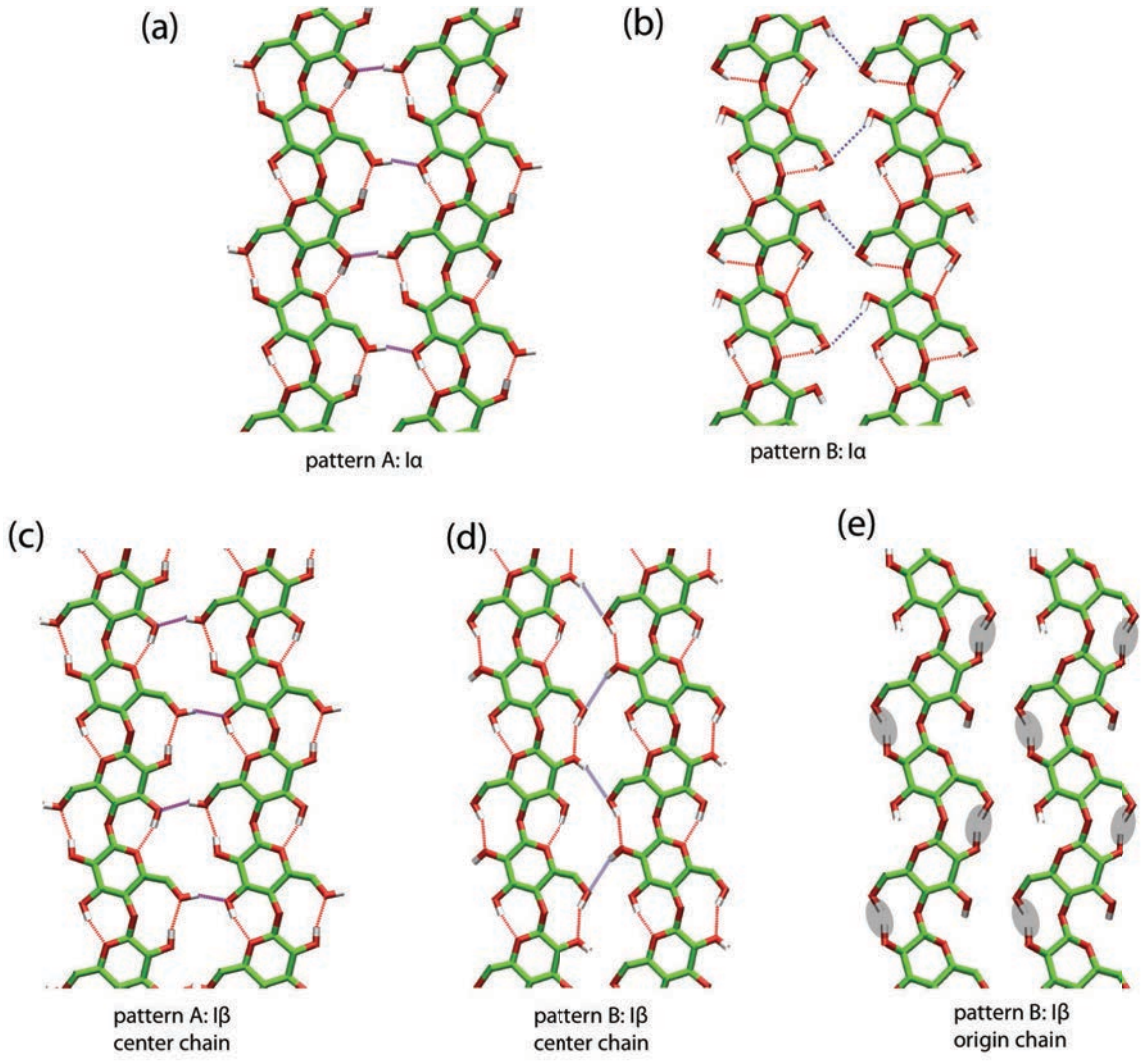

Figure V.2.1. The HB patterns in cellulose Iα and Iβ: a) pattern A in Iα; b) the HB pattern B in Iα; c) pattern A in the center chain of Iβ; d) pattern B in the center chain of Iβ; e) pattern B in the origin chain of Iβ. The HB pattern A in the origin chain of Iβ is similar to that in center chain and was thus not presented. The hydrogen bonds are indicated with red dotted lines. The gray shadow in e indicated the two hydrogen atoms (HO2 and HO3) in very short distance (0.094 nm) when origin chains of Iβ adopts with HB pattern B.

V.2.2.2. System setup

The revised Gr56A$_{carbo}$FF is only different in the repulsive Lennard-Jones parameters of CH1,

comparing to the native Gr56A$_{carbo}$FF, which has been described in **Chapter IV**. In order to compare our results with the structural details observed experimentally at 15 K, the MD simulation was first performed from 0 to 300 K for all models. Subsequently, an annealing process between 300 and 600 K was performed to test the structural stability. Other system setups are all identical to those described in **§ III.2**.

V.2.2.3. Analysis

Hydrogen bonds are detected from geometric criteria: the distance between donor and acceptor with a cutoff of 0.3 nm and the donor-hydrogen-acceptor angle with a cutoff of 140°.
The unit cell parameters at equilibrium were calculated from the last 5 ns of MD simulation.

V.2.3. Crystal structures of model 1 of Iα and Iβ

By using the infinite approach, the Iα and Iβ structures with pattern A were both stable at 15 and 300 K. Some important parameters used to describe the crystal structure of cellulose, such as the conformation of the hydroxymethyl group, unit cell parameters and glycosidic linkage angles, are all in good agreement with experimental data, since the revised Gr56A$_{carbo}$FF is optimized against these parameters.

V.2.4. Crystal structures of models 2 and 3 of Iβ

In the experimental crystal structure of Iβ, the HO2 hydrogen atom in the origin chain was found to have a unique position contrary to the two positions in the center chain. The HO6 hydrogen atom in the origin chain still had two possible positions which were similar to those in the center chain. When O6---HO6 in the origin chain pointed to O3 from the neighbor origin chain, an inter-chain hydrogen bond could form. Together with the intra-chain HB, it formed the HB pattern A. However, when the O6---HO6 pointed to O2, the HO6 and HO2 were

separated by 0.094 nm (**Figure V.2.1e**) and thus caused high electrostatic repulsion, since both of them have the atomic partial charge of 0.41.

After energy minimization, we found that in the model 2 of Iβ, the O6---HO6 did not change its orientation to form the HB pattern A. Instead, the O2---HO2 immediately oriented to toward the O6 of the neighbor origin chain. Consequently, like in the center chain, the HB pattern B was formed in the layer of origin chain.

Figure V.2.2 (middle) shows selected snapshot (last frame at each temperature) of model 2 of Iβ during the MD simulation. The hydroxymethyl group changed its initial *tg* conformation to a mixture of *tg*, *gt* and *gg*, randomly distributed in the crystal. Even at very low temperature (15 K), the unit cell parameters (**Table V.2.1**) became very different with those from pattern A and also different from experimental data, indicating that model 2 was not as stable as model 1.

Figure V.2.2 (right) shows the conformation of the hydroxymethyl groups in the model 3 of Iβ during MD, which indicates a mixture of HB pattern A in the origin chain and HB pattern B in the center chain. After 10 ns at 300 K, the *tg* conformation in the center chain changed to a mix of *tg*, *gg* and *gt*, but the origin chains still kept their initial *tg* conformation and all corresponded to HB pattern A. As shown in **Table V.2.1**, the unit cell parameters departed from the experimental data.

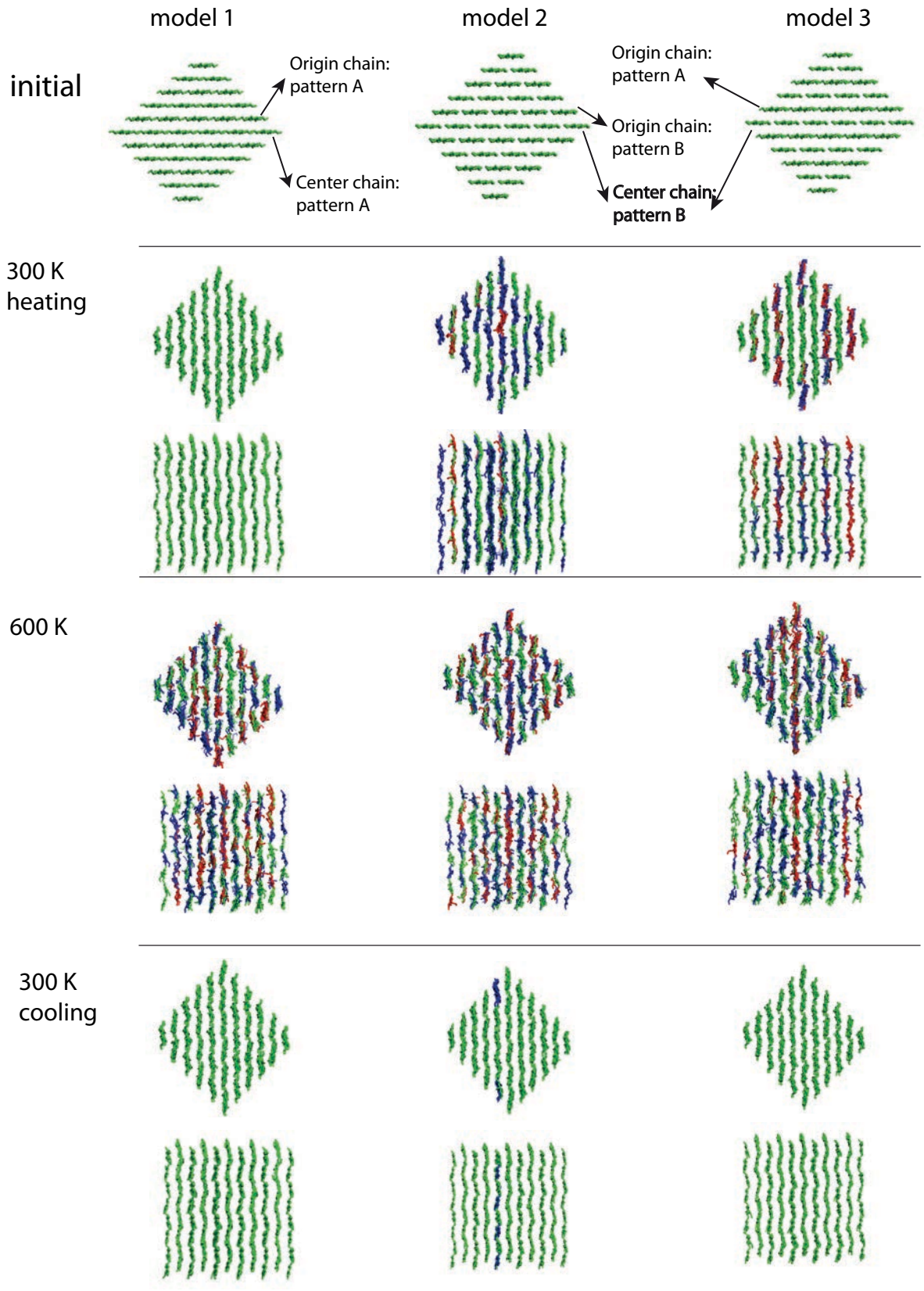

Figure V.2.2. Snapshots of the three models of cellulose Iβ simulated at different temperatures. The conformations of hydroxymethyl groups are colored in green (*tg)*, blue (*gt)* and red (*gg*).

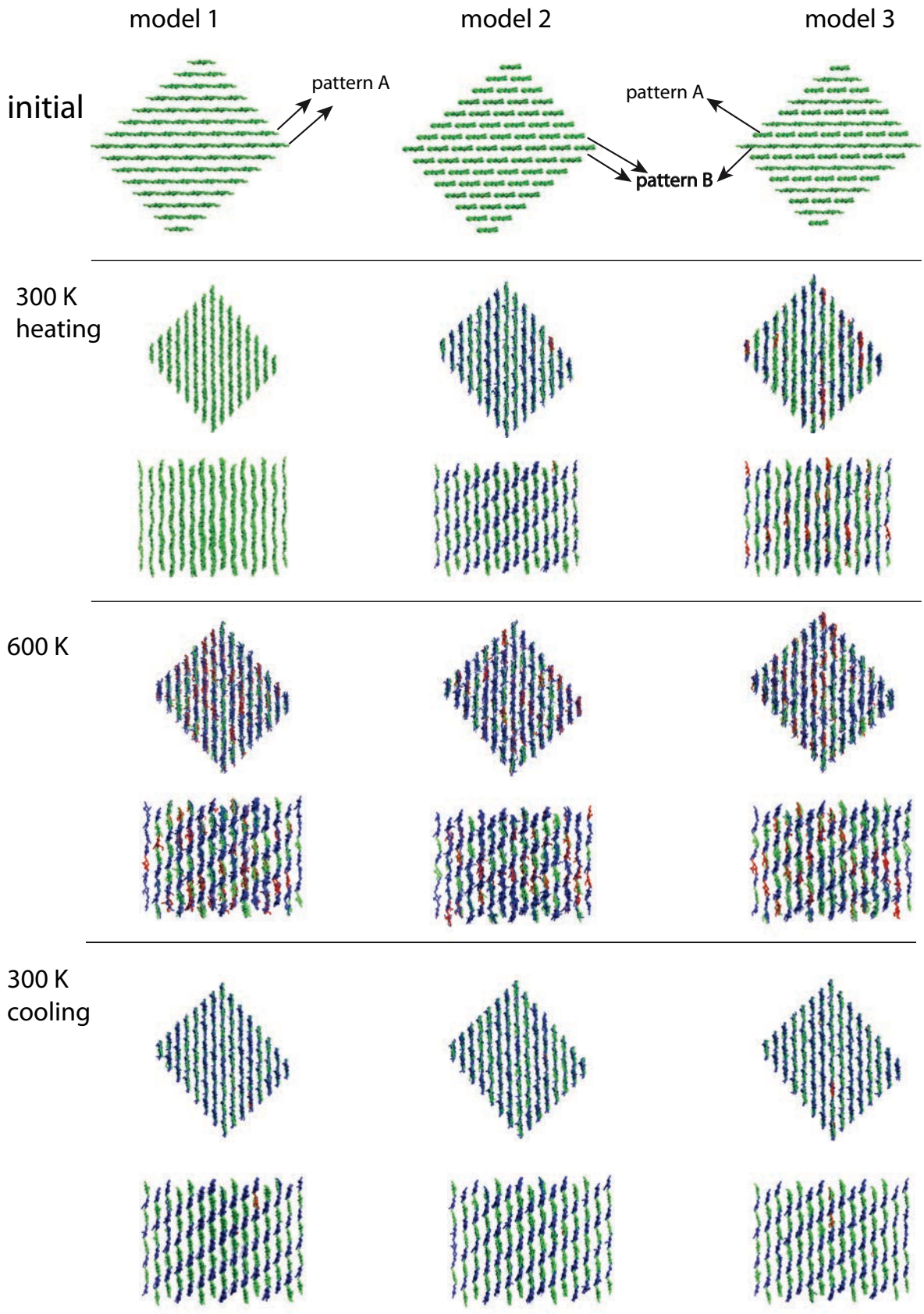

Figure V.2.3. Snapshots of the three models of cellulose Iα at different temperatures. The conformations of hydroxymethyl groups are colored in green (*tg)*, blue (*gt)* and red (*gg*).

Table V.2.1. Unit cell parameters of the three models of Iα and Iβ at different temperatures, calculated by MD simulations from the last 5 ns of the trajectory file.

Iβ		***a* (nm)**	***b* (nm)**	***c* (nm)**	***α* (˚)**	***β* (˚)**	***γ* (˚)**
Heating	Model 1	0.7839(3)	0.8075(2)	1.0348(0)	89.99(3)	90.00(3)	93.50(2)
300 K	Model 2	0.7946(9)	0.8249(8)	1.0335(2)	89.83(10)	90.00(13)	94.80(31)
	Model 3	0.8024(15)	0.8038(4)	1.0301(2)	90.00(6)	90.00(5)	89.27(38)
600 K	Model 1	0.8319(15)	0.8310(15)	1.0264(3)	89.94(57)	90.00(15)	94.39(37)
	Model 2	0.8339(16)	0.8310(15)	1.0259(3)	90.04(60)	89.96(17)	94.11(38)
	Model 3	0.8280(15)	0.8283(13)	1.0263(3)	90.61(67)	89.90(17)	94.74(36)
Cooling	Model 1	0.7839(3)	0.8075(2)	1.0347(0)	89.99(3)	90.04(3)	93.50(2)
300 K	Model 2	0.7851(4)	0.8093(2)	1.0346(1)	90.18(4)	90.04(3)	93.73(2)
	Model 3	0.7851(3)	0.8076(2)	1.0339(1)	89.98(3)	89.99(3)	93.51(3)
Iα		***a* (nm)**	***b* (nm)**	***c* (nm)**	***α* (˚)**	***β* (˚)**	***γ* (˚)**
Heating	Model 1	0.9476(3)	0.8135(2)	1.0416(1)	89.87(3)	124.08(2)	95.10(3)
300 K	Model 2	0.9635(3)	0.8411(2)	1.0357(1)	82.21(3)	124.97(3)	102.10(5)
	Model 3	0.9593(10)	0.8339(4)	1.0369(5)	84.65(13)	124.96(6)	98.50(43)
600 K	Model 1	0.9930(20)	0.8495(7)	1.0292(2)	81.95(19)	124.54(13)	100.04(23)
	Model 2	1.0017(14)	0.8501(7)	1.0227(1)	81.36(20)	124.35(11)	101.20(21)
	Model 3	0.9927(19)	0.8495(7)	1.0292(2)	81.97(19)	124.54(13)	100.02(23)
Cooling	Model 1	0.9645(3)	0.8415(2)	1.0354(1)	82.12(3)	125.04(3)	101.99(5)
300 K	Model 2	0.9652(3)	0.8420(2)	1.0361(2)	82.07(3)	125.31(3)	102.30(3)
	Model 3	0.9637(3)	0.8412(2)	1.0355(0)	82.18(4)	125.02(2)	101.95(3)

V.2.5. Crystal structures of models 2 and 3 of Iα

Figure V.2.3 (middle) shows the snapshots of the three models of Iα at different temperatures. When we calculated the variation of the unit cell parameters from 0 to 300 K, model 2 did not show structural variation at 15 K, indicating that the structure remained stable. With the increase of the temperature, the structure suddenly changed at 34 K. The final frame at 300 K

shows that the structure contained only *tg* (50%) and *gt* (50%) conformations for the hydroxymethyl group. The *tg* and *gt* conformation appears regularly and interactively along the chain direction, which differs from the random distribution in the Iβ (models 2 and 1 at 300 K in **Figure V.2.2**). Furthermore, the fluctuation of the unit cell parameters is constantly evolving during the 10 ns of equilibration at 300 K.

Figure V.2.3 (right) shows the structure of model 3 of Iα during MD simulation, which is initially a mix of HB pattern A (50%) and pattern B (50%). In this structure, the occupancy of HB disorder is close to the reported experimental amount (~55% for A and ~45% for B). From 0 to 300 K, the hydroxymethyl groups in (110) planes that used to form HB pattern A, partly changed from the initial *tg* conformation to a mix of *tg*, *gt* and *gg*. Almost no *gg* conformation appeared in the layer where HB pattern B used to be. Instead, the regular alternation of *tg* and *gt* was observed.

V.2.6. Annealing

All models were heated to 600 K to examine the high temperature structure. The three models for both Iα and Iβ have similar high temperature structures which contain equivalent populations of *tg*, *gg*, and *gt* conformations (**Figure V.2.2** and **V.2.3**). The structures also exhibit similar unit cell parameters, as shown in **Table V.2.1**.

As previously described in **Chapter IV**, a reversible transition was observed for the Iβ allomorph with HB pattern A (model 1) during the heating-cooling process when the LJ-optimized Gr56A$_{carbo}$FF was used. Here, during cooling, we observed that the three models of Iβ all evolved to the same structure that only contained HB pattern A, which is almost identical to the initial structure of model 1. This indicates that the models that contained pattern B, which is not stable at 300 K, evolved to a structure that only contained pattern A (**Figure V.2.2**).

Figure V.2.4 shows the variation of the populations of the different conformations of the hydroxymethyl groups of the three models of Iβ, between 0 and 600 K. The rapid decrease of *tg* population in model 1 indicates that the transition started at 550 K with the revised Gr56A$_{carbo}$FF, which corresponds the experimental transition temperature of 500 K. It is where

started from 350 K, models 2 and 3 exhibit an increase of the *tg* population and a decrease of *gg* and *gt*, yielding only *tg* (almost 100%) appears at 490 K. On an other hand, other conformational parameters, such as unit cell parameters (**Table V.2.1**) and the conformation to the hydroxyl groups, all show that the structure of these three models became very similar, indicating that HB pattern A was dominant in all the models at 490 K. These results revealed that, in these simulation, HB pattern B, that deformed at room temperature, converted to HB pattern A at a temperature just below the phase transition temperature.

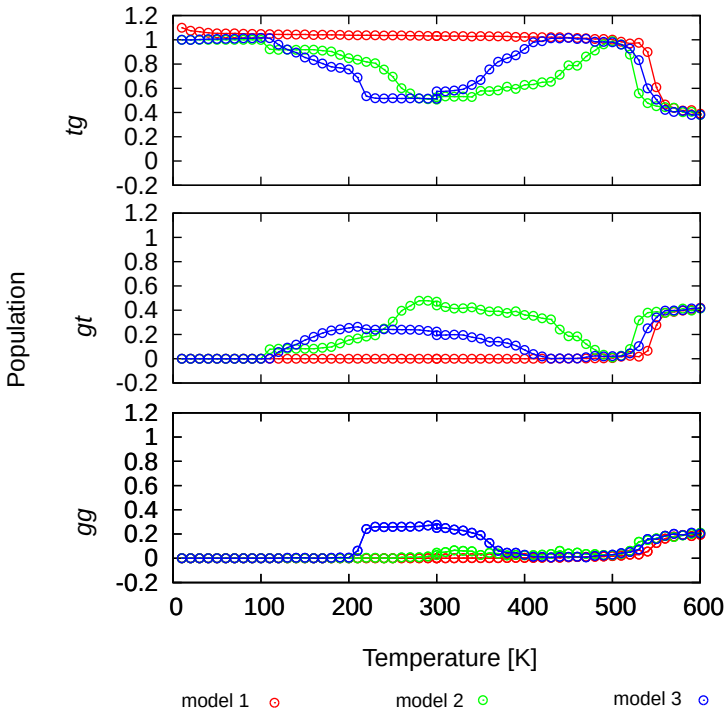

Figure V.2.4. Variation of the three-staggered conformation of hydroxymethyl group of the three model of Iβ allomorph as a function of temperature.

For I_α, we observed an irreversible transition, contrary to the reversible one of I_β. When temperature was decreased from 600 to 300 K, the three models irreversibly evolved to the same structure that was different from any starting model, but similar to the simulated structure of model 2 at 300 K, in which the *tg* and *gt* conformations alternated along the chain direction. Comparing the total energy, we found that this cooling structure at 300 K had a potential energy about 0.66 ± 0.58 kJ/mol per glucose lower than that of the structure of model 1 at 300 K (heating). The lower energy may be due to the formation of new HB networks in the cooling

structure, as shown in **Figure V.2.5.** In this HB network, every hydroxymethyl group that adopted a *tg* conformation participates to one intra-chain HB (O6-HO6---O2) and one inter-chain HB (O2-HO2---O6). This pattern seems to be similar to pattern B in Iβ. On the other hand, the other hydroxymethyl groups that adopted the *gt* conformation also participate to two hydrogen bonds. However, both of them are inter-chain HBs. In average, each glucose has 1.5 inter-chain HBs and 1.5 intra-chain ones, in contrast to 1 inter-chain HB and 2 intra-chain HBs in the experimental pattern A. The hydrogen bonds in the cooling structure are slightly stronger than in the experimental one due to the shorter hydrogen-donor distance (**Figure V.2.5**). The formation of this new HB network requires a relative translation of the neighbor chain in the (110) planes of Iα by about 0.12 nm along the chain direction.

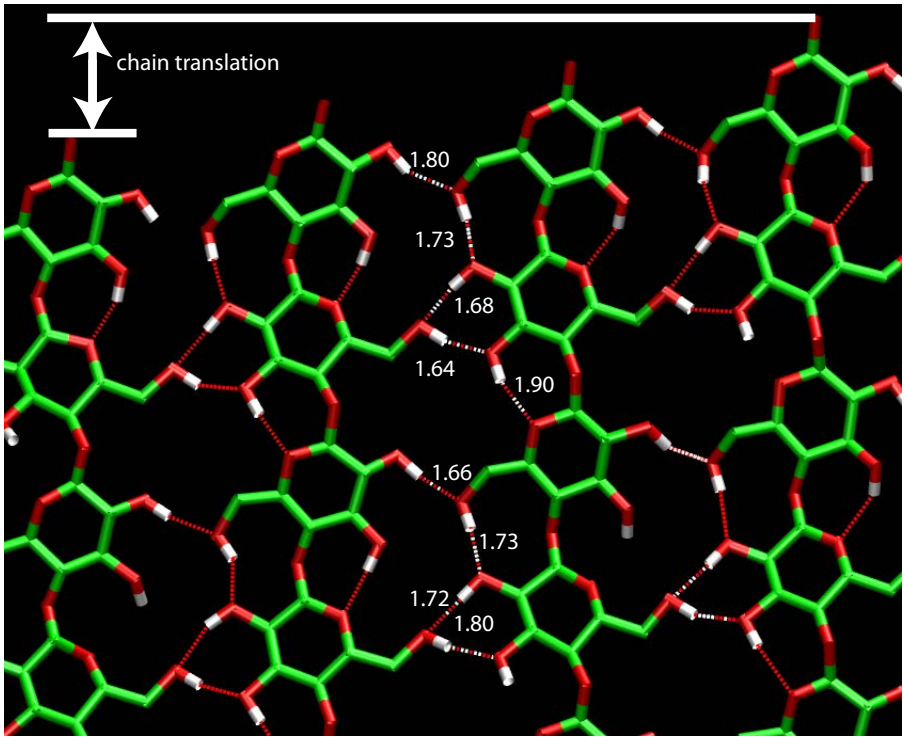

Figure V.2.5. Snapshot of the hydrogen-bonding network in the (110) plane in the cooling (300 K) structure of Iα. The distances between the hydrogen atoms and acceptors are indicated. Hydrogen bonds are indicated by red dotted lines.

V.2.7. Conclusion

The HB disorder, which was expected to be located at the surface or region of crystal defects in cellulose I, was studied by MD simulation using infinite model and revised Gr56A$_{carbo}$FF. The

calculation demonstrated that HB pattern B was not stable in the core region of both Iβ and Iα. The Iβ with pattern B could transform into pattern A at high temperature (490 K), which is just below the phase transition temperature. In addition, the HB disorder, which has a higher occupancy of pattern B (45%) in Iα than that in Iβ, is also studied. We found that with HB pattern A in Iα was stable at room temperature and irreversible through a thermal treatment, which is different to the reversible HB pattern A in Iβ. Another simulated structure in which *gt* and *tg* alternated regularly along the chain direction was energetically more stable. It should be noted that, in our infinite approach, the structure interacted with its mirror structure under periodical boundary condition. This result was expected to reflect the bulk properties without the consideration of the surface structure. MD simulation on finite system should be carried out to study the surface structure and compare with the structure in the core, as mentioned in the following.

V.3 Thermal conversion of cellulose allomorphs

V.3.1. Introduction

V.3.1.1. Iα to Iβ conversion

The coexistence of two allomorphs, namely Iα and Iβ, coexisting in native cellulose was first demonstrated by NMR spectroscopy in 1984 (Atalla and Vanderhart 1984). The spectra revealed that cellulose of various origins generally contained a mixture of both crystal forms but that their ratio varied depending on the source. The same technique was later used to follow the variation of the NMR spectra as a function of annealing temperature. It was found that the triclinic Iα could convert into monoclinic Iβ by annealing to 260° in polar medium or up to 280˚C in non-polar medium (Debzi *et al.* 1991; Horii *et al.* 1987; Wada *et al.* 2003; Yamamoto and Horii 1993). More recently, using X-ray diffraction, Wada *et al.* (2003) have observed that Iα could convert into Iβ by a simple thermal treatment in helium atmosphere. Two main mechanisms, chain translation and chain rotation, were proposed to describe the mechanism of conversion.

Molecular modeling has been used to help clarify the mechanism of this allomorphic conversion at the molecular level. Hardy and Sarko *et. al.* (1996) suggested a break-slip mechanism. Matthew *et. al.* (2012) studied the intermediate high temperature phase by using CHARMM and GLYCAM force fields. Results using the DFT method (Bučko *et al.* 2011) suggested two possible intermediate phases during the transition, which requires the chain translation both along the chain (*c*-axis) and the sheet direction (*b*-axis). All the modeling studies as well as the discussions based on experimental measurements (Nishiyama *et al.* 2003; Wada *et al.* 2003) agree that the chain translation mechanism seems to be more feasible than that involving a chain rotation, since the (110) and (200) lattice spacings in Iα and Iβ, respectively, and the additive 6% expansion by thermal treatment are not sufficient for a 180° chain rotation.

The temperature-induced phase transition of Iα to the high temperature form is the first step for the Iα to Iβ conversion (Wada *et al.* 2003). Then the chains reorganize somehow close to the

high temperature phase of Iβ and cool down to Iβ. Many carbohydrate force fields (CHARMM, GLYCAM, GROMOS) have reproduced this transition of cellulose Iβ to the high temperature phase. However, none of the existing force fields could reproduce the reversibility of this phase transition, and thus there is no way of reproducing the conversion path from Iα to Iβ in MD. In the former chapters, we have found that by slightly revising this parameter, a reversible transition that similar to experiment could be observed. This potentially gives a possibility of reproducing thermal conversion from cellulose Iα to Iβ conversion in MD simulation to extract molecular details of the conversion.

V.3.1.2. Iβ to III_I inter-conversion

The monoclinic unit cell of cellulose III_I contains only one independent chain (Wada *et al.* 2004). X-ray diffraction studies revealed that cellulose III_I can convert to Iβ with thermal treatment in hot water (Chanzy *et al.* 1986) or in helium atmosphere (Wada 2001) around 473 K, and can convert back from Iβ through a liquid ammonia treatment (Barry *et al.* 1936). The transformations require (1) switching of the orientation of the hydroxymethyl group between the *gt* and *tg* conformations, (2) slippage of chains. Based on the structural features of Iβ and III_I, the mechanism of transformation was proposed from experiment (Wada 2001) and molecular dynamics simulations using GLYCAM force field (Bellesia *et al.* 2011; Yui and Hayashi 2007; Yui and Hayashi 2009), but so far, cellulose III_I to Iβ has not yet been reproduced *in-silico*.

To further understand the mechanism of crystalline conversion from Iα to Iβ and III_I to Iβ by thermal treatment, we have performed MD simulation by using GROMOS force fields.

V.3.2. Computational details

The models, computational procedure, system setups and analysis tools are all identical to those described in chapters III and IV. The Gr53a6FF and Gr56A$_{carbo}$FF and two revised versions based (named revised Gr53a6FF and revised Gr56A$_{carbo}$FF) were used. In the step-wise heating and cooling process, the heating rate was 20 ps/K and the cooling rate 40 ps/K. The systems were equilibrated during 200 ps upon heating and 400 ps upon cooling every 10 K.

The translation of the chains was estimated from the difference of z-coordinate of O1 atoms between neighboring chains. The rotation of the chains was estimated by measuring the angle between O3-O6 vectors (using only x- and y-coordinates) in neighboring chains.

V.3.3. Iα to Iβ conversion by thermal treatment

V.3.3.1. Chain translation and chain rotation by using Gr53a6FF

It has been reported that the solid-state conversion of Iα to Iβ experimentally occurred between 530 and 550 K (Debzi *et al.* 1991; Horii *et al.* 1987; Wada *et al.* 2003; Yamamoto and Horii 1993). By performing MD simulation with Gr53a6FF, we have checked the amplitude of chain translation and chain rotation at 600 K, as shown in **Figure V.3.1**. The amplitudes of the translation were no more than 0.15 nm, which was too small compared with the target value of 0.26 nm (which is half length of one glucose residue). Similarly, the chains rotated by nearly 35° for Iα and Iβ, which, again, was far from the expected 90° value. Consequently, these preliminary calculations could not discriminate which is the likely molecular process occurring during the Iα to Iβ thermal conversion.

By using native and revised Gr56A$_{carbo}$FF as well as the revised Gr53a6FF, the amplitude of chain translation and rotation are also not large enough for the Iα to Iβ conversion.

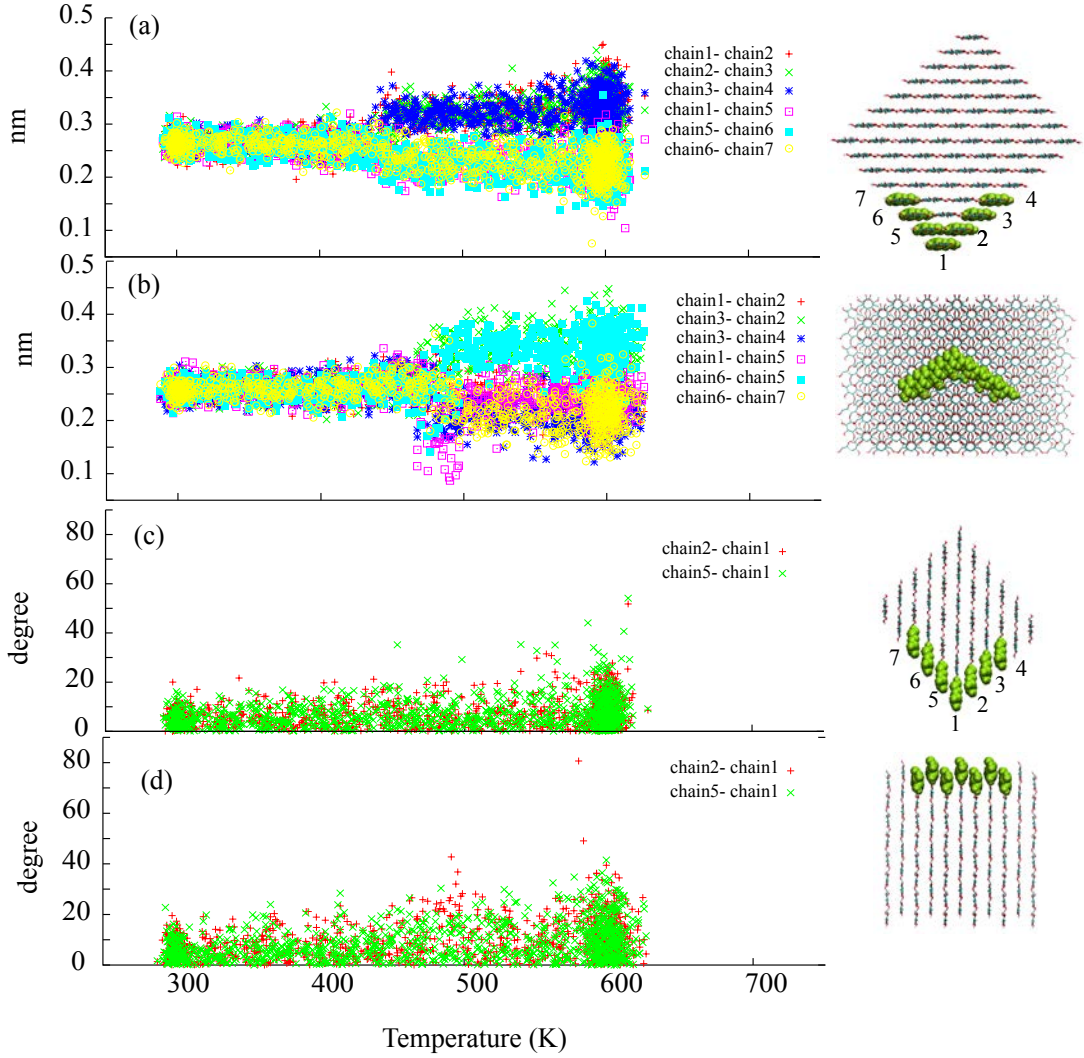

Figure V.3.1. Amplitude of chain translation and rotation in cellulose I as a function of temperature during heating: a) chain translation of cellulose Iα; b) chain translation of cellulose Iβ. c) chain rotation of cellulose Iα; d) chain rotation of cellulose Iβ. "chainN-chainP" means that packing difference or rotation amplitude between chain N and chain P in the chain direction. Chain numbers are indicated in the models on the right side of the figures.

V.3.3.2. Simulated transition temperature of Iα and Iβ

Experimentally, the structural transition of Iα occurs at higher temperature (530 K) than Iβ (500 K). However, our simulations by using the previously mentioned four force fields all give a slightly lower transition temperature for Iα. **Figure V.3.2** shows the population of *tg* conformation as function of temperature by using Gr53a6FF. A sudden decrease of *tg* from almost 100% to about 0% indicated the temperature where phase transition occurred, which is 440 K for Iα and 450 K for Iβ.

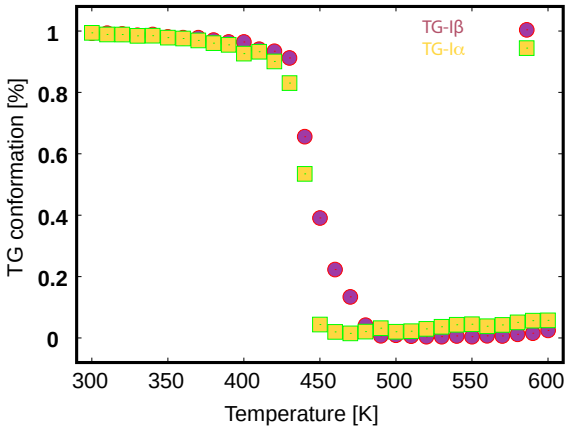

Figure V.3.2. Variation the population of *tg* conformation in Iα (yellow) and Iβ (purple) as a function of temperature during heating process by using Gr53a6FF.

V.3.3.3. Hydrogen-bonding network difference between Iα and Iβ

X-ray diffraction (Wada *et al.* 2003) and IR (Watanabe *et al.* 2006; Watanabe *et al.* 2007) measurements show that the hydrogen bonding significantly change after the phase transition, which could be due to the conformational variation of the of hydroxymethyl groups, since their orientation plays key roles in forming both intra- and inter-chain hydrogen bond networks. **Figure V.3.3** shows a snapshot of cellulose I at different temperatures calculated by using Gr53a6FF. At room temperature, before heating, the simulated HBs are identical to the experimental HB networks (pattern A). At 600 K, after the transition, inter-layer HBs are observed. The occurrence of inter-sheet HBs is higher in Iβ than in Iα (**Figure V.3.3**). After cooling to 300 K, there are almost no inter-sheet HBs in Iα but many remained in Iβ. The cooling structure of Iβ at 300 K is very similar to its high temperature structure.

At high temperature, just over the transition temperature, we have observed two different HB patterns. With native Gr53a6FF and revised Gr56A$_{carbo}$FF, the HB patterns in **Figures V.3.4** (left) and **V.3.4** (right) were found in Iα and Iβ, respectively. When the native Gr56A$_{carbo}$FF was used, the pattern in **Figure V.3.4** (left) was found in both Iα and Iβ, whereas when the revised Gr53a6FF was used, the pattern in **Figure V.3.4** (right) was found in both Iα and Iβ.

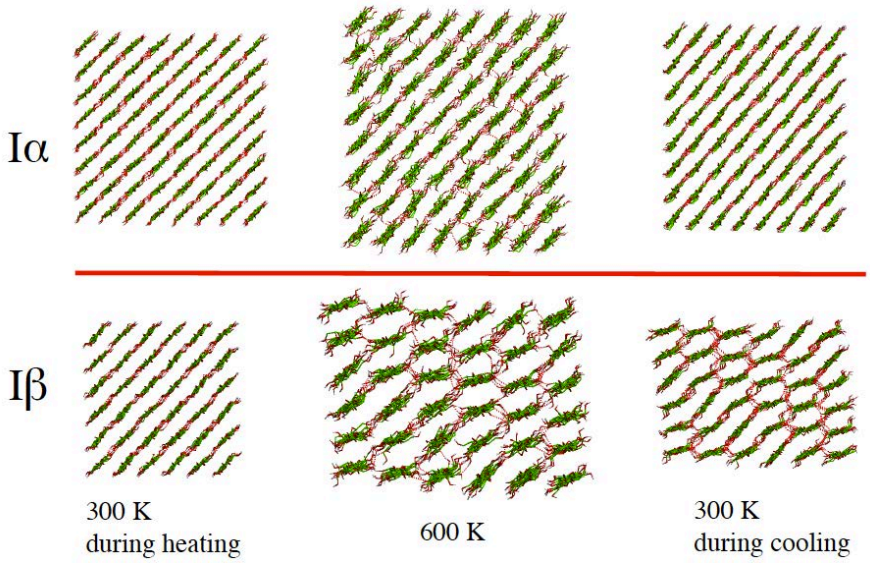

Figure V.3.3. Selected snapshots of cellulose I at 300 and 600 K during heating and cooling phases, simulated using Gr53a6FF. Hydrogen bonds are shown by red dotted lines.

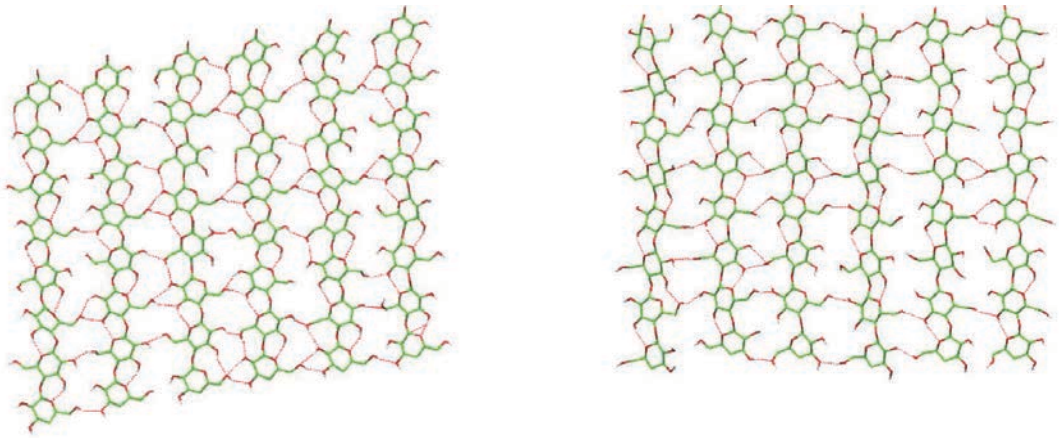

Figure V.3.4. Two types of hydrogen bonding patterns of Iα and Iβ during MD simulations by using four GROMOS force fields. Hydrogen bonds are indicated in red dotted lines. The left pattern is a snapshot from the high temperature structure of Iα by using native Gr53a6FF, while the right pattern is from Iβ by using the same force field

V.3.3.4. Comparison of energy between Iα and Iβ

Allomorph Iβ has been described as more thermodynamically stable than Iα by several authors (Heiner *et al.* 1995; Kroon-Batenburg and Kroon 1997; Wada *et al.* 2003) due to the irreversible conversion from Iα to Iβ. In agreement with previous MD studies (Heiner *et al.* 1995; Kroon-Batenburg and Kroon 1997), we found that the calculated energy of Iα using the various force fields was also higher than that of Iβ. There is a 1 kJ/mol difference between the

values calculated using native and revised Gr53a6FF and native Gr56A$_{carbo}$FF, and a 4 kJ/mol difference with revised Gr56A$_{carbo}$ (**Table V.3.1**), indicating that allomorph Iβ is more favorable. Both allomorphs have shown a sudden variation of each energy component during the phase transition.

By using native Gr53a6FF, Iα had a lower total energy per residue than Iβ when the system was cooled back to 300 K. At high temperature, the energy of Iα became lower than Iβ, which is caused by the larger decrease of the total energy during transition. A detailed analysis of the energy components per residue of Iα and Iβ are shown in **Figure V.3.5** and **Table V.3.1**. At 300 K, the triclinic phase (Iα) is 1.0 kJ/mol per residue higher than that of the monoclinic one (Iβ), which is mainly contributed from the short-range and long-range Coulomb interactions (excluding the 1-4 Coulomb interaction). At 600 K, the 0.9 kJ/mol per residue lower energy per residue of Iα than Iβ is mainly contributed from the 1-4 pair interaction, short and long Coulomb interactions. Since the cooled cellulose structures contain few *tg* conformation, their energy is different from that at 300 K before heating. The difference of energy originates mainly from the proper-dihedral and short- and long-range Coulomb interaction terms.

The energy components of Iα and Iβ per residue from Gr56A$_{carbo}$FF are listed in **Table V.3.1**. At 300 K, the energy of Iα is 1.0 kJ/mol per residue higher than that of Iβ, mainly due to the energy difference of Gangle (energy from bond bending), Coulomb 1-4 (coulomb interaction of 1-4 pairs) and Coulomb reciprocal interaction (coulomb interaction from atoms pairs separated by a distance larger than the defined cut-off radius). The transition temperature of Iα is 440 K, which is slightly lower than that of Iβ (450 K), similar to that observed using Gr53a6FF. Both of the cooled structures of cellulose Iα and Iβ have a higher energy than the never heated ones. The energy of the cooling structure of Iα at 300 K is about 0.7 kJ/mol higher than that of the heating structure, while the cooling one of Iβ is 1.3 kJ/mol higher than heating. Both differences were mainly due to contributions from the proper dihedral, the Coulomb 1-4, Coulomb short-range (coulomb interaction from pairs within the cut-off radius) and Coulomb reciprocal interactions. The energy of the cooling structure of Iα at 300 K is 0.4 kJ/mol higher than that of Iβ, which is mainly contributed from the energy difference of Coulomb 1-4 and Coulomb short-range interactions, whereas the energy of the high temperature structure of Iα is 5.9

kJ/mol lower than that of Iβ, mostly due to contributions from the proper dihedral and Coulomb 1-4 interactions.

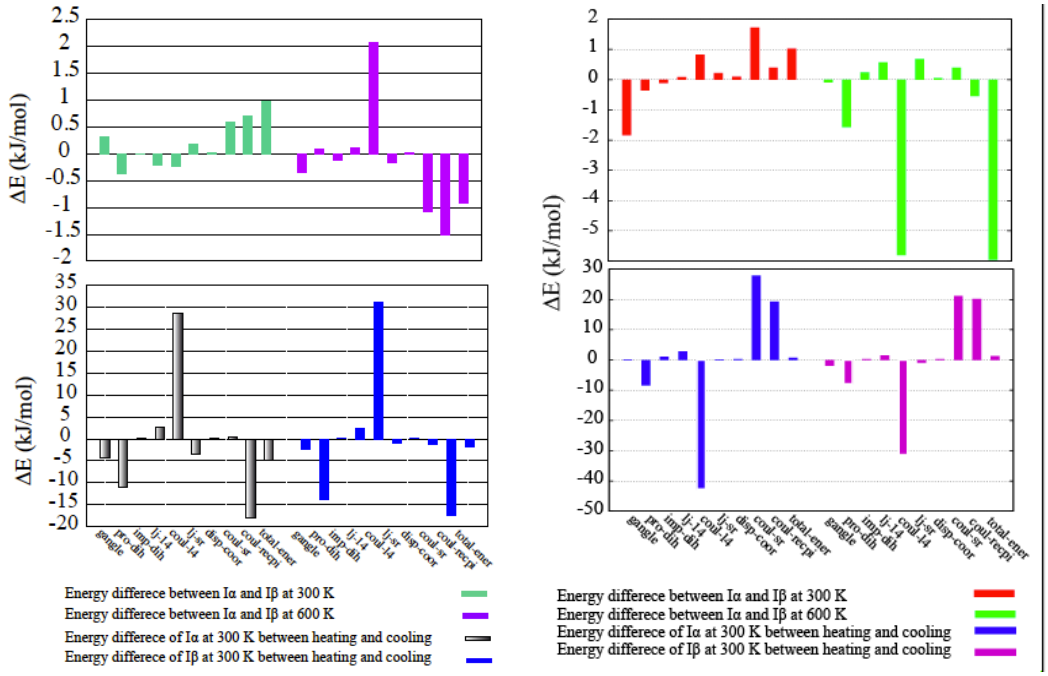

Figure V.3.5. Energy difference of energy components between cellulose Iα and Iβ at 300 and 600 K during heating and cooling by using native Gr53a6FF (left) and Gr56A$_{carbo}$FF (right).

By using the revised Gr53a6FF, we observed a reversible transition for both Iα and Iβ allomorphs, which can be characterized by the equivalent energy components between them. During the annealing process, the total energy of Iα is always 1~2 kJ/mol higher than that of Iβ, which is mainly attributed to the energy difference of Coulomb interactions (Coulomb 1-4, Coulomb SR and Coulomb reciprocal). By using the revised Gr56A$_{carbo}$FF, we observed a reversible transition for Iβ and an irreversible transition for Iα. The total energy of Iα is about 5 kJ/mol higher than that of Iβ, mainly due to the contribution from the proper dihedral and Coulomb interaction energy terms.

From **Table V.3.1**, one can see that the absolute value of the energy of the Coulomb interaction (Coulomb 14, Coulomb short-range and Coulomb reciprocal) account for a large fraction in the total energy. In comparison, the Gr56A$_{carbo}$FF yielded a lower electrostatic energy than Gr53a6FF, but is still significantly high. However, the experimental electrostatic interaction in crystalline cellulose may be not as high as the energy simulated here.

Table V.3.1. Energy components of cellulose Iα and Iβ at 300 and 600 K during heating and cooling by using native and revised GROMOS force fields. "c300 K" indicates the cooling phase of 300 K)

	Gangle	*Pro-dih*	*Imp-dih*	*LJ-14*	*Coul-14*	*LJ-SR*	*Disp-Coor*	*Coul-SR*	*Coul-Recp*	*Total*
Native Gr56A$_{carbo}$FF										
Iα 300K	29.1	52.7	4.6	-18.0	252.9	-30.4	-8.1	-72.0	116.6	361.0
Iα c300K	29.2	44.3	5.6	-15.2	210.2	-30.2	-7.7	-44.0	135.8	361.7
Iα 600K	52.6	51.1	9.9	-13.5	217.4	-32.1	-7.2	-20.8	133.9	470.0
Iβ 600K	52.7	52.7	9.7	-14.1	223.2	-32.8	-7.3	-21.2	134.4	475.9
Iβ 300K	30.9	53.1	4.7	-18.0	252.1	-30.7	-8.2	-73.7	116.2	360.0
Iβ c300K	29.0	45.2	5.0	-16.5	220.7	-31.5	-7.9	-52.5	136.3	361.3
Native Gr53A6FF										
Iα 300K	28.6	49.5	4.2	6.5	632.1	-20.0	-8.1	-126.1	-246.6	353.7
Iαc300K	24.2	38.4	4.4	9.1	660.7	-23.3	-7.9	-125.7	-264.6	349.0
Iα 600K	41.7	41.4	7.2	11.6	665.6	-23.3	-7.5	-109.9	-263.8	430.2
Iβ 600K	42.0	41.4	7.4	11.5	663.5	-23.2	-7.5	-108.9	-262.3	431.1
Iβ 300K	28.3	49.8	4.2	6.7	632.4	-20.2	-8.1	-126.7	-247.3	352.7
Iβc300K	26.0	36.0	4.3	9.3	663.4	-21.2	-7.9	-127.9	-264.6	350.9
Revised Gr56A$_{carbo}$FF										
Iα 300K	28.3	52.5	4.2	-26.1	251.8	-44.9	-8.8	-78.9	117.2	328.9
Iαc300K	26.2	49.6	4.0	-25.3	233.9	-47.2	-8.7	-63.0	125.5	328.6
Iα 600K	43.6	50.5	7.5	-22.8	224.8	-47.9	-8.3	-38.8	131.7	407.7
Iβ 600K	45.4	52.6	7.9	-23.3	230.0	-46.2	-8.3	-41.7	128.9	412.7
Iβ 300K	29.9	53.6	4.3	-26.2	251.2	-45.3	-8.9	-82.2	114.6	324.6
Iβc300K	29.9	53.6	4.4	-26.2	251.2	-45.4	-8.9	-82.2	114.6	324.6
Revised Gr53A6FF (with Gr56A$_{carbo}$FF charges)										
Iα 300K	27.0	33.4	3.9	15.4	746.9	-43.2	-8.8	-133.7	-318.3	356.2
Iαc300K	27.0	33.4	3.9	15.4	746.9	-43.2	-8.8	-133.7	-318.3	356.2
Iα 650K	44.3	39.9	7.6	19.2	749.3	-43.6	-8.1	-104.9	-319.1	457.4
Iβ 650K	44.3	39.7	7.4	19.2	750.3	-43.3	-8.1	-107.1	-319.7	455.6
Iβ 300K	27.3	33.4	4.0	15.8	747.4	-43.5	-8.8	-135.1	-319.4	355.0
Iβc300K	27.4	33.4	4.0	15.8	747.5	-43.4	-8.8	-135.3	-319.4	354.8

The currently employed force fields all yield energy differences between Iα and Iβ. At high temperature, due to the motion of the atoms, the chains are expected to have the probability to move with respect to each other to form either the Iα or the Iβ packing. A translation of 0.52 nm may need to overcome a large energy barrier, which may be due to the short-range interaction (both Coulomb SR and Lennard-Jones SR). As we have noticed, the Coulomb short-range interaction represents a large part of the total energy. The chain translation may be constrained by the electrostatic energy that is over-estimated by the force field. Thus, observation of the chain translation is not likely by using these force fields.

V.3.3.5. Structure deformation of cellulose Iα at high temperature

When Iα was heated up to 700 K, the types of structural deformations were observed using the native or revised Gr56A$_{carbo}$FF, as shown in **Figure V.3.6**. The annealing structure strongly depended on the initial shape of crystal models and on computer resource (the number of cores used for parallel calculation and types of the computers). Comparing the energy per residue among the cooling structures at 300 K, the one shown in **Figure V.3.6b** has is the lowest, while the one shown in **Figure V.3.6a** is the highest (**Table V.3.1**). I have noticed that only the *gt* conformation appeared in structure (b), a mixture of *tg* and *gt* occurred in structure (a) and a mix of *gg* and *gt* occurred in structure (c). The fact that structure (c) has the lowest energy and structure (a) has the highest energy, are probably due to the torsional energy profile of the revised Gr56A$_{carbo}$FF, which defined that the *gt* conformation has the lowest torsional energy and *tg* has the highest energy among the three staggered conformations.

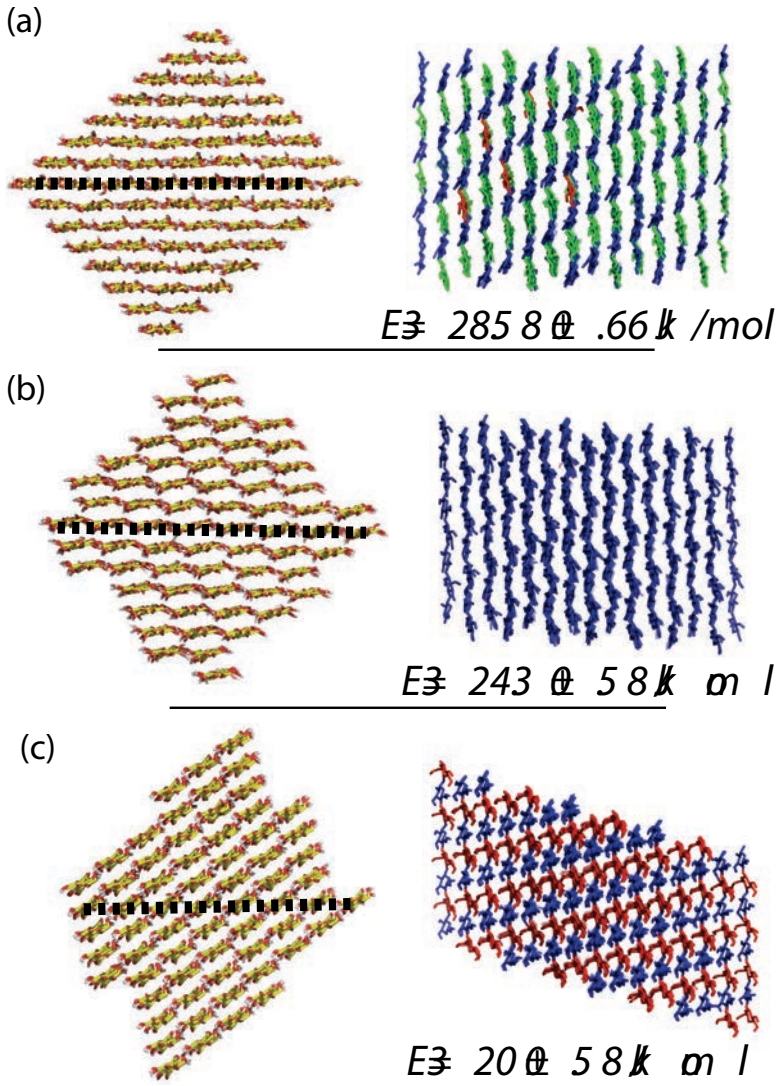

Figure V.3.6. Three types of cooling structures of Iα during annealing by using revised or native Gr56A$_{carbo}$FF. Color code: green: *tg*, blue: *gt*, red: *gg*. The black doted lines indicate the initial 110 phane. Energy per glucose is also shown (right). The energies showing in the graph is from the simulation using revised Gr56A$_{carbo}$FF.

It was only when we use the native Gr56A$_{carbo}$FF on a cluster computer (376 cores), chain translation can be obviously observed in the structure during "some" MD simulations, one of which is shown in **Figure V.3.7**. The chain packing changed to a monoclinic-line packing manner between 490 and 500 K, which is very similar to the packing of Iβ. However, in the following cooling process, both cellulose Iα and Iβ show few *tg* conformation at 300 K, which is not consistent with the experiments. The same MD file prepared for simulation did not produce the same results when using a server-computer (8 cores).

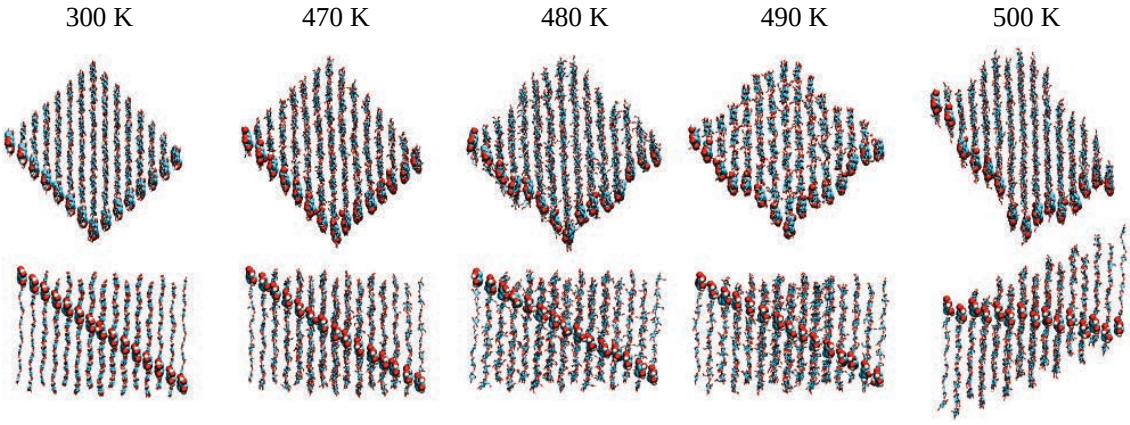

Figure V.3.7. Snapshots of cellulose Iα simulated using native Gr56A$_{carbo}$FF. The sphere models indicated the lattice arrangement of Iα.

V.3.4. Structure of cellulose II at high temperature by using revised Gr56A$_{carbo}$FF

The native Gr56A$_{carbo}$FF gives a cellulose II structure with about 10% of the hydroxymethyl groups exploring the *gg* conformation and 90% the *gt* one at room temperature (**Chapter III**). With the revised Gr56A$_{carbo}$FF, there is no *gg* and *tg* conformations but only *gt*.

A model containing 45 chains was heated up to 500 K and then cooled down to 300 K. As shown **in Figure V.3.8**, the variation of all structural parameters reveals a transition between 350 and 400 K. The unit cell parameters after cooling (300 K) are slightly different from those upon heating (300 K). The simulated unit cell parameters after cooling are closer to the experimental parameters than the ones simulated at 300 K. The variation of the unit cell parameters is different from the experimental data reported by Hori *et al.* (Hori and Wada 2006), but similar to those found in an earlier work (Takahashi and Takenaka 1982).

On an other hand, during the heating process, the hydroxymethyl group varied from 100% of *gt* conformation at 300 K, to a mix of 86% *gt*, 10% *gg* and 3% *tg* conformation at 500 K. During cooling, all the O6 varied back to the initial and experimental *gt* conformation.

The orientation of the hydroxyl group τ3 nearly did not change during heating and cooling. However, the orientation of two other hydroxyl groups τ2 and τ6 are significantly different between heating (300 K) and cooling (300 K), as shown in **Figures V.3.8** and **V.3.9**. After the transition, hydrogen bonds could still be formed with origin-origin, center-center and origin-center schemes. Interestingly, the oxygen atoms involved in HO2 or HO6 that used to

act as donors became acceptors after transition, whereas the oxygen atoms that acted as acceptors became donors, as shown by the gray shadow in **Figure V.3.9**. Upon cooling, the structure did not go back to the initial conformation but the high temperature one was preserved. Curiously, the conformation of τ2 and τ3 were different between the center and origin chains throughout heating and cooling.

The total energy of the cooling structure (300 K) was about 5.4 kJ/mol (**Table V.3.2**) lower than the heating one (300 K), indicating that the cooling structure was more stable than the starting one. This may due to that the formation of the new HB network is more extensive than the initial one. The semi-finite models of these two structures in water have been simulated to further compare the structure stability between them, which will be discussed in the next subchapter.

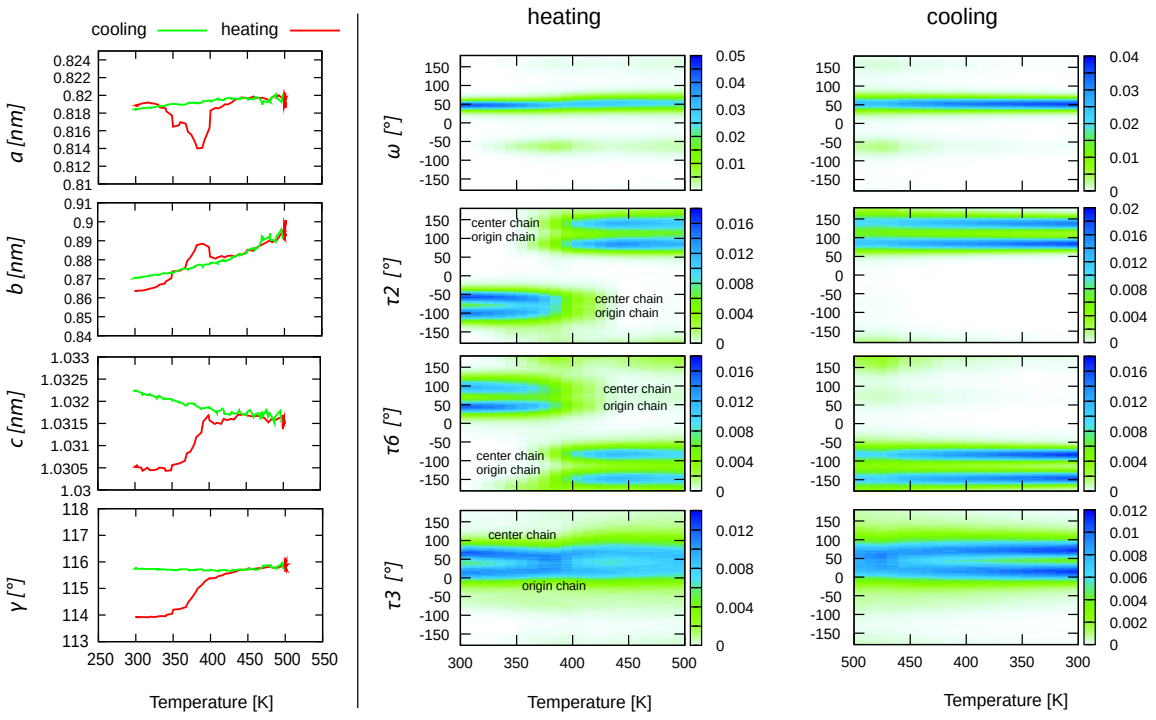

Figure V.3.8. Conformational parameters of cellulose II as a function of temperature: a) unit cell parameters; b) conformation of the hydroxymethyl group and hydroxyl groups during heating; c) conformation of the hydroxymethyl and hydroxyl groups during cooling.

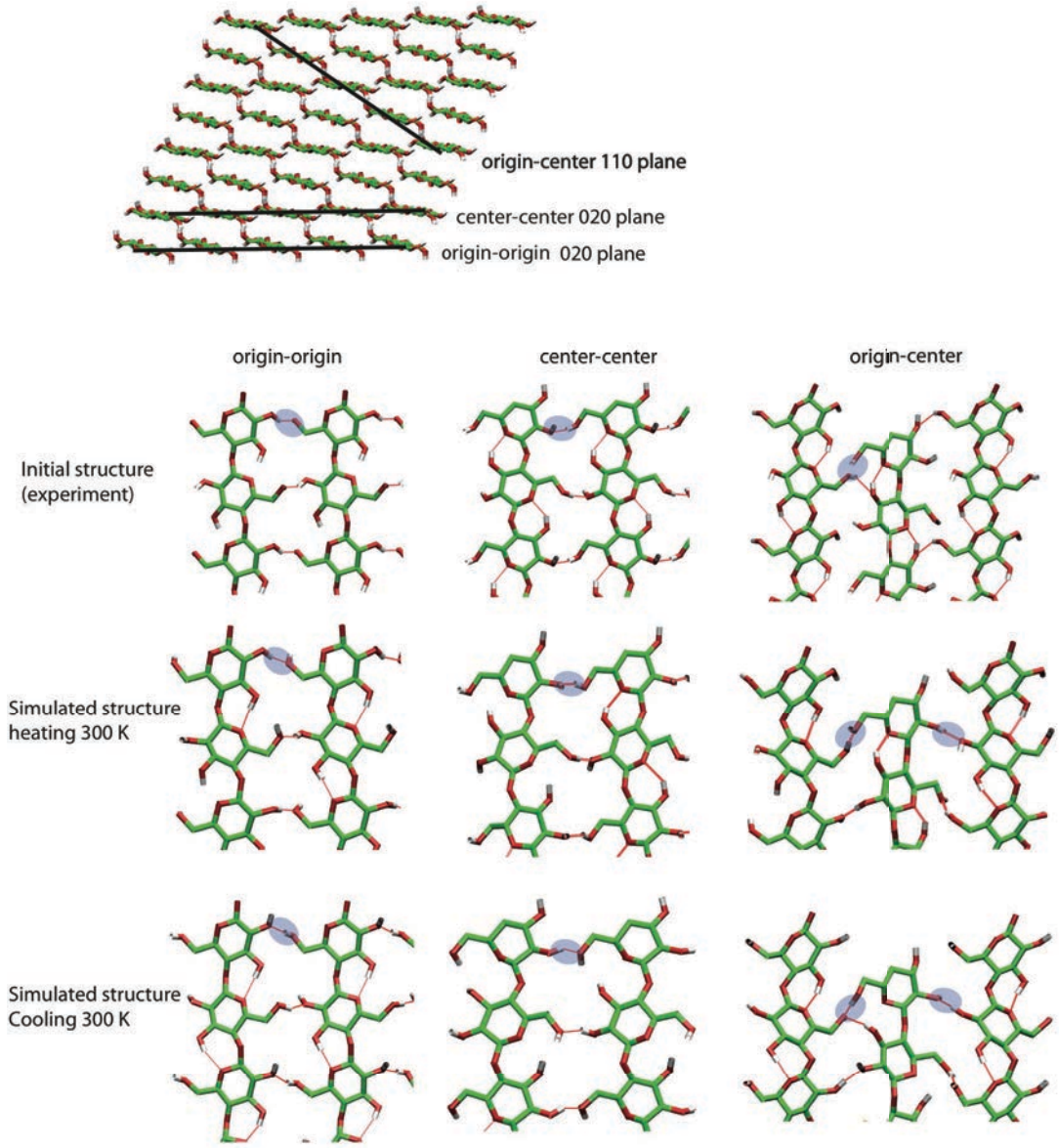

Figure V.3.9. Snapshots of HB patterns of cellulose II from experiment and MD simulation.

V.3.5. Transition from cellulose III_I to $I\beta$

A model of cellulose III_I containing 45 chains was heated up to 600 K and then cooled down to 300 K by using the infinite approach and revised $Gr56A_{carbo}FF$. By following the variation of the unit cell parameter and conformation of the hydroxymethyl group, we have observed two transitions, one around 400 K and the other around 550 K.

As shown in **Figure V.3.10**, upon increasing the temperature, up to 36% of *gg* conformation appeared between 390 and 490 K. Subsequently, all of them went back to *gt* conformation. At 550 K, the *gg* population increased again, yielding a system containing 53% *gg* and 47% *gt* up

to 600 K. The mix of *gg* and *gt* conformation was maintained until the system was cooled down to 300 K.

The hydroxyl groups varied accordingly at the transition temperature (**Figure V.3.10**). All these results indicate that the HB systems at 600 K and after cooling (300 K) are very different from the initial one.

The unit cell parameters slightly varied at the first transition temperature 400 K but significantly changed upon the second transition (550 K). The chains re-arranged to form a layered structure. As shown in **Figure V.3.10c**, the initial (100) planes of III_I formed the (200) planes of Iβ, which is different from the mechanism proposed by Wada *et al.* (2001), in which the $(1\bar{1}0)$ planes form the (200) planes of Iβ. However, none of the hydroxymethyl groups adopts a *tg* conformation, suggesting that the cooling structure after transition was not Iβ. In contrast, the chain packing seems to be very similar to one type of the cooling structures of Iα (**Figure V.3.6.c**).

Experimentally, the thermally-induced III_I to Iβ conversion occurred between 470 and 500 K. We have tried to cool the system from 500 K. The cooling structure is different from the one obtained after cooling down from 600 K. The total energy of each cooling structure are both lower than the initial one, as shown in **Table V.3.2**.

By using two other force fields, the revised Gr53a6FF and native $Gr56A_{carbo}FF$, we have observed that part of infinite model exhibited the layer structure of cellulose I. However, none of them correspond to allomorph Iβ. Furthermore, we have noticed that the reported MD studies on cellulose III_I, which were only carried out by using GLYCAM 06 force field (Yui and Hayashi 2007; Yui and Hayashi 2009), showed that the structure had a tendency to deform as we also observed by using the native Gr53a6FF. The MD studies on cellulose III_I based on empirical force field methods are still a challenge.

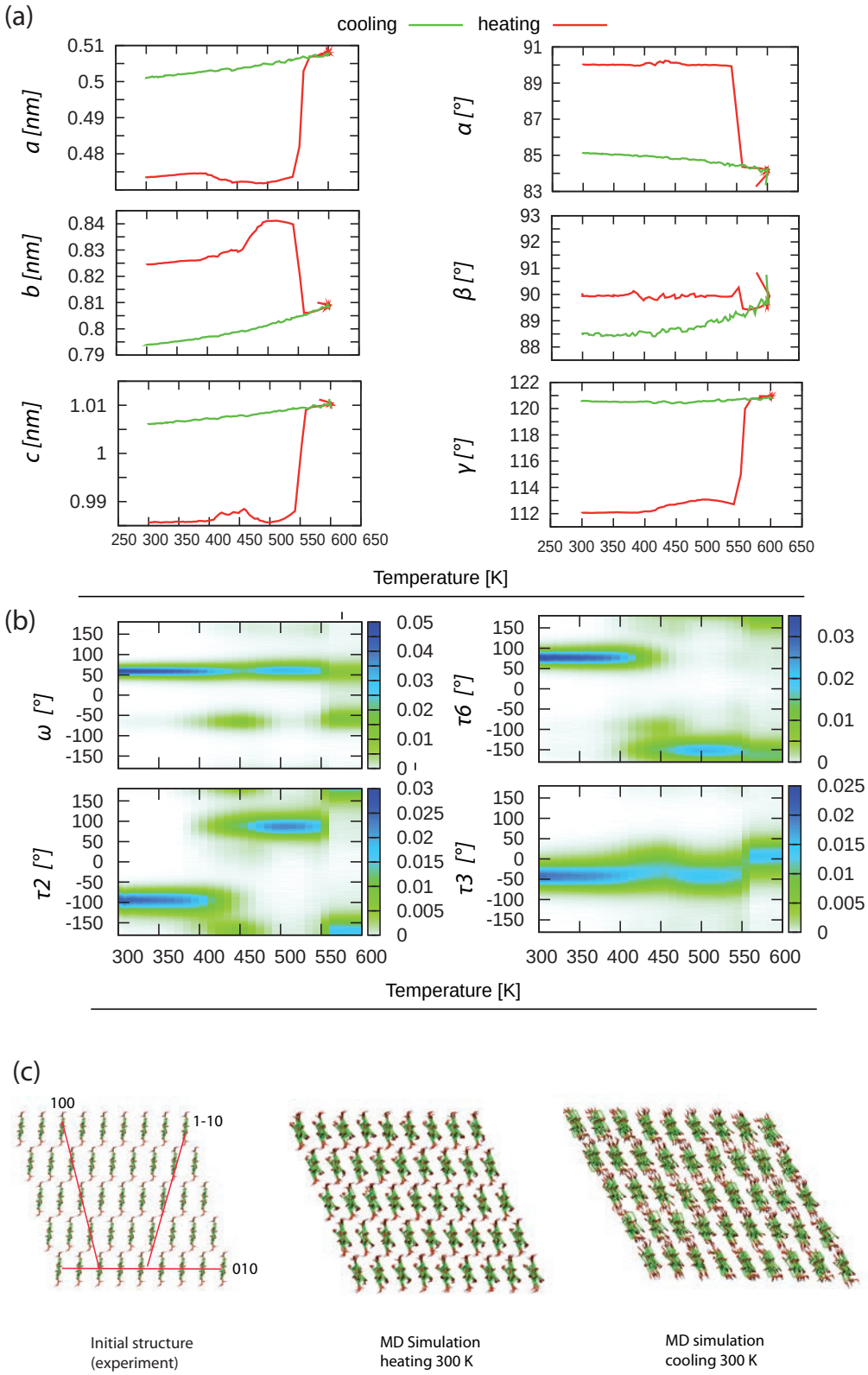

Figure V.3.10. Variation of the conformational parameters of cellulose III_I as a function of temperature (a: the unit cell parameters, b: the hydroxyl and hydroxymethyl groups; c: snapshots of cellulose III_I).

Table V.3.2. Energy per residue of cellulose allomorphs before heating (300 K) and after cooling (300 K) by using revised Gr56A$_{carbo}$FF.

		Energy (kJ/mol)			
		Iα	Iβ	II	III_I
300 K (heating)	Model 1	328.9±0.58	324.6±0.77	329.4±0.74	337±0.69
	Model 2	328.2±0.59	330.7±0.80		
	Model 3	333.1±0.63	329.7±0.81		
300 K (cooling)		324.3±0.58 *			329±0.68 *
		320±0.58 *	324.6±0.77	324.6±0.73	331.3±0.68 *
		328.6±0.6 *			328.9±0.7 *

* *Three sets of cooling energy indicate that three different structures were obtained during cooling.*

V.3.6. Conclusion

The solid-state conversion of Iα to Iβ was not observed during the MD simulation by using GROMOS force fields, probably due to the over-estimated short-range interaction in cellulose I. Several experimental observations could not be reproduced, such as the higher transition temperature of Iα than Iβ. MD simulation on the thermal behaviors of surface chains of cellulose should also be preformed to examine if the chain translation would occur or not.

More experimental observations on both room and high temperature structures of cellulose I are needed to be compared with the simulation in order to develop more accurate force field parameters, such atomic partial charges and LJ parameters as well as torsional parameters. Trajectory files that provided by different computers give different results, indicating that the computational manner referred to the parallel computing have effects on the simulation result when native Gr56A$_{carbo}$FF is used at high temperature.

The MD studies suggest that the hydrogen-bonding network of cellulose II evolved to a more stable type after thermal treatment without the variation of the conformation of the hydroxymethyl group. Diffraction patterns should be calculated to validate the simulated model here.

The allomorphic conversion from cellulose III_I to Iβ was not observed in the simulations using GROMOS force fields, probably due to the poorly reproduced crystal structure of cellulose III_I.

V.4. Models of cellulose allomorphs in water

V.4.1. Introduction

This subchapter describes both surface and core structures of four cellulose allomorphs by performing MD simulation on the solvated models using four GROMOS force fields (native Gr53a6FF and Gr56A$_{carbo}$FF and their revised versions), and compared the results for infinite, semi-finite and finite models.

V.4.2. Computational details

Table V.4.1. Some characteristics of the models. Infinite models: the data are taken from the former chapters.

Code	Phase	Surfaces	Dimension (nm)	Chain Number	SPC water Number	Residues per chain
Semi-finite	Iα	100, 010	4.16, 6, 6	36	3264	8
	Iα	110, 1-10	4.16, 6.5, 6	40	3489	8
	Iα	100, 010	7, 7, 4.16	64	3795	8
	Iβ	110, 1-10	6, 6, 4.152	36	3254	8
	Iβ	200, 010	6, 6, 4.152	40	3078	8
	II	020, 1-10	7, 7, 4.124	40	4586	8
	III$_I$	100, 010	6, 6, 4.124	45	4586	8
Finite	Iα	100, 010	30.8, 7.3, 7.3	36	45703	40
	Iβ	110, 1-10	7.2, 7.6, 30,7	36	46586	40
Infinite	Iα	No		64	0	8
	Iβ	No		36	0	8
	II	No		40	0	8
	III$_I$	No		45	0	8

Our previous studies considered only infinite systems, which contain an infinite number of cellulose chains with infinite length. In this subchapter, two types of finite models have been

used. One type contains a finite number of infinite length cellulose chains, while the other type contains a finite number of finite length cellulose chains. They will be referred to as semi-infinite and finite models, respectively, in the following. All the starting coordinates were taken from the X-ray and neutron diffraction studies (Nishiyama *et al.* 2002, 2003; Langan *et al* 1999, 2001; Wada *et al.* 2004).

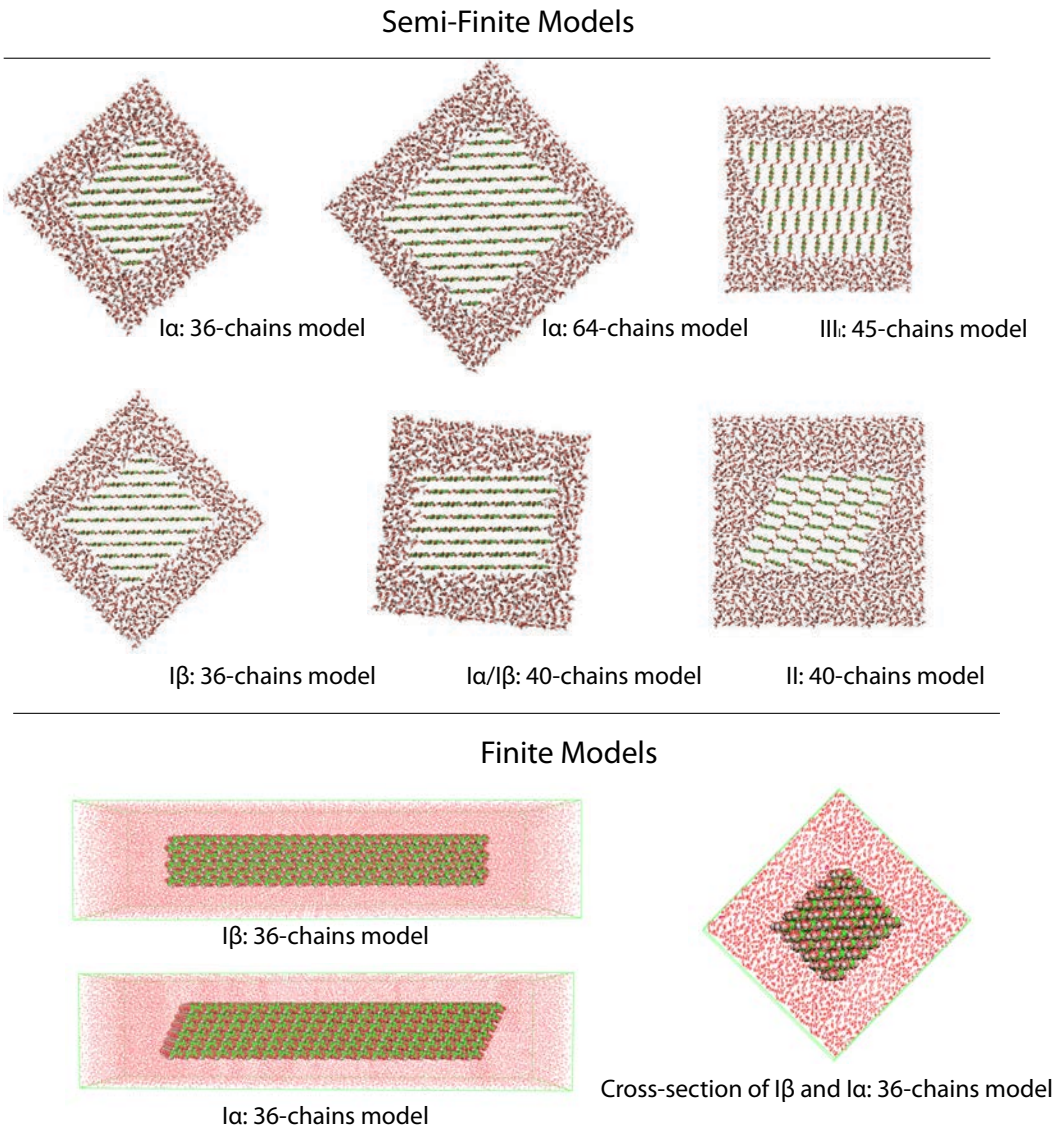

Figure V.4.1. Semi-finite and finite models of four cellulose allomorphs surrounded by water.

V.4.2.1. Semi-finite models

Table V.4.1 summarizes the details of the simulation box, crystal model and water molecules

of four cellulose allomorphs, which are described in the next paragraph. All the models (semi-finite and the finite) are shown in **Figure V.4.1.**

V.4.2.2. Computational procedures

The pressure coupling type was isotropic and a compressibility of 4.5×10^{-5} was used. The coordinates were saved every 20 ps. All other system setups were identical as previously described. The initial models were all firstly subjected to energy minimization by steepest-descent method and followed by a conjugate gradient method. The convergence criterion to stop minimization was a maximum force of 1 kJ mol^{-1} $Å^{-1}$. Subsequently, MD simulation with position constraints was performed for 200 ps to remove the "bad" contact between cellulose molecules and water solvent molecules. Then, the systems were heated from 0 to 300 K in a stepwise manner in 6 ns and extended for 94 ns or 100 ns of MD simulation. Note that the results related to infinite models are from our former chapters. The simulation scales of them are 16 ns, including 6 ns of initially heating from 0 K to 300 K and 10 ns of equilibrium at 300 K.

V.4.2.3. Unit cell calculation

Unit cell parameters were calculated using the coordinates of atom C1 in center four chains of the core extracted from the last 20 ns of the trajectory file.

V.4.3. Results and discussion

The RMSD, the variation of ω and unit cell parameters are used to characterize the modeled structures. When the conformation of ω in the core departed from experimental data, the structure was considered as deformed. **Table V.4.2** lists the stability of solvated model by using four GROMOS force fields. The structure details of each model will be discussed later.

The stability of the different systems can be estimated from the deviation from the crystal structure, a feature quantified by the RMSD. Most RMSD at 100 ns remained reasonably stable (except the semi-finite models of Iβ using native $Gr56A_{carbo}FF$), indicating that the equilibrium

reached for most models.

Table V.4.2. Stability of solvated cellulose allomorphs modeled by using native and revised GROMOS force fields. The ✓ and × symbols were used when the structure were stable and unstable, respectively. n.t.: not tested.

	Allomorphs	Native force field		Revised force field	
		Gr53a6FF	Gr56A$_{carbo}$FF	Gr53a6FF	Gr56A$_{carbo}$FF
Semi-finite	Iα	✓	×	✓	× & ✓
models	Iβ	×	✓	✓	✓
	II	n.t.	n.t.	×	×
	III$_I$	n.t.	n.t.	n.t.	×
Finite models	Iα	n.t.	n.t.	n.t.	×
	Iβ	n.t.	n.t.	n.t.	✓
Infinite	Iα	✓	✓	✓	✓
Models	Iβ	✓	✓	✓	✓
	II	×	✓	✓	✓
	III$_I$	×	✓	×	✓

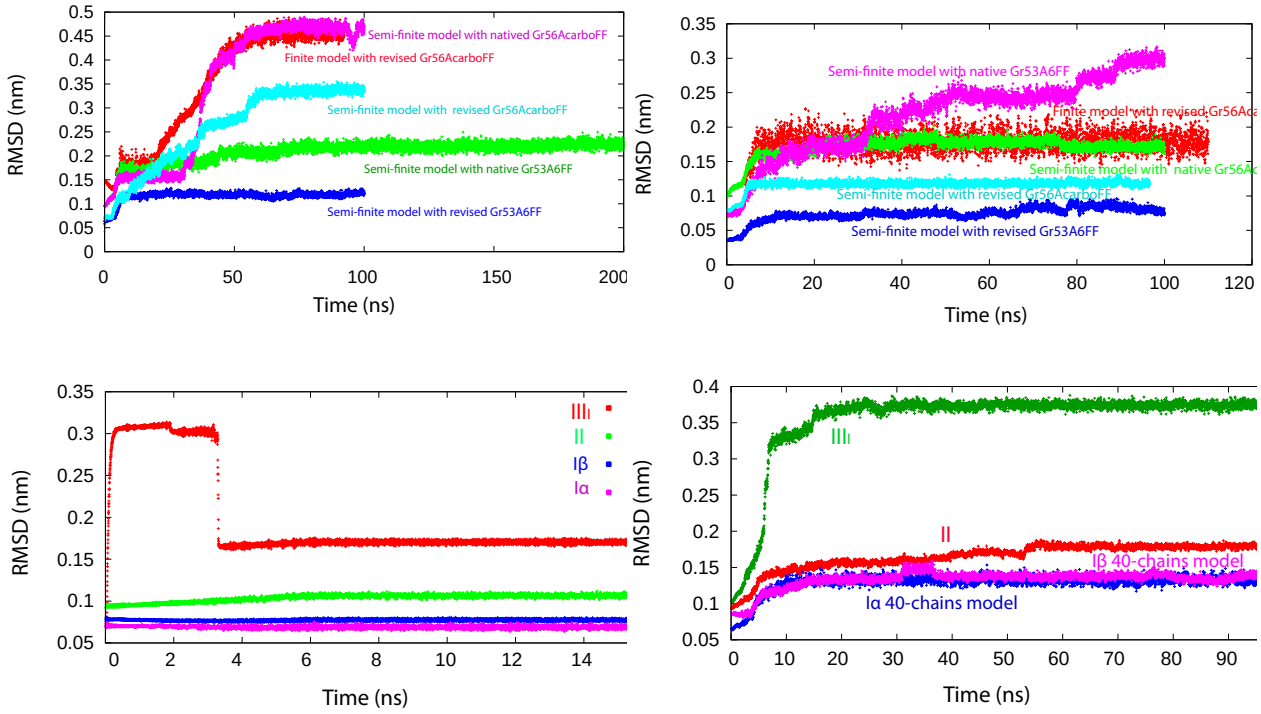

Figure V.4.3.2. RMSD of various models as a function of time: a) Iα; b) Iβ. The semi-finite model of Iβ contains 36 chains. c) infinite models simulated using revised Gr56A$_{carbo}$FF; d) four allomorphs with semi-finite model using revised Gr56A$_{carbo}$FF.

Low values of RMSD at equilibrium indicated that the model remained close to initial structure during MD simulation. The stable structures validated both the modeling strategy and the force field parameters. In contrast, the high values indicated substantial conformational and/or organizational variations from the initial structure. The unstable structures indicated that either modeling strategy or the force field parameters or both of them were inadequate.

V.4.3.1. Deformed structure of Iβ by using native Gr53a6FF

The 36-chain model of Iβ and Iα both consisted of 20 chains in the surface structure and 16 chains in the core. **Figure V.4.3** shows the snapshots of cellulose Iβ at different simulation time by using native Gr53a6FF. At 20 ns, the ω in the interior is dominated by *tg* conformation. However, at 100 ns, the conformation of ω in the interior mostly changed to *gg* and *gt* conformations. The crystal evolved to a parallel-down-like structure, similar to that reported by Matthew *et al.* (2012).

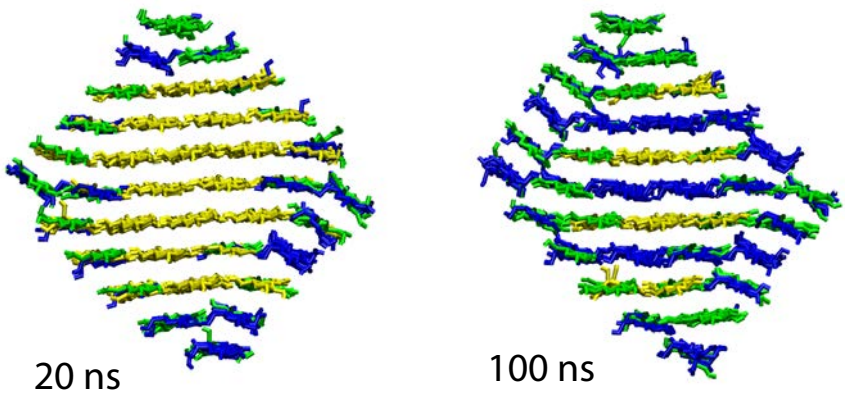

Figure V.4.3. Snapshots of Iβ at 20 ns and 100 ns by using native Gr53a6FF labeled by the three-staggered conformations of ω. Color code: yellow: *tg*, blue: *gg*, green: *gt*).

V.4.3.2. Deformed and undeformed structures of Iα by using native and revised Gr56A$_{carbo}$FF

Two models of Iα with square cross section, which contain 36 and 64 chains, respectively, were modeled using native and revised Gr56A$_{carbo}$FF. Both force fields yield a deformed structure between 20 and 40 ns, as suggested by the variation of the conformation of ω. At 100 ns, by

using native Gr56A$_{carbo}$FF, *tg* is minor in the core, while when using revised Gr56A$_{carbo}$FF, half of the ω evolved to *gg* and *gt* conformation, as shown in **Figure V.4.4**. The packing in the deformed structures is different from the initial one.

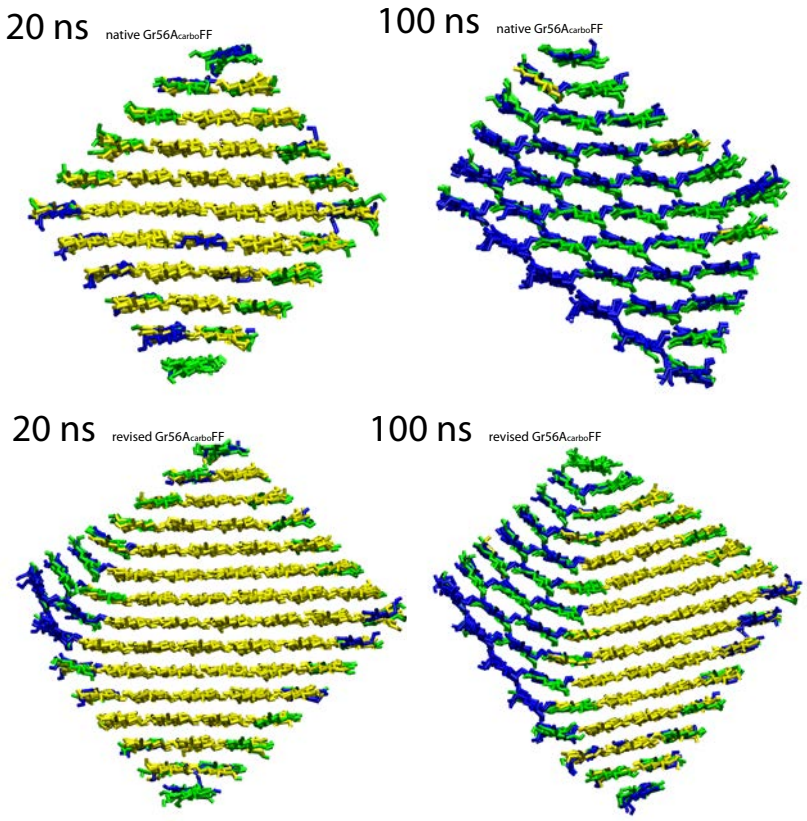

Figure V.4.4. Snapshots of Iα at 20 and 100 ns by using native and revised Gr56A$_{carbo}$FF. Color code: yellow: *tg*, blue: *gg*, green: *gt*.

When another models of Iα, the 40-chains models, was performed MD by using revised Gr56A$_{carbo}$FF, the core of the models contains only *tg* conformation for ω at 100 ns (**Figure V.4.5-bottom**), which differs from the 64-chains model using the same force field. The RMSD (**Figure V.4.3.2d**) shows that the structure already reached equilibrium after 20 ns. These results indicate that the stability of cellulose Iα has the model dependence on the revised Gr56A$_{carbo}$FF. However, this is not the case for Iβ when using revised Gr56A$_{carbo}$FF (**Figure V.4.5-top**), which will be shown later.

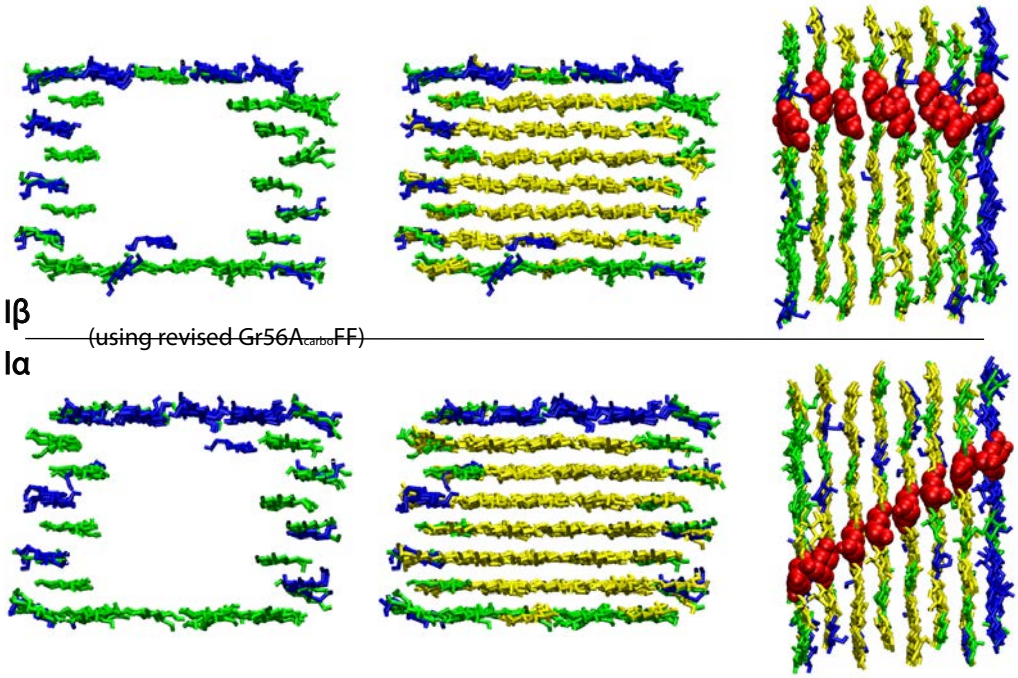

Figure V.4.5. Snapshots of the 40-chains models of Iα and Iβ at 100 ns by using revised Gr56A$_{carbo}$FF. Color code: yellow: *tg*, blue: *gg*, green: *gt*. The sphere models in red show the lattice characteristic of Iα and Iβ.

V.4.3.3. Structure of cellulose Iβ by using the revised Gr53a6FF, native Gr56A$_{carbo}$FF and revised Gr56A$_{carbo}$FF

Three force fields, namely revised Gr53a6FF, native Gr56A$_{carbo}$FF and revised Gr56A$_{carbo}$FF, could all yield structures with the *tg* conformation being dominant in the core region after 100 ns of MD simulation, as shown in **Figure V.4.6.**

The surface glucose residues can be generally divided into two types in the 36-chain model: one type with its hydroxymethyl group oriented to the interior and the other type with the hydroxymethyl group oriented to the solvent molecules. The populations of the three staggered conformations of ω located in surface and core by using different force fields have been plotted in **Figure V.4.6** (right). All of them identically show that the core structure adopted almost only the *tg* conformation (over 94%) and the surface is a mixture of *tg*, *gg* and *gt* conformations.

By using native and revised Gr56A$_{carbo}$FF, the surface glucoses with O6 oriented to water molecules were dominated by *gt* and *gg*, similar to the conformational distribution of a single glucose in water measured from experiment (Nishida *et al.* 1988) and MD simulation (Lins and

Hünenberger 2005). In contrast, by using revised Gr53a6FF, a population of 43% *tg* was observed. Most of the theoretical and experimental studies reveal that *tg* is few in water for glucose. The revised Gr53a6FF employed probably overestimated the stability of *tg* conformation of glucose in water.

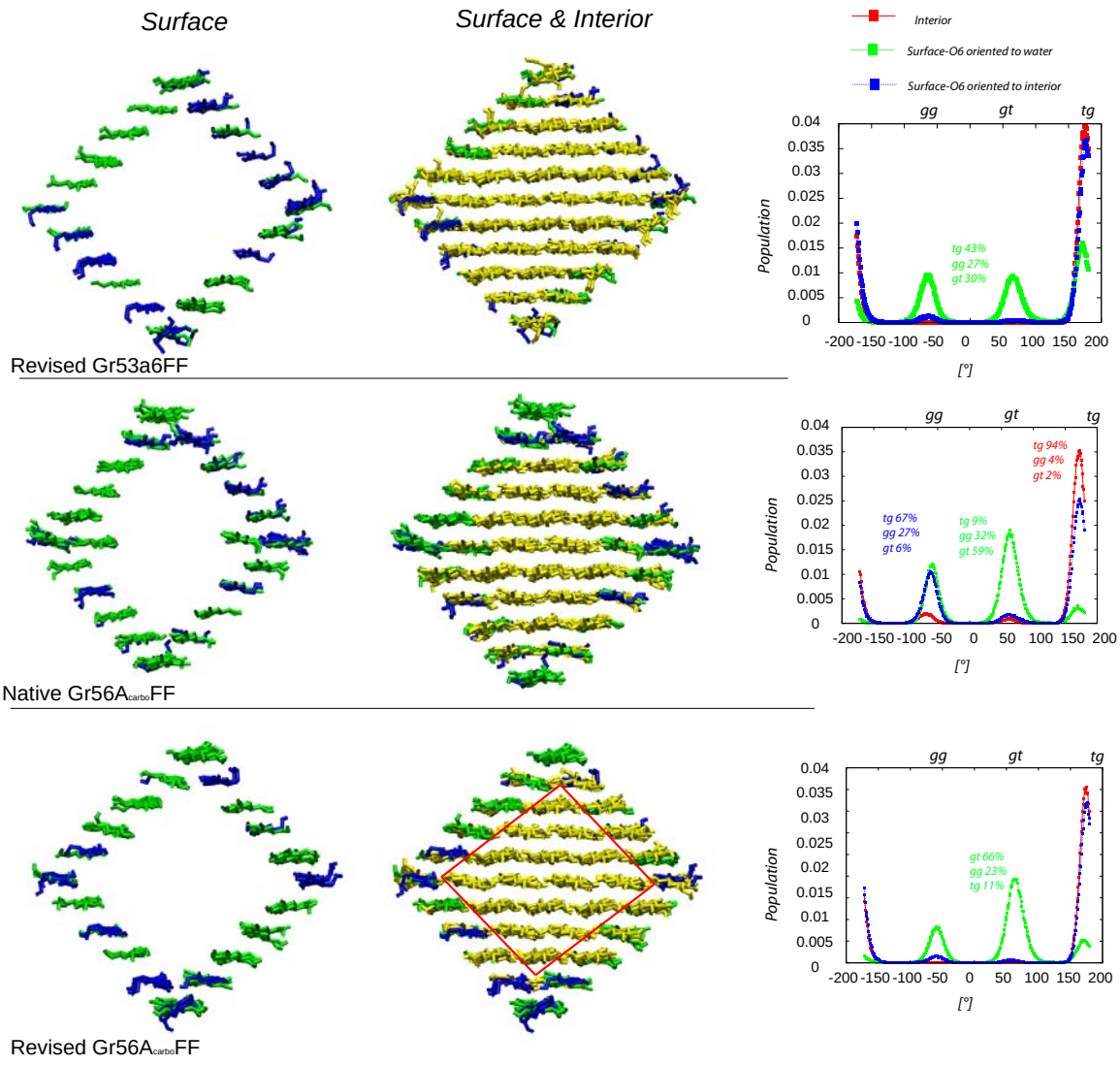

Figure V.4.6. Snapshots of Iβ after 100 ns MD simulation by using revised Gr53a6FF, native Gr56A$_{carbo}$FF and revised Gr56A$_{carbo}$FF (left and middle, yellow: *tg*, blue: *gg*, green: *gt*) and the populations of the conformation of ω in different region of crystal (right).

V.4.3.4. Structure of cellulose Iα by using native and revised Gr53a6FF

The native Gr53a6FF yields a deformed structure for Iβ but not for Iα. After 100 ns simulation, 76% of the hydroxymethyl groups in the core structure retained their *tg* conformation, as shown

in **Figure V.4.6**. Those located at the surface contained about 56% *gt*, 43% *gg* and 1% *tg* (see **Table V.4.3**). These amounts are consistent with the conformational distribution of ω of a single glucose in water simulated by using this force field.

Based on the occurrence of *gg* (6%) and *gt* (18%) conformations in the core, one may conjecture that the structure was gradually varying during simulation. Thus, another 100 ns of MD simulation was extended. The result showed unchanged populations of the three conformations compared to that at 100 ns (**Figure V.4.6**), indicating that the MD had already reached the equilibrium and that the core structure of Iα with *tg* conformation was stable.

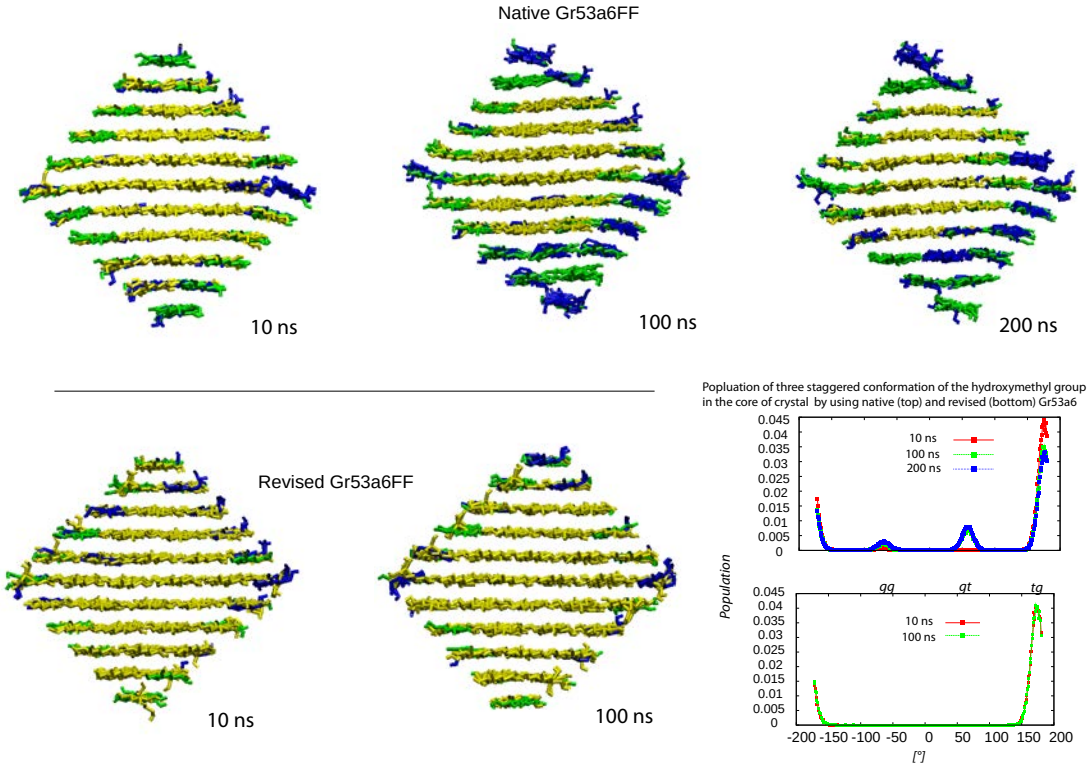

Figure V.4.6. Snapshots of Iα at varied simulation time by using native and revised Gr53a6FF. Color code: yellow: *tg*, blue: *gg*, green: *gt*. Lower right: populations of the conformation of ω in the crystal core.

The simulation with the revised Gr53a6FF showed that the core only contained *tg* conformation (100%) after 100 ns. The surface structure where O6 oriented to solvent contained 35% *tg*, 27% *gg* and 38% *gt* conformation. A larger population of *gt* in plane 010 than in plane 100 was observed for both native and revised Gr53a6FF.

Table V.4.3. Population of the three-staggered conformation of ω at the surface of Iα by using native and revised Gr53a6FF. The values were calculated from the last 20 ns of the 200 or 100 ns trajectory file. Fours chains in the middle of each surface of the 36-chain models are considered for estimating the population in (100) and (010) planes.

Force field	010			100		
	gt	*gg*	*tg*	*gt*	*gg*	*tg*
Native Gr53a6FF	91%	8%	1%	42%	57%	1%
Revised Gr53a6FF	49%	12%	39%	23%	34%	43%
	Surface (O6 oriented to solvent)			Surface (O6 oriented to core)		
	gt	*gg*	*tg*	*gt*	*gg*	*tg*
Native Gr53a6FF	56%	43%	1%	32%	28%	40%
Revised Gr53a6FF	35%	27%	38%	0%	0%	100%

V.4.3.5. The structure of cellulose II and III_I by using revised Gr56A$_{carbo}$FF

As shown in **Figure V.4.8**, the crystal shape of cellulose II was maintained during the solvated MD simulation, which means that the packing remained unchanged. However, the conformation of ω in the core region, partially changed from *gt* to *gg* conformation after 40 ns.

In **§ V.3.4**, I have shown that when cellulose II was heated to 500 K, a more stable structure (lower energy) could be obtained during cooling. Some basically structural characteristics of cooling structure, such as the conformation for the hydroxymethyl group and unit cell parameters, are more similar to experiment. However, the hydrogen bond pattern exchanged in center-center and origin-origin layers. To test the stability of this model, we have also run MD with its solvated models. The RMSD (**Figure V.4.9**) shows that the structure reached the equilibrium in 100 ns. Different to the experimental models, the core of this model still retains the *gt* conformation after 100 ns, indicating that the model in water is also stable than the experimental one.

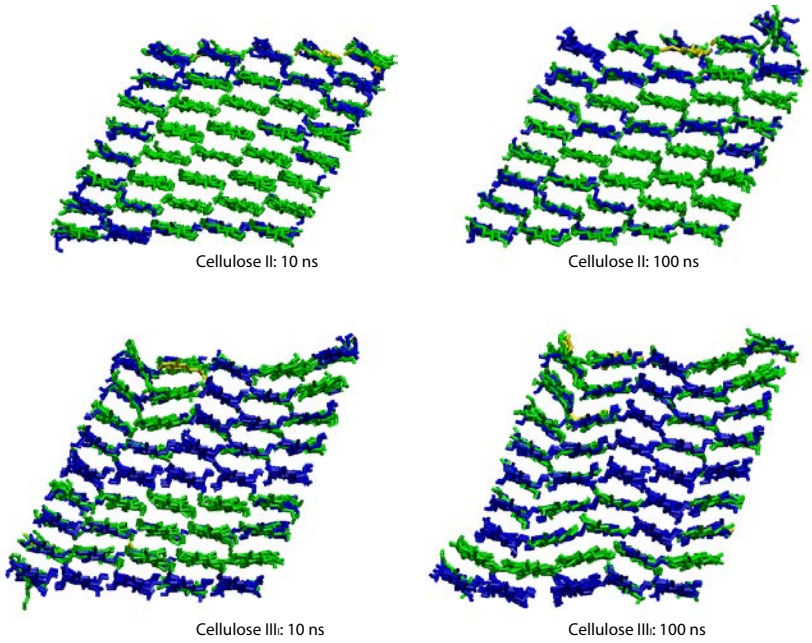

Figure V.4.8. Snapshots of cellulose II and III_I at 10 and 100 ns simulated by using the revised $Gr56A_{carbo}FF$.

From above, the disagreement between experiment and simulation suggests that either the simulated or experimental structure is wrong. However, based only on the MD simulation here, we cannot conclude that experimental structure is or is probably wrong, since a small variation of the force field parameters may give very different results. Thus, more probably, the current force field still needs to be further optimized.

The 45-chain model of cellulose III_I rapidly exhibits structural deformation after 7 ns when the temperature reached 300 K. The packing of cellulose chains was completely changed compared to experiment (**Figure V.4.8**). The deformation of cellulose III_I in solvent has also been observed by using GLYCAM 06 force field (Yui and Hayashi 2009).

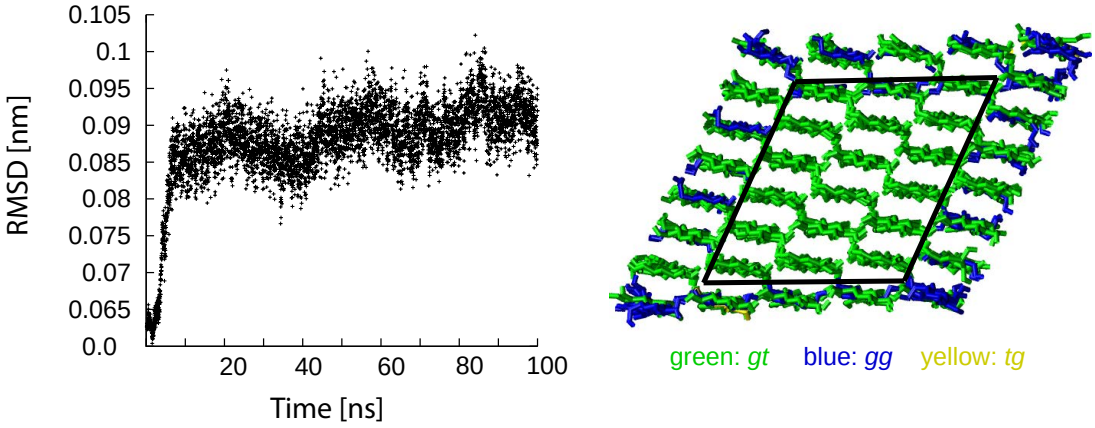

Figure V.4.9. RMSD of cellulose II (the cooling model in **§ V.3.4**) in water (left) and the snapshot at 100 ns.

V.4.3.6. Finite model of cellulose I

Finite models of cellulose I were constructed and subjected to MD by using the revised Gr56A$_{carbo}$FF, which were stable structure for the semi-finite models of Iβ but deformed for Iα. Here, the simulated structure of Iα was also deformed after 20 ns. The final structure is similar to the deformed structure of semi-finite models by using native Gr56A$_{carbo}$FF. However, by using the revised Gr56A$_{carbo}$FF, the finite model of Iβ was still stable, as its semi-finite model. **Figure V.4.10** shows the snapshots of one layer in 200 plane. The right hand twist in the fiber can be observed.

Several MD studies (Matthew *et al.* 2012; Yui *et al.* 2009, 2010) on finite models of cellulose allomorphs have shown that their structures deformed when solvent molecules were present, yielding structures different from the experimental ones. Thus, they proposed that the individual microfibril may have different structure compared with the large crystals which were used for determination of the crystal structure of cellulose. However, their results may change if the torsional parameters of the used force fields or some non-bonded force field force parameters are adjusted. Actually, some of the parameters in their force fields do need to be re-optimized since none of them could reproduce the experimentally reversible phase transition.

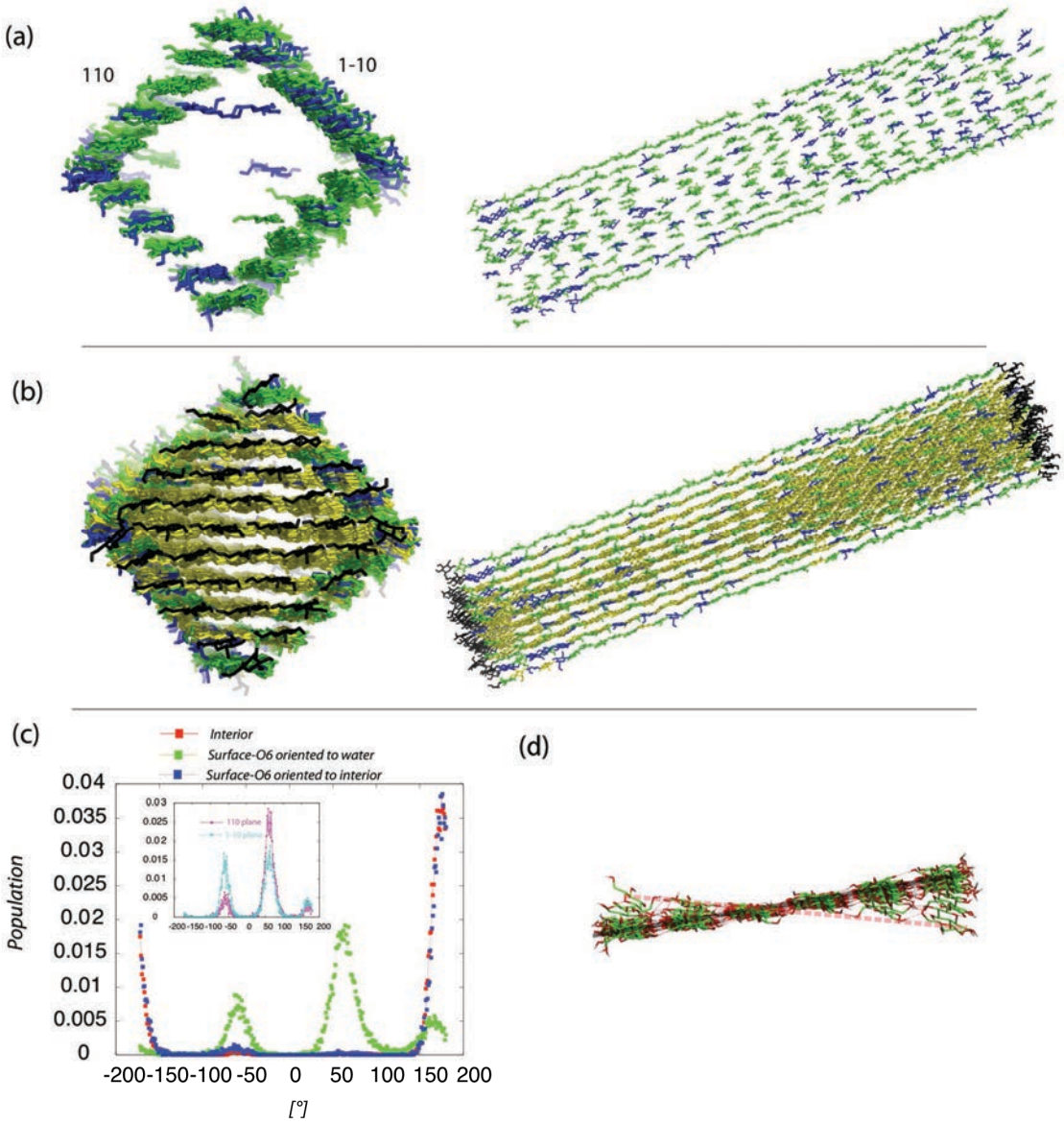

Figure V.4.10. Snapshots of finite models of Iβ at 100 ns by using revised Gr56A$_{carbo}$FF (a and b, yellow: *tg*, blue: *gg*, green: *gt*, terminal groups: black) and the populations of the conformation of ω in different regions of fiber (c).

V.4.3.6.1. Conformation of the hdyroxymethyl groups

The snapshot of Iβ at 100 ns is shown in Figures **V.4.10a** and **V.4.10b**. The core contains almost only *tg* conformation and negligible *gg* and *gt* conformation. The periphery contains a mix of three conformations, in which *gg* and *gt* are dominant (**Table V.4.4** and **Figure V.4.10**). In total, the population of *gt* is larger than that of *gg* in the surface. However, the relative amount of them is different between 110 and ($1\bar{1}0$) planes. More *gt* conformation is present in (

$1\bar{1}0$) plane. In fact, MD simulations with semi-finite models by using other force fields all give the identical result, as shown in **Table V.4.4**.

Table V.4.4. Population of the three-staggered conformation of ω of Iβ in (110) and ($1\bar{1}0$) plane by using GROMOS force fields. Results are calculated from the last 50 ns of the trajectory file. Fours chains in the middle of each surface of the 36-chain models are considered. The terminal groups of the finite models are ignored.

Model type	Force field	110			$1\bar{1}0$		
		gt	*gg*	*tg*	*gt*	*gg*	*tg*
Semi-finite	Revised Gr53A6FF	48%	16%	36%	22%	32%	46%
model	Native Gr56A$_{carbo}$FF	72%	22%	4%	46%	47%	7%
	Revised Gr56A$_{carbo}$FF	72%	18%	9%	53%	33%	14%
Finite model	Revised Gr56A$_{carbo}$FF	78%	14%	8%	53%	37%	10%

V.4.3.6.2. Location of the hydrogen bond disorder in Iβ

In the hydrogen-bonding pattern B, the HO2 oriented to O6 to form an inter-chain HBs and HO6 oriented to O2 to form intra-chain HBs. This pattern was proposed to be located in the surface or non-crystalline domain (Nishiyama *et al.* 2008). From the simulation, the HO2 in the surface resembles that of pattern B. However, this is not the case of the surface HO6. A few inter-chain O2-HO2---O6 HBs and intra-chain O6-HO6---O2 HBs can be found between the surface and core region (**Figure V.4.11a**). The intra-chain O6-HO6---O2 HBs can also be found in the surface (**Figure V.4.11b**). However, their occurrence was much less than the experimental ones (~20%). Furthermore, the inter-chain and intra-chain HBs are not cooperated but separately distributed outside of the core region. Differing from the hydroxymethyl group in the core, the conformation of ω in the surface transformed more rapidly. In addition, the previous simulation suggested that the cooperation of the pattern B is not stable in the core region of single fiber.

V.4.4. Comparison of unit cell parameters of cellulose I for the infinite, semi-finite and finite models

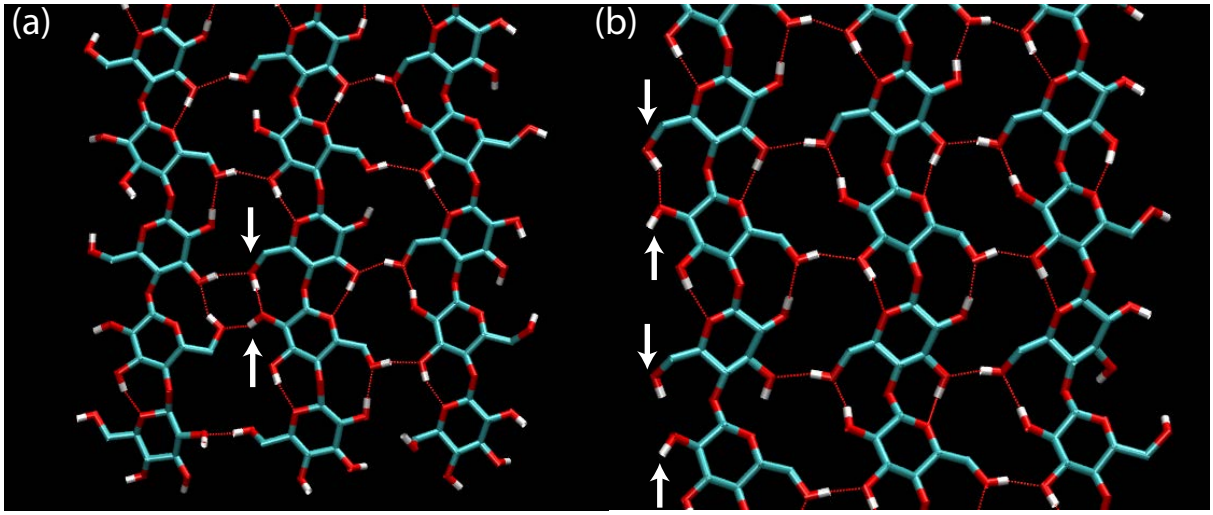

Figure V.4.11. Snapshots of the HB disorder in I_β at 100 ns: a) terminal part of the fiber, b) middle part of the fiber.

Table V.4.5. The unit cell parameters of the semi-finite models, finite models and infinite models of cellulose I calculated from the last 10 ns of the 100 ns trajectory file. Only the models that did not show deformation during MD were shown here.

			a (nm)	*b (nm)*	*c (nm)*	*α (°)*	*β (°)*	*γ (°)*
Finite Models	Iβ	*Revised 56A$_{carbo}$*	0.7762(40)	0.8139(27)	1.035(2)	89.86(1.5)	89.95(1.7)	93.67(2.1)
Semi-finite	Iβ	*Revised 56A$_{carbo}$*	0.7890(5)	0.8079(3)	1.031(2)	89.92(1.3)	90.17(1.9)	93.80(2.1)
models	Iβ	*Native 56A$_{carbo}$*	0.8451(3)	0.8159(3)	1.038(0)	89.78(1.4)	89.60(1.9)	93.29(2.3)
	Iβ	*Revised 53a6*	0.7815(6)	0.8148(3)	1.034(2)	89.86(1.4)	90.22(2.1)	94.41(2.7)
	Iα	*Revised 53a6*	0.9502(4)	0.8139(3)	1.036(2)	89.85(1.3)	123.8(2.1)	94.92(2.1)
	Iα	*Native 53a6*	1.003(5)	0.8165(3)	1.049(2)	89.52(1.3)	121.9(2.3)	90.66(2.2)
Infinite	Iβ	*Revised 56A$_{carbo}$*	0.7839(3)	0.8075(2)	1.035(0)	89.99(3)	90.00(3)	93.50(2)
models	Iβ	*Native 56A$_{carbo}$*	0.8425(4)	0.8137(2)	1.043(0)	89.99(3)	90.00(3)	92.88(3)
	Iβ	*Revised 53a6*	0.7829(5)	0.8164(3)	1.043(2)	90.00(3)	90.00(3)	95.19(4)
	Iα	*Revised 53a6*	0.9553(5)	0.8167(1)	1.044(0)	89.88(0)	124.6(0)	93.56(3)
	Iα	*Native 53a6*	0.9918(4)	0.8164(1)	1.054(0)	89.88(2)	122.30(3)	90.57(5)

Table V.4.5 shows the unit cell parameters of cellulose I for different model types. The fluctuations from the semi-finite and finite models are larger than those from infinite models. By using revised Gr56A$_{carbo}$FF for Iβ, the estimated unit cell parameter *b* from finite models is about 0.005 nm more than that from the infinite model and semi-finite model. The simulated

c-axis from semi-finite models is all slightly smaller than that from infinite models. The *c*-axis from finite model of Iβ is the same with that from the infinite model. In general, all the force fields gave very similar results among the infinite, semi-finite and finite model.

V.4.5. Conclusion

The crystal structures of native cellulose in water are very sensitive to the used force fields. The revised Gr53a6FF, native and revised Gr56A$_{carbo}$FF could give a stable structure for Iβ and the native Gr53a6FF could not, while the native and revised Gr53a6FF could give a stable structure for Iα and those of Gr56A$_{carbo}$FF could not. The revised Gr56A$_{carbo}$FF gives a stable structure of Iα with the 40-chains model but a deformed one with the 64-chains or 36-chains model.

The finite microfibril of cellulose Iβ is stable during 100 ns MD simulation using revised Gr56A$_{carbo}$FF and have no tendency to deform. The structure is very sensitive to the torsional parameters and non-bonded force field parameters, which largely determined the conformation of the hydroxymethyl group in the core.

The infinite, semi-finite and finite models of cellulose I from simulation give identical unit cell parameters, which is close to experimental data, when the structure is stable by using selected force field.

With the diagonal model of cellulose I (the 36-chains model), the ratio between the population of *gt* and *gg* conformation on the low dense plane ((110) of Iβ or (010) of Iα) is larger than that in the high dense surface plane ((1-10) of Iβ or (100) of Iα). This result is also force field independent.

The occurrence of the hydrogen pattern B of native cellulose found in the surface is not as much as the experimentally observed amount, probably due to the designed models that only considered one single microfibril and thus ignoring the surface-surface interaction between fibrils.

None of the GROMOS force fields could well reproduce the experimental structure of cellulose II and III$_I$ with the semi-finite model. The revised Gr56A$_{carbo}$FF suggested that another structure of cellulose II, which is obtained by cooling and mainly different in hydrogen-bonding pattern of 020 plane, is more stable than the experimental one, in the case of both finite and infinite models.

Conclusion

During this PhD work, we have studied the structure and physical properties of cellulose by atomistic calculations, using GROMOS force fields and their modifications using the GROMACS package. The aim of the study was to try to understand the basic interactions and thermodynamics governing the structures of different crystalline allomorphs of cellulose for which precise and reliable experimental data was available. During this study, we have identified crucial force field parameters that had to be carefully adjusted to reproduce experimental results, such as molecular conformation, lattice parameters and phase transition.

Cellulose is composed of fairly rigid pyranose rings linked by glycosidic linkages. Each residue has a labile exocyclic hydroxymethyl group that can take three staggered positions, namely *tg*, *gg* or *gt*. Native cellulose is known to adopt the *tg* conformation, which is very rare in the glucose-based structures. In many force fields, the structure tends to deviate from *tg* conformation during the simulation. This conformation is crucial in maintaining the hydrogen-bonding pattern of native cellulose and thus its resulting physical properties. We have carefully analyzed the dihedral energy term describing the rotation of the hydroxymethyl group and systematically varied its parameters to see their effect on thermal behavior of simulated native cellulose. We have found that the experimentally reversible transition could be observed if one parameter accounting for the energetic penalty of *tg* conformation embedded in the force field was reduced below a critical value and that this transition temperature varied linearly with this parameter value. The slope of this linear variation suggested an entropic gain of 26 kJ mol^{-1} K^{-1} compensating the enthalpy penalty of deviating from the original structure at ambient conditions. The increase in calculated torsional entropy from the trajectory associated with the transition summed to the same value of 26 kJ mol^{-1} K^{-1}. We have thus concluded that the reversible phase transition occurring experimentally at 220°C for cellulose Iβ could be explained in terms of classical thermodynamics where the increase in enthalpy was compensated by the gain in torsional entropy in the high temperature phase.

Another aspect we have focused on was the discrepancy between the experimental unit cell parameters and the simulated ones. In the literature, even within the same family of force fields, the unit cell parameters significantly varied depending on the version. Using the last version of GROMOS, the most significant deviation was observed for the *a*-parameter of native cellulose,

which was longer by more than 8% compared to experimental value, while an older version gave a much better prediction. In addition, the monoclinic angle was close to 90° compared to the 96.5° in the experimental unit cell. The *a*-axis is the direction along which flat hydrogen-bonded sheets of cellulose are stacked by weak van der Waals interactions. We have analyzed the parameters that should be responsible for this discrepancy, and found that the Lennard-Jones parameters that approximates the van der Waals interaction significantly changed among different versions. In particular, the repulsion term had little theoretical justification and was just defined to fit different properties. One of the repulsion term values of hydrocarbon in GROMOS was defined too high to model compact crystal structures, as many of the non-bonded carbon-carbon distances in the Cambridge Structural Database would have prohibitively high energy. Moreover, the partial charges that are directly related to the Coulomb interaction varied among different force fields, depending on the charge assignment strategy. By choosing the right set of partial charges and refining one Lennard-Jones repulsion term of hydrocarbon, we have significantly improved the prediction of the unit cell parameters of all four allomorphs of cellulose.

Having identified some of the critical parameters in GROMOS and optimized them to model crystalline cellulose, we carried out a series of exploratory studies on aspects where experimental approaches are powerless. For example, as we could not define an experimental method to deduce the molecular motion during the allomorphic conversions of cellulose, we have simulated the conversion between cellulose allomorphs by thermal treatment, such as $I_\alpha \rightarrow I_\beta$ and $III_I \rightarrow I_\beta$. We have also modeled interfaces of cellulose crystals with water as well as the hydrogen-bonding disorder.

The experimental $I_\alpha \rightarrow I_\beta$ conversion is generally achieved in three steps. First, I_α is heated above its phase transition to reach the high temperature phase. Second, the chains in the high temperature phase rearrange into a monoclinic organization Lastly, upon cooling, the high temperature phase converts to the I_β structure. We expected that the revised versions of GROMOS force fields that reproduce the reversibility of the phase transition of I_β, should reproduce the $I_\alpha \rightarrow I_\beta$ conversion. However, in the simulation conditions we tested, none of the force fields could reproduce all three steps. Our revised version especially failed in the chain rearrangement at high temperature. It was surprising that the native $Gr56A_{carbo}FF$ could sometimes produce a chain translation at high temperature when MD simulations were performed on a cluster computer. However, we have not yet understood how the computer architecture had an influence on the results, but a detailed analysis of the trajectory at the

moment of transition would give some hints on the necessary conditions to model this transition. It should be also noted that this work was done by using a periodic boundary box of about 4 nm that constitutes a strong constraint on the system. Another unsolved issue is the transition point to go to the high temperature phase: the transition temperature is experimentally higher for $I\alpha$ than $I\beta$ but all simulations we carried out showed the contrary.

The revised Gr56A$_{carbo}$FF could best model the structure of cellulose III_I among all the force fields we used when using infinite models, although the MD simulation from it showed that the unit cell parameter *c*-axis was about 3% smaller than the experimental values and hydrogen-to-acceptor distance of HO2…O6 was 1.6 Å, too short for an alcohol. When the III_I model was heated to 600 K, the chains re-arranged to form a layered structure with intra-layer hydrogen bonds. Although this hydrogen-bonded layer structure is conceptually similar to that in Iβ structure, the translation along the chain direction did not occur to form the staggered structure. In addition, all the hydroxymethyl groups remained *gg* or *gt* instead of adopting *tg* when the system was cooled down to 300 K. However, this new structure had higher energy than $I\beta$, and thus the fact that we did not observe transformation to $I\beta$ might be purely a kinetic problem similar to the problem we faced for $I\alpha \rightarrow I\beta$ conversion.

To account for the finite size of the crystals, we have modeled systems where a small crystal was placed at the center of water-filled periodic boundary box (finite-model), or where the chain goes through a periodic boundary but the crystal model is laterally finite, separated by water molecules across the periodic boundary (2D-finite).

While studying the thermal behavior of cellulose II using revised Gr56A$_{carbo}$FF using infinite and semi-finite approaches, we found that an alternative structure to the current neutron structure was more stable in both cases. This alternative structure has the backbone similar to the experimental structure but differs in the hydrogen-bonding pattern involving O2 and O6. DFT calculations are being carried out to evaluate the energy of these two structures for comparing their stability. Another way to check is to calculate its neutron diffraction pattern directly compared to the experimental one.

Neutron diffraction refinement studies suggested that the hydrogen bonding system of native cellulose was not uniquely determined in the whole system but disordered. The major hydrogen bonding pattern A and another alternative pattern B had been proposed. Thus, we simulated the effect of the putative hydrogen bonding disorder in native cellulose by using revised Gr56A$_{carbo}$FF with infinite, semi-finite and finite models. HB pattern A in $I\beta$ is stable in all the models. When initial structure has pattern B, the related hydroxymethyl groups oscillate between the three staggered conformations. We visually detected the occurrence at the surface

although the quantity seems to be much less than the proposed occupancy of pattern B in the literature. In the infinite model, the pattern B of Iβ converted to pattern A when heated up to 490 K using revised Gr56A$_{carbo}$FF.

For infinite Iα models, irrespective of the starting hydrogen bonding scheme, pattern A or B, an alternative structure that is more enthalpically favorable (0.66±0.58 kJ mol^{-1}per residue lower than pattern A structure) was reached in the simulation. In the case of pattern A, the structure was stable at room temperature during 10 ns, but converted to a new structure by annealing at 600 K. The pattern B readily went to the same new structure at 300K. This structure is characterized by alternated *tg* and *gt* hydroxymethyl conformation along each chain. This resulted in average to 1.5 intra-chain and 1.5 inter-chain hydrogen bonds per glucosyl residue.

Several attempts were made to model cellulose Iα using semi-finite and finite models using the revised Gr56A$_{carbo}$FF. For the semi-finite model, we considered two types of model: a 36-chains model with 110 and 1-10 planes exposed to the environment and a 40-chains model with 200 and 010 planes. The former readily deformed drastically to a new structure whereas the latter was stable at 300 K for 100 ns. The finite model with the former morphology underwent a similar deformation to the semi-finite model.

The conformation of the hydroxymethyl group mainly adopts *gg* and *gt* conformation in the surface of microfibril of Iβ with the 36-chains model with 110 and $1\bar{1}0$ surfaces, which is in agreement with the previous prediction from other force fields. The population of *gt* conformation was higher in the less dense surface (110 of Iβ and 010 of Iα) than that in the denser surface ($1\bar{1}0$ of Iβ and 100 of Iα) with all force fields employed in this study.

During my work, most of my time was spent on chasing the origin of the discrepancy between experimental results and simulation in the force field parameters. We could fix some parameters that were aberrant in the original force field, but reproducing the crystal structures and their behaviors at both room temperature and high temperature with different types of models (infinite, semi-finite and finite) turned out to be still difficult. We still do not have an explanation to the unexpected behaviors of different allomorphs, and a careful analysis of the trajectories obtained in the exploratory studies has to be carried out. Another approach we have not tried in this study is the density functional theory (DFT) analysis that became more and more popular during my PhD work and showed good agreements with experimental crystal structures. DFT itself cannot provide information related to dynamics of molecule but would help the assignment of charge distributions that turned out to be essential in the dense crystalline phase.

References

Alder BJ, Wainwright TE (1957) Phase transition for a hard-sphere system. J Chem Phys 27:1208

Alemdar A, Sain M (2008) Isolation and characterization of nanofibers from agricultural residues - Wheat straw and soy hulls. Bioresource Technol 99:1664-1671

Atalla RH, Vanderhart DL (1984) Native cellulose: A composite of two distinct crystalline forms. Science 223:283-285

Baker AA, Helbert W, Sugiyama J, Miles MJ (2000) New insight into cellulose structure by atomic force microscopy shows the Iα crystal phase at near-atomic resolution. Biophys J 79:1139-1145

Barry AJ, Peterson FC, King AJ (1936) X-ray studies of reactions of cellulose in non-aqueous systems. I.interaction of cellulose and liquid ammonia. J Am Chem Soc 58:333-337

Bellesia G, Chundawat SPS, Langan P, Dale BE, Gnanakaran S (2011) Probing the early events associated with liquid ammonia pretreatment of native crystalline cellulose. J Phys Chem B 115:9782-9788

Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) Interaction models for water in relation to protein hydration. Intermolecular Forces 331-342

Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684-3690

Bergenstråhle M, Berglund LA, Mazeau K (2007) Thermal Response in crystalline Iβ cellulose: a molecular dynamics study. J Phys Chem B 111:9138-9145

Bergenstråhle M, Mazeau K, Berglund LA (2008) Molecular modeling of interfaces between cellulose crystals and surrounding molecules: Effects of caprolactone surface grafting. Eur Polym J 44:3662-3669

Bergenstråhle M, Thormann E, Nordgren N, Berglund LA (2009) Force pulling of single cellulose chains at the crystalline cellulose-liquid interface: a molecular dynamics study. Langmuir 25:4635-4642

Bock K, Duus JØ (1994) A conformational study of hydroxymethyl groups in carbohydrates investigated by ^{1}H NMR spectroscopy. J Carbohydr Chem 13:513-543

Brochier-Salon M-C, Morin C (2000) Conformational analysis of 6-deoxy-6-iodo-D-glucose in aqueous solution. Magnet Reson Chem 38:1041-1042

Brown JW, Wladkowski BD (1996) Ab initio studies of the exocyclic hydroxymethyl rotational surface in α-D-glucopyranose. J Am Chem Soc 118:1190-1193

Bučko T, Tunega D, Ángyán JG, Hafner J (2011) Ab initio study of structure and interconversion of native cellulose phases. J Phys Chem A 115:10097-10105

Bussi G, Parrinello M (2007) Canonical sampling through velocity rescaling. J Chem Phys 126:014101-014107

Case DA, Cheatham III TE, Darden T, Gohlke H, Luo R, Merz KM, Jr., Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The amber biomolecular simulation programs. J Comput Chem 26:1668-1688

Chanzy H, Henrissat B, Vuong R, Revol JF (1986) Structural changes of cellulose crystals during the reversible transformation cellulose I ⇆ III_I in Valonia. Holzforschung 40:25-30

Chen P, Nishiyama Y, Mazeau K (2012) Torsional entropy at the origin of the reversible temperature-induced phase transition of cellulose. Macromolecules 45:362-368

Chundawat SPS, Bellesia G, Uppugundla N, Sousa LD, Gao DH, Cheh AM, Agarwal UP, Bianchetti CM, Phillips GN, Langan P, Balan V, Gnanakaran S, Dale BE (2011) Restructuring the crystalline cellulose hydrogen bond network enhances its depolymerization rate. J Am Chem Soc 133:11163-11174

Cremer D, Pople JA (1975) General definition of ring puckering coordinates. J Am Chem Soc 97:1354-1358

Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. J Chem Phys 98:10089-10092

Debzi EM, Chanzy H, Sugiyama J, Tekely P, Excoffier G (1991) The Iα → Iβ transformation of highly crystalline cellulose by annealing in various mediums. Macromolecules 24:6816-6822

Dunfield LG, Burgess AW, Scheraga HA (1978) Energy parameters in polypeptides. 8. Empirical potential energy algorithm for the conformational analysis of large molecules. J Phys Chem 82:2609-2616

Eichhorn SJ, Davies GR (2006) Modelling the crystalline deformation of native and regenerated cellulose. Cellulose 13:291-307

Eichhorn SJ, Young RJ (2001) The Young's modulus of a microcrystalline cellulose. Cellulose 8:197-207

Eichhorn SJ, Young RJ, Davies GR (2005) Modeling crystal and molecular deformation in regenerated cellulose fibers. Biomacromolecules 6:507-513

Foley BL, Tessier MB, Woods RJ (2011) Carbohydrate force fields. WIREs Comput. Mol. Sci 2: 652-697

French AD (2012) In "Advances in carbohydrate chemistry and biochemistry", Chapter 2: Combining computational chemistry and crystallography for a better understanding of the structure of cellulose 67: 19-93

French AD, Howley PS (1989) In "Cellulose and Wood - Chemistry and Technology", Comparisons of structures proposed for cellulose, C. Schuerch Ed, Wiley, New York, pp 159–167

French AD, Johnson GP (2006) Quantum mechanics studies of cellobiose conformations. Can J Chem 84:603-612

Gasteiger J, Marsili M (1980) Iterative partial equalization of orbital electronegativity: a rapid access to atomic charges. Tetrahedron 36:3219-3228

Gross AS, Bell AT, 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

Gross AS, 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

Guvench O, Hatcher E, Venable RM, Pastor RW, MacKerell AD (2009) CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. J Chem Theory Comput 5:2353-2370

Hansen HS, Hünenberger PH (2011) A reoptimized GROMOS force field for hexopyranose-based carbohydrates accounting for the relative free energies of ring conformers, anomers, epimers, hydroxymethyl rotamers, and glycosidic linkage conformers. J Comput Chem 32:998-1032

Hardy BJ, Sarko A (1996) Molecular dynamics simulations and diffraction-based analysis of the native cellulose fiber: structural modeling of the Iα and Iβ phases and their interconversion. Polymer 37:1833-1839

Heiner AP, Sugiyama J, Teleman O (1995) Crystalline cellulose Iα and Iβ studied by molecular dynamics simulation. Carbohydr Res 273:207-223

Heinz H, Lin TJ, Kishore Mishra R, Emami FS (2012) Thermodynamically consistent force fields for the assembly of inorganic, organic, and biological nanostructures: The INTERFACE force field. Langmuir 29:1754-1765

Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. J Comput Chem 18:1463-1472

Hess B, Kutzner C, Van Der Spoel, D., Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4:435-447

Hidaka H, Kim UJ, Wada M (2010) Synchrotron X-ray fiber diffraction study on the thermal expansion behavior of cellulose crystals in tension wood of Japanese poplar in the low-temperature region. Holzforschung 64:167-171

Hockney RW, Goel SP, Easwd JW (1974) Quiet high-resolution computer models of a plasma. J Comput Phys 14:148-158

Hori R, Wada M (2005) The thermal expansion of wood cellulose crystals. Cellulose 12:479-484

Hori R, Wada M (2006) The thermal expansion of cellulose II and III_{II} crystals. Cellulose 13:281-290

Horii F, Yamamoto H, Kitamaru R, Tanahashi M, Higuchi T (1987) Transformation of native cellulose crystals induced by saturated steam at high temperatures. Macromolecules 20:2946-2949

Horikawa Y, Sugiyama J (2009) Localization of crystalline allomorphs in cellulose microfibril. Biomacromolecules 10:2235-2239

Horta BAC, Fuchs PFJ, van Gunsteren WF, Hünenberger PH (2011) New interaction parameters for oxygen compounds in the GROMOS force field: Improved pure-liquid and solvation properties for alcohols, ethers, aldehydes, ketones, carboxylic acids, and esters. J Chem Theory Comput 7:1016-1031

Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. J Mol Graph 14:33-38

Imai T, Sugiyama J (1998) Nanodomains of Iα and Iβ cellulose in algal microfibrils. Macromolecules 31:6275-6279

Jarvis MC (2000) Interconversion of the Iα and Iβ crystalline forms of cellulose by bending. Carbohydr Res 325:150-154

Jorgensen WL, Tirado-Rives J (2005) Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. P Natl Acad Sci USA 102:6665-6670

Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, Woods RJ (2008) GLYCAM06: A generalizable biomolecular force field. Carbohydrates. J Comput Chem 29:622-655

Koehler JEH, Saenger W, Gunsteren WF (1987) A molecular dynamics simulation of crystalline Œ±-cyclodextrin hexahydrate. Eur Biophys J 15:197-210

Kroon-Batenburg LMJ, Kroon J (1997) The crystal and molecular structures of cellulose I and II. Glycoconjugate J 14:677-690

Kroon-Batenburg LMK, Bouma B, Kroon J (1996) Stability of cellulose structures studied by MD simulations: Could mercerized cellulose II be parallel. Macromolecules 29:5695-5699

Langan P, Nishiyama Y, Chanzy H (1999) A revised structure and hydrogen-bonding system in cellulose II from a neutron fiber diffraction analysis. J Am Chem Soc 121:9940-9946

Langan P, Nishiyama Y, Chanzy H (2001) X-ray structure of mercerized cellulose II at 1 Å resolution. Biomacromolecules 2:410-416

Langan P, Sukumar N, Nishiyama Y, Chanzy H (2005) Synchrotron X-ray structures of cellulose Iβ and regenerated cellulose II at ambient temperature and 100 K. Cellulose 12:551-562

Li D-W, Brüschweiler R (2009) In silico relationship between configurational entropy and soft degrees of freedom in proteins and peptides. Phys Rev Lett 102:118108

Lins RD, Hünenberger PH (2005) A new GROMOS force field for hexopyranose-based carbohydrates. J Comput Chem 26:1400-1412

Mackerell AD (2004) Empirical force fields for biological macromolecules: Overview and issues. J Comput Chem 25:1584-1604

Matthews J, Himmel M, Crowley M (2012a) Conversion of cellulose Iα to Iβ via a high temperature intermediate (I-HT) and other cellulose phase transformations. Cellulose 19:297-306

Matthews JF, Beckham GT, Bergenstra M, Brady JW, Himmel ME, Crowley MF (2012) Comparison of cellulose Iβ simulations with three carbohydrate force fields. J Chem Theory Comput 8:735-748

Matthews JF, Bergenstrahle M, Beckham GT, Himmel ME, Nimlos MR, Brady JW, Crowley MF (2011) High-temperature behavior of cellulose I. J Phys Chem B 115:2155-2166

Matthews JF, Skopec CE, Mason PE, Zuccato P, Torget RW, Sugiyama J, Himmel ME, Brady JW (2006) Computer simulation studies of microcrystalline cellulose Iβ. Carbohydr Res 341:138-152

Mazeau K (2005) Structural Micro-heterogeneities of crystalline Iβ-cellulose. Cellulose 12:339-349

Mazeau K, Rivet A (2008) Wetting the (110) and (100) surfaces of Iβ cellulose studied by molecular dynamics. Biomacromolecules 9:1352-1354

Moon RJ, Martini A, Nairn J, Simonsen J, Youngblood J (2011) Cellulose nanomaterials review: structure, properties and nanocomposites. Chem Soc Rev 40:3941-3994

Nelder JA, Mead R (1965) A simplex method for function minimization. Comput J 7:308-313

Neyertz S, Pizzi A, Merlin A, Maigret B, Brown D, Deglise X (2000) A new all-atom force field for crystalline cellulose I. J Appl Polym Sci 78:1939-1946

Nishida Y, Hori H, Ohrui H, Meguro H (1988) ^{1}H NMR Analyses of rotameric distribution of C5-C6 bonds of D-glucopyranoses in solution. J Carbohydr Chem 7:239-250

Nishiyama Y (2009) Structure and properties of the cellulose microfibril. J Wood Sci 55:241-249

Nishiyama Y, Johnson GP, French AD, Forsyth VT, Langan P (2008) Neutron crystallography, molecular dynamics, and quantum mechanics studies of the nature of hydrogen bonding in cellulose Iβ. Biomacromolecules 9:3133-3140

Nishiyama Y, Langan P, Chanzy H (2002) Crystal structure and hydrogen-bonding system in cellulose Iβ from synchrotron X-ray and neutron fiber diffraction. J Am Chem Soc 124:9074-9082

Nishiyama Y, Sugiyama J, Chanzy H, Langan P (2003) Crystal structure and hydrogen bonding system in cellulose Iα from synchrotron X-ray and neutron fiber diffraction. J Am Chem Soc 125:14300-14306

Oostenbrink C, Villa A, Mark AE, Gunsteren WFV (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field Parameter sets 53A5 and 53A6. J Comput Chem 25:1656-1676

Saito T, Hirota M, Tamura N, Kimura S, Fukuzumi H, Heux L, Isogai A (2009) Individualization of nano-Sized plant cellulose fibrils by direct surface carboxylation using TEMPO catalyst under neutral conditions. Biomacromolecules 10:1992-1996

Schuler LD, Daura X, van Gunsteren WF (2001) An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. J Comput Chem 22:1205-1218

Sugiyama J, Vuong R, Chanzy H (1991) Electron diffraction study on the two crystalline phases occurring in native cellulose from an algal cell wall. Macromolecules 24:4168-4175

Takahashi M, Takenaka H (1982) X-ray study of thethermal expansion and transition of crystalline cellulose. Polym J 14:675-679

Thibaudeau C, Stenutz R, Hertz B, Klepach T, Zhao S, Wu Q, Carmichael I, Serianni AS (2004) Correlated C-C and C-O bond conformations in saccharide hydroxymethyl groups: Parametrization and application of redundant ^{1}H-^{1}H, ^{13}C-^{1}H, and ^{13}C-^{13}C NMR

J-Couplings. J Am Chem Soc 126:15668-15685

van Gunsteren WF, Bakowies D, Baron R, Chandrasekhar I, Christen M, Daura X, Gee P, Geerke DP, Glättli A, Hünenberger PH, Kastenholz MA, Oostenbrink C, Schenk M, Trzesniak D, van der Vegt NFA, Yu HB (2006) Biomolecular modeling: goals, problems, perspectives. Angew Chem Int Edit 45:4064-4092

Viëtor RJ, Mazeau K, Lakin M, Pérez S (2000) A priori crystal structure prediction of native celluloses. Biopolymers 54:342-354

Viëtor RJ, Newman RH, Ha M-A, Apperley DC, Jarvis MC (2002) Conformational features of crystal-surface cellulose from higher plants. Plant J 30:721-731

Wada M (2001) In situ observation of the crystalline transformation from cellulose III_I to Iβ. Macromolecules 34:3271-3275

Wada M (2002) Lateral thermal expansion of cellulose Iβ and III_I polymorphs. J Polym Sci, Part B: Polym Phys 40:1095-1102

Wada M, Chanzy H, Nishiyama Y, Langan P (2004) Cellulose III_I crystal structure and hydrogen bonding by synchrotron X-ray and neutron fiber diffraction. Macromolecules 37:8548-8555

Wada M, Hori R, Kim U-J, Sasaki S (2010) X-ray diffraction study on the thermal expansion behavior of cellulose Iβ and its high-temperature phase. Polym Degrad Stabil 95:1330-1334

Wada M, Kondo T, Okano T (2003) Thermally induced crystal transformation from cellulose Iα to Iβ. Polym J 35:155-159

Wada M, Nishiyama Y, Langan P (2006) X-ray structure of ammonia-cellulose I: New insights into the conversion of cellulose I to cellulose III. Macromolecules 39:2947-2952

Wada M, Okano T (2001) Localization of Iα and Iβ phases in algal cellulose revealed by acid treatments. Cellulose 8:183-188

Watanabe A, Morita S, Ozaki Y (2006) Study on temperature-dependent changes in hydrogen bonds in cellulose Iβ by infrared spectroscopy with perturbation-correlation moving-window two-dimensional correlation spectroscopy. Biomacromolecules 7:3164-3170

Watanabe A, Morita S, Ozaki Y (2007) Temperature-dependent changes in hydrogen bonds in cellulose Iα studied by infrared spectroscopy in combination with perturbation-correlation moving-window two-dimensional correlation spectroscopy: comparison with cellulose Iβ. Biomacromolecules 8:2969-2975

Wohlert J, Bergenstråhle-Wohlert M, Berglund L (2012) Deformation of cellulose nanocrystals: entropy, internal energy and temperature dependence. Cellulose 19:1821-1836

Woods RJ, Dwek RA, Edge CJ, Fraser-Reid B (1995) Molecular mechanical and molecular dynamic simulations of glycoproteins and Oligosaccharides. 1. GLYCAM_93 parameter development. J Phys Chem 99:3832-3846

Yamamoto H, Horii F (1993) CPMAS ^{13}C NMR analysis of the crystal transformation induced for Valonia cellulose by annealing at high temperatures. Macromolecules 26:1313-1317

Yui T, Hayashi S (2007) Molecular dynamics simulations of solvated crystal models of cellulose Iα and III_I. Biomacromolecules 8:817-824

Yui T, Hayashi S (2009) Structural stability of the solvated cellulose III_I crystal models: a molecular dynamics study. Cellulose 16:151-165

Yui T, Nishimura S, Akiba S, Hayashi S (2006) Swelling behavior of the cellulose Iβ crystal models by molecular dynamics. Carbohydr Res 341:2521-2530

Zhang Q, Bulone V, Aagren H, Tu Y (2011) A molecular dynamics study of the thermal response of crystalline cellulose Iβ. Cellulose 18:207-221

Appendix

Tables S1. Optimizing the LJ parameters of the 5 types of atom: CH1, CH2, C1, OA and Or.

Table S1a. Optimized parameters $C6^{1/2}$ in $(kJ\ mol^{-1}\ nm^{6})^{1/2}$ and $C12^{1/2}$ in $(kJ\ mol^{-1}\ nm^{12})^{1/2}$, final values of target function F_{min}.

	CH1		CH2		C1		OA		Or		F_{min}
	$C6^{1/2}$	$C12^{1/2}$	$C6^{1/2}$	$C12^{1/2}$	$C6^{1/2}$	$C12^{1/2}$	$C6^{1/2}$	$C12^{1/2}$	$C6^{1/2}$	$C12^{1/2}$	
Initial	0.0779	0.00985	0.08642	0.005828	0.0779	0.00985	0.04756	0.00110	0.04756	0.00110	329.06
Ex. III	0.1685	0.00925	0.0635	0.00595	0.0782	0.01693	0.0178	0.00098	0.0468	0.00074	11.619
Ex. II	0.0895	0.00578	0.0925	0.00845	0.0882	0.01214	0.0404	0.00126	0.0607	0.00063	5.065
Ex. Iβ	0.1346	0.00864	0.0725	0.00498	0.2754	0.02375	0.0157	0.00093	0.0189	0.00004	2.903
Ex. Iα	0.0811	0.00761	0.0420	0.00595	0.1405	0.01303	0.0478	0.00112	0.0688	0.00030	6.558

Table S1b. Distance σ in nm and energy ε in $kJ\ mol^{-1}$ at the minimum obtained by using the optimized LJ parameters.

	CH1		CH2		C1		OA		Or		F_{min}
	σ	ε	σ	ε	σ	ε	σ	ε	σ	ε	
Initial	0.5019	0.0949	0.4070	0.4105	0.5019	0.0949	0.2849	1.05711	0.2849	1.05711	329.06
Ex. III	0.3801	2.3554	0.4542	0.1148	0.6005	0.0326	0.3804	0.0261	0.2510	2.1901	11.619
Ex. II	0.4013	0.4792	0.4504	0.2564	0.5163	0.1028	0.3147	0.4208	0.2180	8.5880	5.065
Ex. Iβ	0.4003	1.1004	0.4095	0.2787	0.4418	2.5483	0.3897	0.0176	0.1329	16.1780	2.903
Ex. Iα	0.4543	0.1871	0.5215	0.0219	0.4526	0.5744	0.2857	1.0520	0.1641	60.6146	6.558

Table S1c. Predicted unit cell parameters of four cellulose models and by using the optimized L-J parameters.

Crystal		a (nm)	b (nm)	c (nm)	α (°)	β (°)	γ (°)
Cellulose Iβ	Ex. III	0.7918	0.8098	1.040	90.01	90.01	97.37
	Ex. II	0.7828	0.8241	1.038	90.11	89.95	97.59
	Ex. Iβ	0.7865	0.8092	1.020	90.08	89.97	97.50
	Ex. Iα	0.7846	0.8128	1.016	90.00	89.99	96.64
Cellulose Iα	Ex. III	0.681	0.5965	1.041	117.17	114.68	80.31
	Ex. II	0.6657	0.5978	1.038	117.28	114.60	81.62
	Ex. Iβ	0.6673	0.5917	1.022	117.37	114.62	80.32
	Ex. Iα	0.6691	0.5884	1.017	117.09	114.08	81.21
Cellulose II	Ex. III	0.8144	0.8951	1.032	90.03	90.00	117.95
	Ex. II	0.8434	0.8767	1.031	90.03	90.01	118.02
	Ex. Iβ	0.8069	0.8967	1.021	90.03	90.01	117.68
	Ex. Iα	0.8158	0.8927	1.020	90.01	90.00	117.00
Cellulose III	Ex. III	0.4475	0.7918	1.029	90.00	90.02	104.37
	Ex. II	0.4394	0.7908	1.033	90.15	90.04	104.38
	Ex. Iβ	0.4464	0.7901	1.021	89.99	90.00	105.16
	Ex. Iα	0.4494	0.7751	1.019	90.73	89.74	105.14

Table S1d. Deviation (in %) between the predicted unit cell parameters and the experimental ones.

Crystal		a	b	c	α	β	γ
Cellulose Iβ	Ex. III	1.72	-1.26	0.19	0.01	0.01	0.85
	Ex. II	0.57	0.49	0.00	0.12	-0.06	1.08
	Ex. Iβ	1.04	-1.33	-1.73	0.09	-0.03	0.98
	Ex. Iα	0.80	-0.89	-2.12	0.00	-0.01	0.09
Cellulose Iα	Ex. III	1.38	0.05	0.10	-0.77	-0.10	-0.07
	Ex. II	-0.89	0.27	-0.19	-0.68	-0.17	1.56
	Ex. Iβ	-0.66	-0.75	-1.73	-0.60	-0.16	-0.06
	Ex. Iα	-0.39	-1.31	-2.21	-0.84	-0.63	1.05
Cellulose II	Ex. III	1.67	-0.87	0.10	0.03	0.00	0.73
	Ex. II	5.29	-2.91	0.00	0.03	0.01	0.79
	Ex. Iβ	0.74	-0.70	-0.97	0.03	0.01	0.50
	Ex. Iα	1.85	-1.14	-1.07	0.01	0.00	-0.09
Cellulose III	Ex. III	0.56	0.87	-0.19	0.00	0.02	-0.69
	Ex. II	-1.26	0.74	0.19	0.17	0.04	-0.69
	Ex. Iβ	0.31	0.65	-0.97	-0.01	0.00	0.06
	Ex. Iα	0.99	-1.26	-1.16	0.81	-0.29	0.04

Tables S2. Optimizing the LJ parameters of the 4 types of atom: CH1, CH2, C1 and OA.

Table S2a. Optimized parameters: $C6^{1/2}$ in (kJ mol^{-1} nm^{6} $)^{1/2}$ and $C12^{1/2}$ in (kJ mol^{-1} nm^{12} $)^{1/2}$.

	CH1		CH2		C1		OA	
	$C6^{1/2}$	$C12^{1/2}$	$C6^{1/2}$	$C12^{1/2}$	$C6^{1/2}$	$C12^{1/2}$	$C6^{1/2}$	$C12^{1/2}$
Initial	0.0779	0.00985	0.08642	0.005828	0.0779	0.00985	0.04756	0.001100
Ex. III	0.1195	0.005360	0.07309	0.08224	0.07241	0.01403	0.02611	0.0009538
Ex. II	0.07618	0.003335	0.11287	0.009002	0.08473	0.01305	0.04002	0.001149
Ex. Iβ	0.16913	0.008872	0.06809	0.001408	0.03596	0.01395	0.03537	0.0008167
Ex. Iα	0.08635	0.006092	0.08680	0.01106	0.08472	0.01040	0.04948	0.0009362

Table S2b. Distance σ in nm and energy ε in kJ mol^{-1} at the minimum obtained by using the optimized LJ parameters, final values of target function F_{min}.

	CH1		CH2		C1		OA		
	σ	ε	σ	F_{min}	σ	ε	σ	ε	F_{min}
Initial	0.5019	0.0949	0.4070	0.4105	0.5019	0.0949	0.2849	1.05711	329.06
Ex. III	0.3553	1.7745	1.0401	0.0011	0.5787	0.0349	0.3318	0.2161	18.8
Ex. II	0.3524	0.7570	0.4304	0.5007	0.5360	0.0757	0.3062	0.4598	22.0
Ex. Iβ	0.3743	2.5988	0.2745	2.7106	0.7293	0.0021	0.2848	0.5623	9.55
Ex. Iα	0.4132	0.3745	0.5032	0.1160	0.4970	0.1191	0.2665	0.3167	24.1

Table S2c. Predicted unit cell parameters of four cellulose models and by using the optimized LJ parameters.

Crystal		**a (nm)**	**b (nm)**	**c (nm)**	**α (°)**	**β (°)**	**γ (°)**
Cellulose Iβ	Ex. III	0.8006	0.8110	1.046	90.11	89.88	96.55
	Ex. II	0.7939	0.8168	1.047	89.89	90.17	96.16
	Ex. Iβ	0.7143	0.7674	1.023	89.46	90.42	90.86
	Ex. Iα	0.7949	0.8024	1.042	89.98	90.00	94.45
Cellulose Iα	Ex. III	0.6590	0.5972	1.047	116.31	113.85	80.33
	Ex. II	0.6492	0.6031	1.047	116.59	114.82	80.40
	Ex. Iβ	0.6656	0.5991	1.043	116.41	113.69	79.68
	Ex. Iα	0.6594	0.5852	1.041	116.62	113.14	80.07
Cellulose II	Ex. III	0.8157	0.8741	1.037	90.03	90.02	116.23
	Ex. II	0.8031	0.8002	1.019	90.01	89.99	118.49
	Ex. Iβ	0.8120	0.8984	1.037	90.05	90.02	118.30
	Ex. Iα	0.8065	0.8785	1.036	90.02	89.99	116.69
Cellulose III	Ex. III	0.4498	0.7791	1.034	89.93	90.26	103.22
	Ex. II	0.4497	0.7778	1.033	90.03	89.87	103.41
	Ex. Iβ	0.4475	0.7858	1.031	89.94	90.26	103.29
	Ex. Iα	0.4489	0.7772	1.031	90.01	89.98	103.77

Table S2d. Deviation (in %) between the predicted and experimental unit cell parameters.

Crystal		a	b	c	α	β	γ
Cellulose Iβ	Ex. III	2.85	-1.11	0.75	0.12	-0.13	0.00
	Ex. II	1.99	-0.40	0.84	-0.12	0.19	-0.40
	Ex. Iβ	-8.23	-6.43	-1.43	-0.60	0.47	-5.89
	Ex. Iα	2.12	-2.16	0.36	-0.03	0.00	-2.17
Cellulose Iα	Ex. III	-1.89	0.17	0.62	-1.50	-0.83	-0.05
	Ex. II	-3.35	1.16	0.67	-1.26	0.02	0.04
	Ex. Iβ	-0.91	0.49	0.28	-1.42	-0.97	-0.86
	Ex. Iα	-1.83	-1.85	0.11	-1.23	-1.45	-0.37
Cellulose II	Ex. III	1.84	-3.20	0.59	0.03	0.02	-0.74
	Ex. II	0.26	-11.4	-1.13	0.01	-0.01	1.19
	Ex. Iβ	1.37	-0.51	0.56	0.06	0.02	1.02
	Ex. Iα	0.69	-2.71	0.46	0.02	-0.01	-0.35
Cellulose III	Ex. III	1.08	-0.75	0.27	-0.08	0.29	-1.79
	Ex. II	1.06	-0.92	0.19	0.03	-0.14	-1.61
	Ex. Iβ	0.56	0.10	0.01	-0.07	0.29	-1.72
	Ex. Iα	0.88	-0.99	-0.01	0.01	-0.02	-1.27

Table S3. Predicted unit cell parameters of the different crystals of cellulose by transferring the partial atomic charges of different force fields into GROMOS 53a6.

	Force field	*a (nm)*	*b (nm)*	*c (nm)*	α (°)	β (°)	γ (°)	SD_γ
Iβ	CHARMM C35	0.8399(7)	0.8165(2)	1.0518(1)	90.01(3)	89.98(3)	93.27(8)	-3.4
	GLYCAM 06	0.8319(5)	0.8126(2)	1.0510(1)	90.00(3)	89.99(2)	92.74(2)	-3.9
	GROMOS 53a6	0.8384(6)	0.8159(2)	1.0515(1)	90.00(4)	90.00(4)	90.82(7)	-5.9
	GROMOS 56A$_{carbo}$	0.8379(8)	0.8178(2)	1.0503(1)	89.99(4)	90.01(4)	93.09(7)	-3.6
	GROMOS 87	0.8513(5)	0.8247(3)	1.0503(1)	90.05(4)	89.95(5)	93.28(9)	-3.4
	Experiments	0.7784(8)	0.8201(8)	1.038(1)	90.00	90.00	96.55	
Iα	CHARMM C35	0.6614(2)	0.6288(3)	1.0527(1)	115.34(3)	114.04(3)	78.49(3)	-2.3
	GLYCAM 06	0.6574(2)	0.6280(4)	1.0526(1)	115.47(2)	114.38(3)	78.50(3)	-2.3
	GROMOS 53a6	0.6470(3)	0.6387(4)	1.0531(1)	114.47(4)	114.31(4)	78.82(3)	-1.9
	GROMOS 56A$_{carbo}$	0.6647(4)	0.6291(4)	1.0515(1)	115.57(3)	114.25(3)	78.42(3)	-2.4
	GROMOS 87	0.6653(9)	0.6361(13)	1.0508(1)	114.85(6)	113.78(9)	78.66(4)	-2.1
	Experiments	0.6717(7)	0.5962(6)	1.0400(1)	118.08(5)	114.80(5)	80.37(5)	
II	CHARMM C35	0.8028(1)	0.9439(3)	1.0455(1)	90.03(1)	90.02(1)	113.10(1)	-3.4
	GLYCAM 06	0.8046(1)	0.9252(2)	1.0436(1)	90.01(1)	90.00(1)	112.54(1)	-3.9
	GROMOS 53a6	0.8023(1)	0.9477(2)	1.0462(1)	90.01(1)	90.01(1)	113.03(1)	-3.5
	GROMOS 56A$_{carbo}$	0.8045(1)	0.9441(2)	1.0445(1)	90.021(1)	90.01(1)	113.13(1)	-3.4
	GROMOS 87	0.8087(1)	0.9584(2)	1.0463(1)	90.2(1)	90.00(1)	113.29(1)	-3.3
	Experiments	0.801(3)	0.903(3)	1.031(5)	90	90	117.10(5)	
III	CHARMM C35	0.4918(1)	0.7777(1)	1.0446(1)	89.99(1)	90.03(1)	105.94(1)	0.8
	GLYCAM 06	0.4901(1)	0.7728(1)	1.0425(1)	89.64(5)	88.66(15)	105.98(1)	0.8
	GROMOS 53a6	0.4935(1)	0.7786(1)	1.0438(1)	89.99(1)	90.03(1)	105.85(1)	0.7
	GROMOS 56A$_{carbo}$	0.4933(1)	0.7805(1)	1.0430(1)	90.00(1)	90.03(1)	106.10(2)	1.0
	GROMOS 87	0.4888(4)	0.7956(5)	1.0463(1)	90.37(4)	91.96(13)	110.81(12)	5.4
	Experiments	0.4450(4)	0.7850(8)	1.0310(0)	90	90	105.10(5)	

Tables S4. Optimizing the LJ parameters of the 3 types of atom: CH1, C1 and OA.

Table S4a. Optimized parameters: $C6^{1/2}$ in (kJ mol^{-1} nm^{6})$^{1/2}$ and $C12^{1/2}$ in (kJ mol^{-1} nm^{12})$^{1/2}$, final values of target function F_{min}.

	CH1		C1		OA		F_{min}
	$C6^{1/2}$	$C12^{1/2}$	$C6^{1/2}$	$C12^{1/2}$	$C6^{1/2}$	$C12^{1/2}$	
Initial	0.0779	0.00985	0.0779	0.00985	0.04756	0.001100	329.06
Ex. III	0.1392	0.0065989	0.1213	0.0135886	0.0091	0.0009368	20.29
Ex. II	0.1314	0.0058694	0.0982	0.0133463	0.0269	0.0013206	19.84
Ex. Iβ	0.1382	0.0063017	0.1012	0.0138368	0.0067	0.0008681	13.1
Ex. Iα	0.1051	0.0074433	0.0767	0.0107072	0.0503	0.0011994	13.6
Four crystals	0.1337	0.0065698	0.1202	0.0110944	0.0080	0.0008983	27.38

Table S4b. Distance σ in nm and energy ε in kJ mol^{-1} at the minimum obtained by using the optimized LJ parameters.

	CH1		C1		OA	
	σ	E	σ	ε	σ	ε
Initial	0.5019	0.09489	0.5019	0.09489	0.2849	1.05711
Ex. III	0.3619	2.15702	0.4820	0.29336	0.4679	0.00199
Ex. II	0.3548	2.16221	0.5142	0.13047	0.3661	0.07512
Ex. Iβ	0.3572	2.29926	0.5152	0.13694	0.5050	0.00069
Ex. Iα	0.4373	0.32518	0.5178	0.06527	0.2906	1.21149
Four crystals	0.3663	1.84932	0.4520	0.42369	0.4826	0.00126

Table S4c. predicted unit cell parameters of four cellulose models and by using the optimized LJ parameters.

Crystal		a (nm)	b (nm)	c (nm)	α (°)	β (°)	γ (°)
Cellulose Iβ	ex. Iα	0.7754	0.8152	1.048	90.02	89.99	94.28
	ex. Iβ	0.8009	0.8035	1.044	90.19	89.78	95.57
	ex. II	0.7805	0.8264	1.049	90.01	89.99	94.94
	ex. III	0.7943	0.8076	1.046	90.10	89.82	95.67
	four crystals	0.7925	0.8043	1.044	90.07	89.93	95.30
Cellulose Iα	ex. Iα	0.6428	0.6040	1.049	116.50	114.78	81.65
	ex. Iβ	0.6532	0.5971	1.046	116.11	113.94	79.93
	ex. II	0.6553	0.6111	1.051	117.18	115.80	81.53
	ex. III	0.6538	0.5962	1.047	116.17	113.94	80.45
	four crystals	0.6510	0.5925	1.046	116.01	113.62	80.53
Cellulose II	ex. Iα	0.8123	0.8873	1.036	90.01	90.01	116.03
	ex. Iβ	0.8063	0.8942	1.036	90.02	90.01	117.29
	ex. II	0.8405	0.8752	1.037	90.01	90.01	117.77
	ex. III	0.8133	0.8914	1.037	90.02	90.00	117.49
	four crystals	0.8125	0.8919	1.036	90.02	90.01	117.88
Cellulose III	ex. Iα	0.4511	0.7756	1.035	89.81	90.01	105.89
	ex. Iβ	0.4458	0.7834	1.034	90.00	90.02	104.21
	ex. II	0.4350	0.8036	1.038	89.95	90.18	103.69
	ex. III	0.4446	0.7884	1.033	90.01	90.02	104.11
	four crystals	0.4423	0.7865	1.033	90.01	90.01	103.92

Table S4d. Deviation (in %) between the predicted and experimental unit cell parameters.

Crystal		*a*	*b*	*c*	α	β	γ
Cellulose Iβ	ex. Iα	-0.39	-0.60	0.92	0.02	-0.01	-2.35
	ex. Iβ	2.89	-2.02	0.59	0.21	-0.25	-1.01
	ex. II	0.27	0.77	1.01	0.01	-0.01	-1.67
	ex. III	2.04	-1.52	0.72	0.11	-0.20	-0.91
	four crystals	1.81	-1.93	0.59	0.08	-0.08	-1.30
Cellulose Iα	ex. Iα	-4.30	1.31	0.86	-1.34	-0.02	1.59
	ex. Iβ	-2.75	0.15	0.54	-1.67	-0.75	-0.54
	ex. II	-2.44	2.50	1.03	-0.76	0.87	1.44
	ex. III	-2.66	0.00	0.66	-1.62	-0.75	0.10
	four crystals	-3.08	-0.62	0.54	-1.75	-1.03	0.20
Cellulose II	ex. Iα	1.41	-1.74	0.51	0.01	0.01	-0.92
	ex. Iβ	0.66	-0.97	0.50	0.02	0.01	0.16
	ex. II	4.93	-3.08	0.60	0.01	0.01	0.57
	ex. III	1.54	-1.28	0.55	0.02	0.00	0.33
	four crystals	1.44	-1.23	0.52	0.02	0.01	0.66
Cellulose III	ex. Iα	1.37	-1.20	0.36	-0.21	0.01	0.75
	ex. Iβ	0.18	-0.20	0.26	0.00	0.02	-0.85
	ex. II	-2.25	2.37	0.69	-0.06	0.20	-1.34
	ex. III	-0.09	0.43	0.23	0.01	0.02	-0.94
	four crystals	-0.61	0.19	0.18	0.01	0.01	-1.12

Tables S5. Optimizing the LJ parameters of the 2 types of atom: CH1 and OA, using the charge distribution of CHARMM C35.

Table S5a. Optimized parameters: $C6^{1/2}$ in (kJ mol^{-1} nm^{6} $)^{1/2}$ and $C12^{1/2}$ in (kJ mol^{-1} nm^{12} $)^{1/2}$, final values of target function F_{min}.

	CH1		OA		
	$C6^{1/2}$	$C12^{1/2}$	$C6^{1/2}$	$C12^{1/2}$	F_{min}
Initial	0.0779	0.00985	0.0476	0.00110	329.06
ex. III	0.1539	0.00956	0.0140	0.00099	13.71
ex. II	0.1173	0.00801	0.0374	0.00141	16.52
ex. Iβ	0.1094	0.00827	0.0441	0.00127	12.65
ex. Iα	0.1295	0.00894	0.0309	0.00123	20.02

Table S5b. Distance σ in nm and energy ε in kJ mol^{-1} at the minimum obtained by using the optimized LJ parameters.

	CH1		OA	
	σ	ε	σ	ε
Initial	0.5019	0.09489	0.2849	1.05711
ex. III	0.3960	1.53576	0.4138	0.00976
ex. II	0.4087	0.73882	0.3351	0.24684
ex. Iβ	0.4229	0.52314	0.3062	0.59097
ex. Iα	0.4103	0.87935	0.3415	0.15040

Table S5c. Predicted unit cell parameters of four cellulose models and by using the optimized LJ parameters.

Crystal		a (nm)	b (nm)	c (nm)	α (°)	β (°)	γ (°)
Cellulose Iβ	ex. Iα	0.7862	0.8200	1.050	89.95	90.02	94.94
	ex. Iβ	0.7768	0.8161	1.048	89.98	90.00	94.42
	ex. II	0.7830	0.8275	1.051	90.01	89.95	95.06
	ex. III	0.7894	0.8093	1.048	90.02	89.96	96.05
Cellulose Iα	ex. Iα	0.6553	0.6038	1.052	116.64	114.47	81.19
	ex. Iβ	0.6461	0.6037	1.050	116.62	114.76	81.51
	ex. II	0.6578	0.6031	1.053	116.59	114.31	81.90
	ex. III	0.6592	0.5928	1.049	116.98	113.98	80.38
Cellulose II	ex. Iα	0.8254	0.8968	1.038	90.04	90.05	117.13
	ex. Iβ	0.8167	0.8877	1.037	90.03	90.00	116.51
	ex. II	0.8403	0.8843	1.040	90.08	90.05	116.79
	ex. III	0.8164	0.8969	1.038	90.03	90.02	118.16
Cellulose III	ex. Iα	0.4511	0.7923	1.035	89.91	90.00	104.82
	ex. Iβ	0.4496	0.7843	1.034	90.49	89.96	105.22
	ex. II	0.4384	0.7990	1.042	89.67	90.79	105.22
	ex. III	0.4453	0.7885	1.034	90.00	90.02	103.76

Table S5d. Deviation (in %) between the predicted and experimental unit cell parameters.

Crystal		a	b	c	α	β	γ
Cellulose Iβ	ex. Iα	1.00	-0.01	1.16	-0.06	0.03	-1.67
	ex. Iβ	-0.21	-0.49	0.97	-0.02	0.00	-2.21
	ex. II	0.59	0.90	1.28	0.01	-0.05	-1.54
	ex. III	1.41	-1.32	0.93	0.02	-0.05	-0.52
Cellulose Iα	ex. Iα	-2.44	1.27	1.12	-1.22	-0.29	1.02
	ex. Iβ	-3.81	1.26	0.92	-1.24	-0.04	1.41
	ex. II	-2.07	1.16	1.24	-1.26	-0.43	1.90
	ex. III	-1.86	-0.57	0.89	-0.94	-0.71	0.01
Cellulose II	ex. Iα	3.05	-0.69	0.70	0.04	0.06	0.02
	ex. Iβ	1.96	-1.69	0.55	0.03	0.00	-0.50
	ex. II	4.91	-2.07	0.87	0.09	0.06	-0.26
	ex. III	1.92	-0.68	0.71	0.04	0.02	0.90
Cellulose III	ex. Iα	1.37	0.93	0.37	-0.10	0.00	-0.27
	ex. Iβ	1.03	-0.09	0.30	0.54	-0.04	0.11
	ex. II	-1.48	1.78	1.04	-0.37	0.88	0.11
	ex. III	0.07	0.45	0.33	0.00	0.02	-1.27

Tables S6. Optimizing the LJ parameters of the CH1 atom type, using the charge distribution of CHARMM C35.

Table S6a,b. Optimized parameters: $C6^{1/2}$ in (kJ mol^{-1} nm^{6} $)^{1/2}$ and $C12^{1/2}$ in (kJ mol^{-1} nm^{12} $)^{1/2}$, distance σ in nm and energy ε in kJ mol^{-1} at the minimum, deviations (in %) between the optimized parameters and their original values, final values of target function F_{min}.

	$C6^{1/2}$	$C12^{1/2}$	σ	ε	$\frac{C6^{1/2}_{optimized} - C6^{1/2}_{initial}}{C6^{1/2}_{initial}}$	$\frac{C12^{1/2}_{optimized} - C12^{1/2}_{initial}}{C12^{1/2}_{initial}}$	F_{min}	$\frac{C12^{1/2}_{optimized} - C12^{1/2}_{initial}}{C12^{1/2}_{initial}}$
Initial	0.0779	0.00985	0.5019	0.0949				
Iα (ex.)	0.1102	0.00898	0.4334	0.4582	41.5	-8.8	10.3	-8.8
Iβ (ex.)	0.0971	0.00831	0.4407	0.3222	24.6	15.6	19.6	15.6
II (ex.)	0.1080	0.00854	0.4293	0.4655	38.6	13.3	21.3	13.3
III (ex.)	0.0684	0.00626	0.4508	0.1394	-12.2	-36.4	17.9	-36.4
Four crystals	0.0762	0.00661	0.4427	0.1930	-2.1	-32.9	30.4	-32.9

Table S5c. Predicted unit cell parameters of four cellulose models and by using the optimized LJ parameters.

Crystal		a (nm)	b (nm)	c (nm)	α (°)	β (°)	γ (°)
Cellulose Iβ	Ex. Iα	0.7690	0.8132	1.044	90.01	89.98	94.88
	Ex. Iβ	0.7800	0.8144	1.045	89.99	89.99	94.81
	Ex. II	0.7733	0.8136	1.044	89.98	90.00	94.77
	Ex. III	0.7876	0.8153	1.044	90.04	89.95	95.27
	four crystals	0.7803	0.8148	1.044	90.00	89.96	95.33
Cellulose Iα	Ex. Iα	0.6437	0.5984	1.045	116.77	114.79	81.74
	Ex. Iβ	0.6478	0.6011	1.046	116.48	114.50	81.33
	Ex. II	0.6436	0.6010	1.046	116.66	114.75	81.58
	Ex. III	0.6535	0.5955	1.045	116.17	113.61	81.34
	four crystals	0.6504	0.5956	1.045	116.42	114.01	81.55
Cellulose II	Ex. Iα	0.8163	0.8721	1.033	90.08	90.07	115.71
	Ex. Iβ	0.8105	0.8970	1.034	90.00	90.00	116.32
	Ex. II	0.8097	0.8917	1.034	90.01	90.00	116.57
	Ex. III	0.8122	0.9027	1.034	90.02	90.02	116.18
	four crystals	0.8144	0.8903	1.034	90.01	90.01	116.18
Cellulose III	Ex. Iα	0.4541	0.7787	1.029	89.99	90.04	103.45
	Ex. Iβ	0.4584	0.7840	1.032	89.99	90.01	104.80
	Ex. II	0.4522	0.7812	1.032	90.02	90.01	105.11
	Ex. III	0.4609	0.7835	1.032	90.00	90.01	104.71
	four crystals	0.4579	0.7837	1.031	89.99	90.01	104.57

Table S6d. Deviation (in %) between the predicted unit cell parameters and the experimental ones.

Crystal		a	b	c	α	β	γ
Cellulose Iβ	Ex. Iα	-1.21	-0.84	0.55	0.01	-0.02	-1.73
	Ex. Iβ	0.21	-0.70	0.64	-0.01	-0.01	-1.81
	Ex. II	-0.66	-0.79	0.61	-0.02	0.00	-1.85
	Ex. III	1.18	-0.59	0.55	0.04	-0.05	-1.32
	four crystals	0.24	-0.65	0.53	0.00	-0.04	-1.27
Cellulose Iα	Ex. Iα	-4.17	0.37	0.47	-1.11	-0.01	1.71
	Ex. Iβ	-3.56	0.82	0.54	-1.36	-0.26	1.19
	Ex. II	-4.18	0.81	0.53	-1.21	-0.05	1.51
	Ex. III	-2.71	-0.12	0.43	-1.62	-1.03	1.21
	four crystals	-3.17	-0.10	0.43	-1.41	-0.69	1.47
Cellulose II	Ex. Iα	1.91	-3.42	0.22	0.08	0.08	-1.19
	Ex. Iβ	1.19	-0.66	0.29	0.00	0.00	-0.67
	Ex. II	1.09	-1.25	0.26	0.01	0.00	-0.45
	Ex. III	1.40	-0.03	0.24	0.02	0.03	-0.78
	four crystals	1.67	-1.41	0.24	0.02	0.01	-0.79
Cellulose III	Ex. Iα	2.04	-0.80	-0.18	-0.01	0.04	-1.57
	Ex. Iβ	3.01	-0.13	0.05	-0.01	0.01	-0.29
	Ex. II	1.62	-0.48	0.06	0.02	0.01	0.01
	Ex. III	3.57	-0.19	0.07	0.00	0.01	-0.37
	four crystals	2.90	-0.17	-0.01	-0.01	0.01	-0.50

Tables S7. Optimizing only the repulsive LJ parameter of the CH1 atom type, using the charge distribution of Charmm C35.

Table S7a-b. Optimized parameter: $C12^{1/2}$ in (kJ mol^{-1} nm^{12} $)^{1/2}$, distance σ in nm and energy ε in kJ mol^{-1} at the minimum, deviations (in %) between the optimized parameter and its original value, final values of target function F_{min}.

	$C6^{1/2}$	$C12^{1/2}$	σ	E	$\frac{C12^{1/2}_{optimized} - C12^{1/2}_{initial}}{C12^{1/2}_{initial}}$	F_{min}
Initial	0.0779	0.00985	0.5019	0.0949		
Iα	0.0779	0.00659	0.4389	0.2122	-33.1	11
Iβ	0.0779	0.00757	0.4598	0.1606	23.1	2.2
II	0.0779	0.00726	0.4535	0.1745	26.3	3.4
III	0.0779	0.00584	0.4217	0.2699	40.7	0.91
Four crystals	0.0779	0.00673	0.4420	0.2034	31.7	30.38

Table S7c. Predicted unit cell parameters of four cellulose models and by using the optimized LJ parameters.

Crystal		a (nm)	b (nm)	c (nm)	α (°)	β (°)	γ (°)
Cellulose Iβ	Iα	0.7962	0.8162	1.045	89.97	89.97	94.90
	Iβ	0.7778	0.8146	1.043	90.07	89.92	95.27
	II	0.7914	0.8154	1.045	90.03	89.94	95.07
	III	0.7613	0.8131	1.042	89.99	89.97	95.54
	four crystals	0.7803	0.8149	1.044	89.98	90.03	95.27
Cellulose Iα	Iα	0.6551	0.6050	1.046	116.11	114.05	80.66
	Iβ	0.6493	0.5945	1.044	116.44	113.94	81.69
	II	0.6533	0.6013	1.046	116.21	113.98	81.00
	III	0.6460	0.5855	1.043	116.72	113.89	82.50
	four crystals	0.6505	0.5958	1.045	116.45	114.03	81.52
Cellulose II	Iα	0.8106	0.9120	1.036	90.02	90.02	115.44
	Iβ	0.8149	0.8877	1.033	90.02	90.01	116.31
	II	0.8123	0.9037	1.035	90.03	90.01	115.70
	III	0.8148	0.8709	1.033	90.02	90.00	117.17
	four crystals	0.8147	0.8923	1.034	90.01	90.01	116.14
Cellulose III	Iα	0.4664	0.7841	1.034	90.00	90.01	105.10
	Iβ	0.4566	0.7836	1.031	90.00	90.00	104.51
	II	0.4631	0.7837	1.033	90.00	90.01	104.91
	III	0.4480	0.7839	1.028	89.99	90.01	104.04
	four crystals	0.4580	0.7835	1.031	90.00	90.01	104.61

Table S7d. Deviation (in %) between the predicted and experimental unit cell parameters.

Crystal		a	b	c	α	β	γ
Cellulose Iβ	Iα	2.29	-0.48	0.70	-0.04	-0.03	-1.71
	Iβ	-0.08	-0.67	0.50	0.08	-0.09	-1.32
	II	1.67	-0.57	0.65	0.03	-0.07	-1.53
	III	-2.20	-0.85	0.38	-0.01	-0.03	-1.04
	four crystals	0.24	-0.63	0.54	-0.02	0.03	-1.32
Cellulose Iα	Iα	-2.47	1.48	0.62	-1.67	-0.65	0.36
	Iβ	-3.33	-0.29	0.41	-1.39	-0.75	1.64
	II	-2.74	0.86	0.53	-1.59	-0.72	0.79
	III	-3.83	-1.79	0.28	-1.15	-0.80	2.65
	four crystals	-3.16	-0.07	0.44	-1.38	-0.67	1.43
Cellulose II	Iα	1.20	1.00	0.47	0.02	0.02	-1.42
	Iβ	1.74	-1.69	0.22	0.02	0.02	-0.68
	II	1.41	0.08	0.37	0.03	0.01	-1.20
	III	1.72	-3.55	0.16	0.02	0.00	0.06
	four crystals	1.71	-1.18	0.25	0.01	0.01	-0.82
Cellulose III	Iα	4.81	-0.11	0.27	0.00	0.01	0.00
	Iβ	2.61	-0.18	-0.05	0.00	0.00	-0.56
	II	4.07	-0.17	0.16	0.00	0.01	-0.18
	III	0.67	-0.14	-0.29	-0.01	0.01	-1.01
	four crystals	2.92	-0.19	0.00	0.00	0.01	-0.47

Table S7e. Comparison of unit cell parameters predicted by original LJ parameters and optimized repulsive parameter of CH1. The optimized results have been used two different simulation scale: 10 ns and 50 ps, which showed the results are similar.

		a (nm)	b (nm)	c (nm)	α (°)	β (°)	γ (°)
Cellulose Iβ	Exp.	0.7784	0.8201	1.038	90	90	96.55
10 ns	Opt.	0.7789	0.8114	1.045	89.99	90	94.27
50 ps	Opt.	0.7803	0.8149	1.044	89.98	90.03	95.27
	Ori.	0.8384	0.8159	1.051	90	90	90.82
Cellulose Iα	Exp.	0.6717	0.5962	1.040	118.08	114.8	80.37
10 ns	Opt.	0.6452	0.6014	1.047	116.53	114.59	81.16
50 ps	Opt.	0.6505	0.5958	1.045	116.45	114.03	81.52
	Ori.	0.6470	0.6387	1.053	114.47	114.31	78.82
Cellulose II	Exp.	0.8010	0.903	1.031	90	90	117.1
10 ns	Opt.	0.8162	0.8741	1.035	90.03	90.01	114.53
50 ps	Opt.	0.8147	0.8923	1.034	90.01	90.01	116.14
	Ori.	0.8023	0.9477	1.046	90.02	90.01	113.03
Cellulose III	Exp.	0.4450	0.785	1.031	90	90	105.1
10 ns	Opt.	0.4637	0.7721	1.032	89.81	90.62	103.48
50 ps	Opt.	0.4580	0.7835	1.031	90.00	90.01	104.61
	Ori.	0.4935	0.7786	1.044	89.99	90.03	105.85

Table S7f. Deviation (in %) between the predicted and experimental unit cell parameters.

Crystal		a	b	c	α	β	γ
Cellulose Iβ	10 ns Opt.	0.06	-1.06	0.72	-0.01	0.00	-2.36
	50 ps Opt.	0.24	-0.63	0.54	-0.02	0.03	-1.32
	Ori.	7.71	-0.51	1.30	0.00	0.00	-5.93
Cellulose Iα	10 ns Opt.	-3.95	0.87	0.63	-1.31	-0.18	0.98
	50 ps Opt.	-3.16	-0.07	0.44	-1.38	-0.67	1.43
	Ori.	-3.68	7.13	1.26	-3.06	-0.43	-1.93
Cellulose II	10 ns Opt.	1.90	-3.20	0.42	0.03	0.02	-2.19
	50 ps Opt.	1.71	-1.18	0.25	0.01	0.01	-0.82
	Ori.	0.16	4.95	1.47	0.02	0.01	-3.47
Cellulose III	10 ns Opt.	4.20	-1.64	0.09	-0.21	0.69	-1.54
	50 ps Opt.	2.92	-0.19	0.00	0.00	0.01	-0.47
	Ori.	10.90	-0.82	1.24	-0.01	0.03	0.71

Tables S8. Optimizing only the repulsive LJ parameter of the CH1r atom type for GROMOS 56Acarbo, using the charge distribution of CHARMM C35.

Table S8a-b. Optimized parameter: $C12^{1/2}$ in (kJ mol^{-1} nm^{12} $)^{1/2}$, distance σ in nm and energy ε in kJ mol^{-1} at the minimum, deviations (in %) between the optimized parameter and its original value, final values of target function F_{min}.

	$C6^{1/2}$	$C12^{1/2}$	σ	E	$\frac{C12^{1/2}_{optimized} - C12^{1/2}_{initial}}{C12^{1/2}_{initial}}$	F_{min}
Initial	0.0779	0.00985	0.5019	0.0949		
Four crystals	0.0779	0.006633	0.4399	0.2093	-32.7	62.9

Table S8c. Comparison of unit cell parameters predicted by original LJ parameters and optimized repulsive parameter of CH1.

		a (nm)	b (nm)	c (nm)	α (°)	β (°)	γ (°)
Cellulose Iβ	Exp.	0.7784	0.8201	1.038	90.00	90.00	96.55
10 ns	Opt.	0.7799	0.8089	1.044	89.99	90.00	94.17
	Ori.	0.8426	0.8137	1.043	89.99	90.01	92.87
Cellulose Iα	Exp.	0.6717	0.5962	1.040	118.08	114.80	80.37
10 ns	Opt.	0.6481	0.5954	1.048	116.25	114.03	81.43
	Ori.	0.6661	0.6241	1.047	114.91	113.33	78.66
Cellulose II	Exp.	0.8010	0.9030	1.031	90.00	90.00	117.10
10 ns	Opt.	0.8146	0.8628	1.039	90.02	90.01	114.00
	Ori.	0.8058	0.9380	1.041	90.10	90.05	112.74
Cellulose III	Exp.	0.4450	0.7850	1.031	90.00	90.00	105.10
10 ns	Opt.	0.4715	0.8203	1.000	90.00	89.97	111.86
	Ori.	0.5008	0.8192	1.006	89.99	89.99	111.90

Table S8d. Deviation (in %) between the predicted unit cell parameters and the experimental ones.

Crystal		a	b	c	α	β	γ
Cellulose Iβ	10 ns Opt	0.19	-1.37	0.53	-0.01	0.00	-2.47
	Ori.	8.25	-0.78	0.48	-0.01	0.01	-3.81
Cellulose Iα	10 ns Opt	-3.51	-0.13	0.81	-1.55	-0.67	1.31
	Ori.	-0.83	4.68	0.69	-2.69	-1.28	-2.13
Cellulose II	10 ns Opt	1.70	-4.45	0.78	0.02	0.01	-2.65
	Ori.	0.60	3.88	0.94	0.11	0.05	-3.72
Cellulose III	10 ns Opt	5.96	4.50	-2.99	0.00	-0.03	6.43
	Ori.	12.54	4.36	-2.42	-0.01	-0.01	6.47

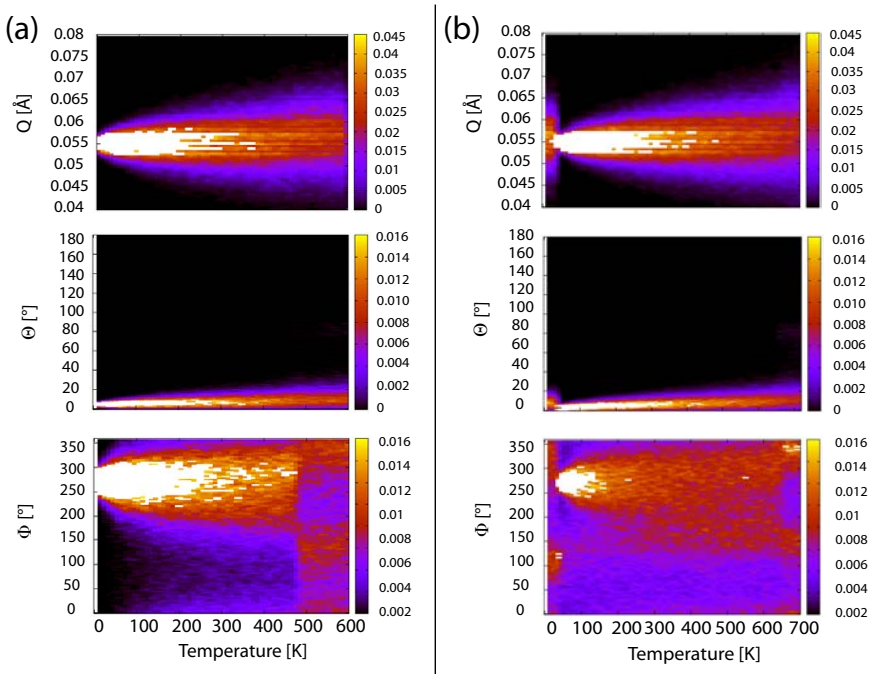

Figure S1. Histogram of puckering parameters as a function of temperature during heating by using original (a, K_1=9.35 kJ/mol) and modified (b, K_1=4.5 kJ/mol) torsional parameters.

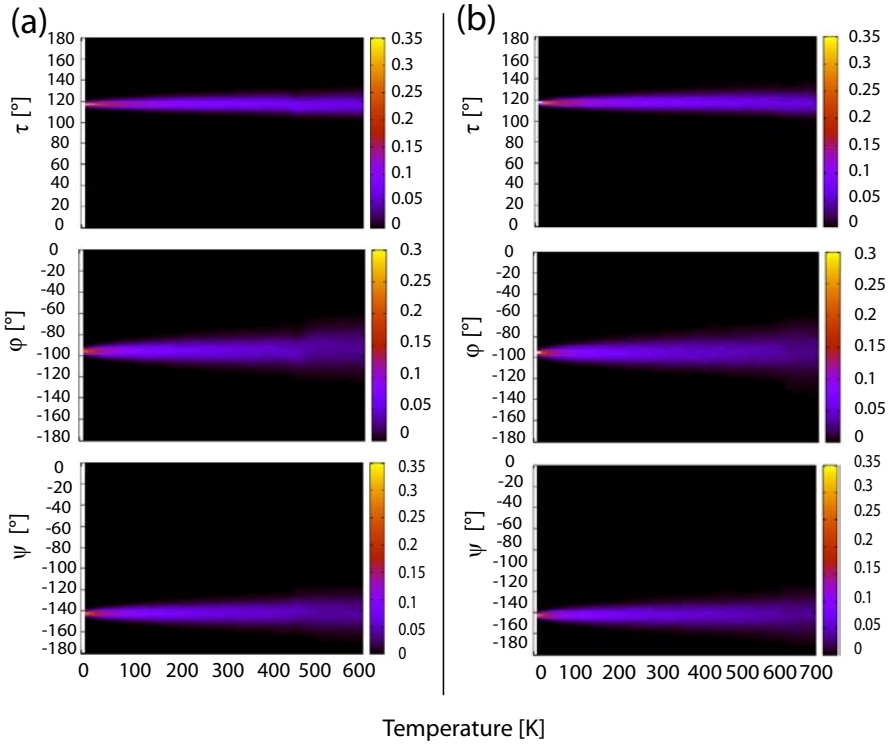

Figure S2. Histogram of glycosidic angles as a function of temperature during heating by using original (a, K_1=9.35 kJ/mol) and modified (b, K_1=4.5 kJ/mol) torsional parameters.

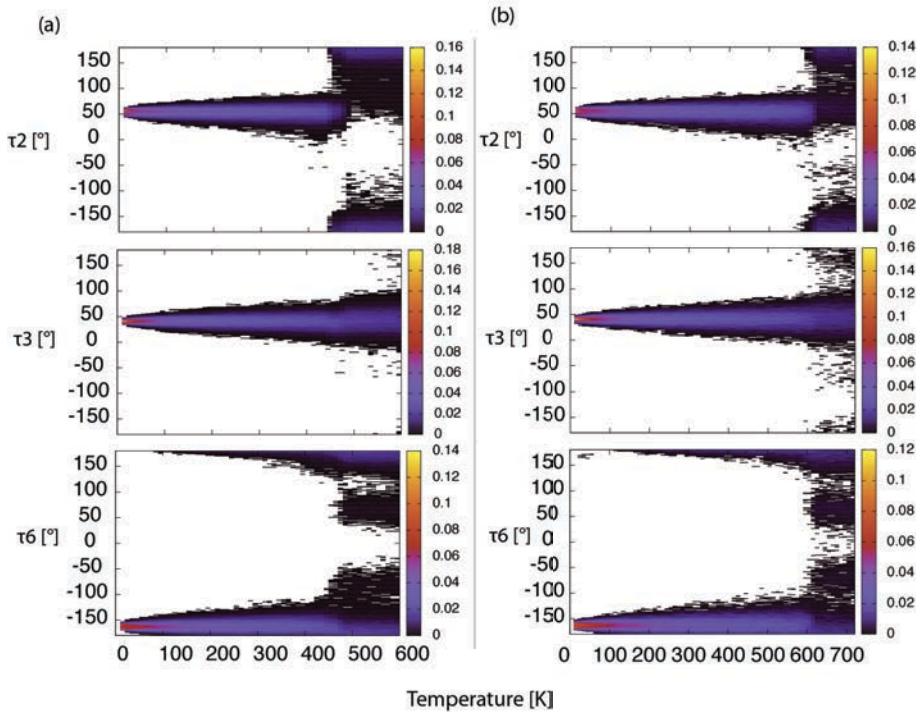

Figure S3. Histogram of hydroxyl groups as a function of temperature during heating by using original (a, K_1=9.35 kJ/mol) and modified (b, K_1=4.5 kJ/mol) torsional parameters.

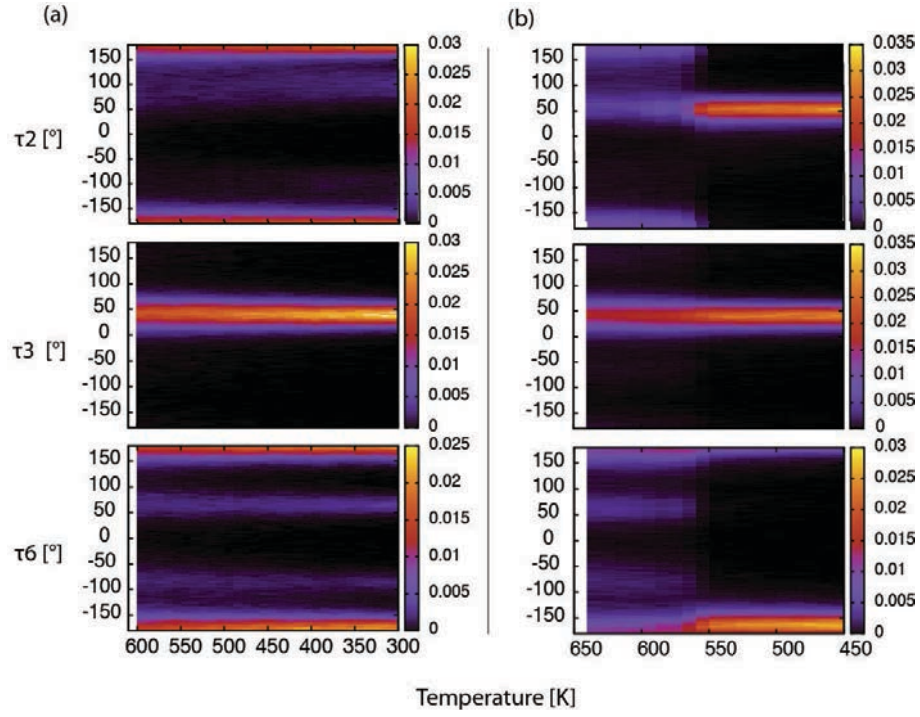

Figure S4. Histogram of hydroxyl groups as a function of temperature during cooling by using original (a, K_1=9.35 kJ/mol) and modified (b, K_1=4.5 kJ/mol) torsional parameters.

Résumé

La structure et les propriétés physiques de quatre types de cellulose cristalline (Iα, Iβ, II et III_I), incluant les caractéristiques des régions de cœur et de surface, leur comportement thermique, les propriétés mécaniques et les transitions allomorphiques, ont été étudiées au moyen de simulations de dynamique moléculaire (DM). Bien que ce type simulations soit en principe conçu pour reproduire des systèmes réels en se basant sur les lois physiques du mouvement, nous avons observé que les paramètres de maille cristalline et des propriétés telles que les transitions structurales thermiques n'étaient pas reproduites avec la précision escomptée. Afin de rendre les simulations de DM plus réalistes vis-à-vis des résultats expérimentaux, nous avons tout d'abord identifié l'origine de ces divergences. En utilisant l'un des plus simples champs de force de première génération, celui de GROMOS, nous avons trouvé qu'un paramètre relatif à la rotation du groupement hydroxyméthyle jouait un rôle significatif dans la capacité à reproduire les transitions de phases thermiquement activées de la cellulose de type I. Nous avons de plus trouvé que les charges atomiques partielles autour de la liaison glycosidique influençaient l'angle monoclinique de la cellulose Iβ et que la composante répulsive du terme de Lennard-Jones de l'atome unifié CH1 déterminait la distance entre les plans des glucopyranoses dans la cellulose cristalline. En utilisant les paramètres optimisés du champ de force, nous avons simulé la structure des quatre allomorphes de la cellulose de manière satisfaisante. Par ailleurs, nous avons réalisé des simulations exploratoires utilisant les paramètres optimisés dont les résultats ont été comparés à ceux obtenus avec les paramètres d'origine.

Abstract

The structure and physical properties of four types of crystalline cellulose (Iα, Iβ, II and III_I), including the structural details of bulk and surface regions, thermal behavior, mechanical properties and allomorphic conversion, were studied by molecular dynamics (MD) simulation. Although MD simulation is in principle designed to reproduce the real world based on the basic physical laws of motion, we observed that the unit cell parameters and properties such as temperature-induced phase transitions were not reproduced to the expected precision. In order to make MD more reliable to complement experiments, we first tried to identify the origin of these discrepancies. Using one of the simplest first generation force fields, namely GROMOS, we found that one parameter related to the dihedral angle of the hydroxymethyl group directly affects the possibility to reproduce temperature-induced phase transitions of cellulose I. We also found that the partial atomic charges around the glycosidic linkage influence the monoclinic angle of cellulose Iβ and that the repulsive Lennard-Jones term of the united atom CH1 determines the distance between glucopyranose planes of crystalline cellulose. By using the optimized force field parameters, the structure of the four cellulose allomorphs was better reproduced. In addition, exploratory simulations based on the optimized parameters were performed and the results are compared with the ones using native force field parameters.