

Diss. ETH No. 18557

Computer simulations of carbohydrates

A dissertation submitted to the
ETH ZÜRICH

for the degree of
Doctor of Sciences

presented by
LOVORKA PERIĆ-HASSLER
Dipl. Ing. Chem.
born April 6, 1979
citizen of Croatia

accepted on the recommendation of
Prof. Dr. Wilfred F. van Gunsteren, examiner
Dr. Philippe H. Hünenberger and Prof. Dr. Roland Riek, co-examiners

2009

Acknowledgments

I would like to thank both Prof. Wilfred van Gunsteren and Dr. Philippe Hünenberger for accepting me into IGC. I'm utterly thankful to Philippe, for being my *de facto* supervisor and for his guidance and support throughout my PhD. I consider myself extremely lucky that I had a chance to work with him. The combination of his genuine modesty, understanding, ease of communication on a personal level and brilliant sense of humor with the sharpness and thoroughness of his critical mind is a rare and beautiful virtue. I would like to thank Wilfred for being my Doktorvater, for his outstanding knowledge as well as his willingness to transfer it to us. He taught us that it is not only the content of our work that is important but also the skill with which we present it and he gave us many opportunities to do that. IGC always seemed to me to be an example of an excellent place to work, but that's largely due to his wisdom to guide and balance all the aspects of this group.

I wish also to thank Dr. Roland Riek, for being my co-examiner.

Special thanks goes to Phil's angels; Halvor with his radiating good energy, Maria with her gracious, unobtrusive ways of being simply great, charming Cristina with her Latino temperament and Bruno with his always intriguing stories and "Guten Morgen" every morning.

Thanks to Zrinka and her more than a decade long friendship and support, lovely Jozi for many nice moments, Daan for jazz&chess and his genial jokes, Bettina for the great time inside and outside the ETH, Clara for her always supportive attitude and music, Moritz for being so cool, Dirk for interesting lunch-coffee-break conversations, Bojan for presenting Croatia in an American way, Anna-Pitschna for her so appealing "Coffee?!!!" and Nathan for being so witty and relaxed. Many thanks to the "older" generation of IGC members: Vincent, Tim, Vreni, Chris, Mika, Michel, Merijn, Markus, Riccardo and Daniel. It was a pleasure to start my PhD in such an atmosphere with so many creative and friendly people. Big thanks to the "new" generation: Denise, Katharina, Jane, Dongqi, Andreas, Alexandra, Zhixiong, Sereina. I appreciate very much their support while finishing my PhD. Thanks to Daniela for her optimism and her millions of useful, friendly pieces of advice and Ana for her spontaneity and for bringing pianos into the group room and into my apartment. IGC is a great team; these years are so precious for me and it is not easy to express my gratitude as I would like to. The many farewell parties we had together bear witness to that authentic IGC spirit and I would like to thank Jolande very much for her hospitality, for welcoming us into her and Wilfred's house, for caring for each of us and for being here at the most important moments of IGC life. A lot of thanks goes to Valerie, Achille and Gaspard, for their patience with Phil while he is working on our manuscripts, for being a great team, and for the happy playful moments we had.

I am deeply grateful to my family and Željka, who always had a lot of trust in me. Lastly, I would like to thank Fabian, my husband, my smart physicist and philosopher. Your love is a blessing. I dedicate this thesis to you.

Contents

| | |
|--|------------|
| Acknowledgments | i |
| Kurzfassung | vii |
| Summary | ix |
| Publication | xi |
| 1 Introduction | 1 |
| 1.1 The complexity of carbohydrates | 1 |
| 1.2 Structure and conformation of carbohydrates | 2 |
| 1.3 Molecular modeling of carbohydrates | 8 |
| 1.3.1 Available molecular models | 8 |
| 1.3.2 Molecular model used in this work | 9 |
| 1.3.3 Configuration generation | 10 |
| 1.3.4 Boundary conditions | 12 |
| 1.4 References | 12 |
| 2 Conformational properties of glucose-based disaccharides investigated using molecular dynamics simulations with local elevation umbrella sampling | 27 |
| 2.1 Summary | 27 |
| 2.2 Introduction | 28 |
| 2.3 Computational details | 32 |
| 2.3.1 Molecular dynamics simulations | 32 |
| 2.3.2 Analysis | 34 |
| 2.4 Results | 38 |
| 2.4.1 Unbiased MD simulations | 38 |
| 2.4.2 LEUS protocol | 39 |
| 2.4.3 Free-energy maps | 41 |
| 2.4.4 Conformational states: definition and sampling | 47 |
| 2.4.5 Comparison with experimental data | 51 |
| 2.4.6 Conformational states: thermodynamics and kinetics | 53 |
| 2.4.7 Hydrogen bonding | 55 |
| 2.4.8 Configurational entropy | 57 |
| 2.5 Conclusions | 60 |
| 2.6 References | 65 |

| | | |
|----------|---|------------|
| 3 | Interaction of the disaccharides trehalose and gentiobiose with lipid bilayers: a comparative molecular dynamics study | 77 |
| 3.1 | Summary | 77 |
| 3.2 | Introduction | 78 |
| 3.3 | Computational details | 82 |
| 3.3.1 | Molecular dynamics simulations | 82 |
| 3.3.2 | Molecular systems | 83 |
| 3.3.3 | Trajectory analysis | 84 |
| 3.4 | Results | 86 |
| 3.4.1 | Distribution profiles | 86 |
| 3.4.2 | Order parameters | 89 |
| 3.4.3 | Bilayer structure | 90 |
| 3.4.4 | Hydrogen-bonding | 92 |
| 3.5 | Conclusion | 97 |
| 3.6 | References | 101 |
| 4 | Conformation, dynamics and ion-binding properties of single-chain polyuronates | 109 |
| 4.1 | Summary | 109 |
| 4.2 | Introduction | 109 |
| 4.3 | Computational details | 114 |
| 4.3.1 | Molecular Dynamics Simulation | 114 |
| 4.3.2 | Analysis of the trajectories | 119 |
| 4.4 | Results and discussion | 121 |
| 4.4.1 | Ring conformation and carboxylate group orientation | 121 |
| 4.4.2 | Glycosidic linkage conformation | 122 |
| 4.4.3 | Helical parameters | 124 |
| 4.4.4 | Intramolecular hydrogen bonding | 132 |
| 4.4.5 | Ion binding | 134 |
| 4.4.6 | Diffusion properties | 139 |
| 4.5 | Conclusion | 140 |
| 4.6 | References | 143 |
| 4.7 | Appendix | 148 |
| 5 | Interaction of alginate single-chain polyguluronate segments with mono- and divalent metal cations: a comparative molecular dynamics study | 163 |
| 5.1 | Summary | 163 |
| 5.2 | Introduction | 164 |
| 5.3 | Computational methods | 167 |
| 5.3.1 | Molecular dynamics simulation | 167 |
| 5.3.2 | Analysis of the trajectories | 170 |
| 5.4 | Results | 172 |
| 5.4.1 | Chain conformation and hydrogen-bonding | 172 |
| 5.4.2 | Distribution of the ions around the chain | 172 |
| 5.4.3 | Ion binding to the chain | 177 |
| 5.4.4 | Chain and ion diffusion | 182 |

| | | |
|----------|----------------------------------|------------|
| 5.4.5 | Effective chain charge | 183 |
| 5.5 | Conclusions | 185 |
| 5.6 | References | 187 |
| 6 | Outlook | 195 |
| | Curriculum Vitæ | 197 |

Kurzfassung

Auf Grund ihrer gewaltigen strukturellen und konformationellen Vielseitigkeit erfüllen Kohlenhydrate ein grosses Spektrum an Funktionen in lebenden Zellen. Diese Komplexität führt jedoch auch zu starken Einschränkungen in ihrer experimentellen Beschreibung. Molekulardynamiksimulationen (MD-Simulationen) stellen ein Werkzeug dar, welches komplementär zum Experiment Einblicke in die Struktur, Flexibilität und Dynamik von (Bio)Molekülen auf atomarer Ebene liefert. Die vorliegende Arbeit beschäftigt sich mit Anwendungen von MD-Simulationen, mit dem Ziel, die Eigenschaften bestimmter Di- und Polysaccharide zu untersuchen. Kapitel 1 liefert eine Einführung in die relativ komplexen Gebiete der Kohlenhydratnomenklatur, -struktur und -konformation. Ausserdem gibt es eine Übersicht über die verschiedenen Prozesse und physikalischen Effekte, welche die Konformationseigenschaften von Mono-, Oligo- und Polysacchariden bestimmen. Diese Einleitung beinhaltet zudem einen kleinen Überblick über die verfügbaren Methoden im Bereich der Kohlenhydratmodellierung (mit dem Schwerpunkt auf grundlegenden Entscheidungen hinsichtlich der Definition eines Molekularmodells) und eine Beschreibung jener Methode (atomare MD-Simulation in expliziten Lösungsmitteln), welche in der vorliegenden Arbeit verwendet wird. Das zweite Kapitel befasst sich mit den Konformationseigenschaften der elf aus Glukose aufgebauten Dissaccharide, d.h. den einfachsten Kohlenhydratmolekülen, welche alle rotatorischen Freiheitsgrade komplexerer Kohlenhydrate aufweisen. Diese Eigenschaften werden im Rahmen einer MD-Simulation zusammen mit einem verbesserten Sampling-Schema, der “local elevation umbrella sampling”-Methode (LEUS), untersucht. Dieser Ansatz erlaubt die Darstellung der freien Energie von Konformationen als Funktion der glykosidischen Dihedralwinkel, sowie die Bestimmung von stabilen und metastabilen Konformationen, intramolekularen Wasserstoffbrücken und konfigurationellen Entropien. Das Verfahren ermöglicht zudem eine Validierung der Kraftfeldparameter, indem die gewonnenen Resultate mit vorhandenen experimentellen Daten verglichen werden. Das dritte Kapitel legt eine vergleichende Studie der Wechselwirkung der Disaccharide Trehalose und Gentiobiose mit einer Lipiddoppelschicht vor. Diese zwei Disaccharide, welche sich in der Art und Flexibilität ihrer glykosidischen Bindung unterscheiden, zeigen eine unterschiedliche Fähigkeit, die Doppelschichtstruktur unter thermischer Belastung zu erhalten. Die Ursache dieses unterschiedlichen Verhaltens wird auf der Grundlage von MD-Simulationen bei Raumtemperatur erörtert und dabei in den Zusammenhang mit möglichen anderen Mechanismen, welche schon früher zur Erklärung der bioprotektiven Rolle von Zuckern in lebenden Zellen unter potentiell schädlichen Bedingungen herangezogen worden sind, gestellt. Kapitel 4 untersucht die Konformationen, Flexibilität und ionenbindenden Eigenschaften von vier verschiedenen Arten von Homopolyuronatketten (Grundbausteine der natürlichen Polysaccharide Alginate, Pektine und Glukuronane). Die Ergebnisse dieser Simulationen werden mittels ihrer Auswirkungen auf die Bildung von Polyuronat-Gelen in der Gegenwart divalenter Metallkationen erklärt. Das fünfte Kapitel erweitert diese Untersuchung und liefert eine vergleichende Studie der Wechsel-

wirkung von einzelnen Homopolyglukuronatketten (einem Grundbaustein von Alginaten) mit unterschiedlichen mono- und divalenten Kationen. Zuletzt wird im Kapitel 6 ein kurzer Ausblick auf mögliche zukünftige Entwicklungen im Forschungsgebiet, welches in dieser Arbeit behandelt wird, gegeben.

Summary

Due to their immense structural diversity and conformational versatility, carbohydrates fulfill a wide range of functions in living cells. However, this complexity also imposes severe limitations on their experimental characterization. Molecular dynamics (MD) simulation represents a powerful tool complementary to experiment to gain insight into the structure, flexibility and dynamics of (bio)molecules on the atomic scale. The present thesis is concerned with the application of MD to study the properties of specific di- and polysaccharide systems.

Chapter 1 provides an introduction into the relatively complex domains of carbohydrate nomenclature, structure and conformation, including an overview of the different processes and physical effects determining the conformational properties of mono-, oligo- and polysaccharides. This introduction also includes a brief review of the available methods in the field of carbohydrate modeling (with the emphasis on the basic choices defining a molecular model) and a description of the modeling approach relevant for the present thesis (atomistic explicit-solvent MD simulation).

Chapter 2 is concerned with the conformational properties of the 11 glucose-based disaccharides, *i.e.* the simplest carbohydrate molecules presenting all rotational degrees of freedom of more complex carbohydrates. These properties are investigated using MD along with a sampling enhancement scheme, the local elevation umbrella sampling (LEUS) method. This approach permits the determination of free-energy maps in the space of the glycosidic dihedral angles, stable and metastable conformational states, intramolecular hydrogen-bonds and solute configurational entropies. It also provides validation for the force-field employed by comparison with available experimental data.

Chapter 3 presents a comparative study of the interaction of the disaccharides trehalose and gentiobiose with a lipid bilayer. These two disaccharides, which differ only in the type and flexibility of their glycosidic linkage, reveal unequal abilities to preserve the integrity of the bilayer structure under the thermal stress. The reason for this difference is discussed on the basis of MD simulations at room temperature and placed in the context of the possible mechanisms suggested previously to explain bioprotective role of sugar in living cells exposed to potentially damaging conditions.

Chapter 4 investigates the conformation, flexibility and ion-binding properties of four different types of homopolyuronate chains (building blocks of the natural polysaccharides alginates, pectins and glucuronans). The result of these simulations are discussed in terms of their implication concerning the formation of gels by polyuronates in the presence of divalent metal cations.

Chapter 5 expands this study by performing a comparative investigation of the interaction of single chain homopolyglucuronate (a building block of alginates) with different mono- and divalent cations.

Finally, Chapter 6 provides a short outlook on possible future directions in the research field covered by the present thesis.

Publication

This thesis has led to the following publication:

Chapter 4:

Lovorka Perić, Crististina S. Pereira, Serge Pérez, and Philippe H. Hünenberger,
“Conformation, dynamics and ion-binding properties of single-chain polyuronates: a molecular
dynamics study”
Molecular Simulation **34**, 421–446 (2008).

Chapter 1

Introduction

1.1 The complexity of carbohydrates

Carbohydrates are the most abundant class of biomolecules on earth¹⁻³. In addition to their structural and energetic function in living organisms^{4,5}, they are protagonists in many other biological processes⁶⁻¹³ including fertilization, immune defense, viral replication, parasitic infection, cell growth, cell-cell adhesion, inflammation, malignant transformations and cryptobiosis. In addition, they play a key role in numerous technological applications related to *e.g.* the food, textile, paper and cosmetics industries^{14,15}, in many novel biomedical applications^{9,16,17} as well as in the design of new (biodegradable and biocompatible) materials¹.

The broad range of functions fulfilled by carbohydrates in living cells is certainly related to their amazing *structural diversity*^{7,18-21}. This class of compounds encompasses a huge variety of possible monomeric units (differing in stereochemistry and functionalization), that can be connected in chains presenting a virtually infinite number of possible residue sequences, linkage types and degrees of branching. Unlike proteins, nucleic acids and lipids, which tend to predominantly adopt a well-defined (native) conformation under the conditions where they are biologically functional, carbohydrates are typically associated with a high extent of *conformational heterogeneity*^{16,18,20,22-24}, the corresponding conformational ensembles being determined by an intricate interplay between steric, stereoelectronic, electrostatic (*e.g.* hydrogen bonds) and solvation effects²¹. As a result of this structural diversity and conformational heterogeneity, carbohydrates arguably represent the most challenging class of biomolecules in terms of experimental characterization and elucidation of structure-function relationships.

Static structures of carbohydrates may often be obtained from (x-ray) crystallography^{22,25-28} (of crystals or fibers), but it is always uncertain whether these molecules adopt similar conformations in solution as in the solid state. On the other hand, nuclear magnetic resonance (NMR) spectroscopy provides information about carbohydrates in solution^{22,26,28-30}, but only in the form of averages over all the populated conformational states present in solution (*i.e.* over all the molecules or molecular segments in the sample as well as over the timescale of the NMR experiment). Many other experimental techniques (*e.g.* electron microscopy, light or neutron diffraction, circular dichroism, infrared spectroscopy, or rheology) provide useful, but even more indirect information about carbohydrate conformations^{27,31}.

Due to these difficulties, molecular modeling has emerged as a powerful tool to comple-

ment experiments for the investigation of carbohydrate-based systems^{22,31–35,35–46}. In particular, explicit-solvent molecular-dynamics (MD) simulations⁴⁷ provide information over a length- and timescale inaccessible to any experimental method (atomic and femtosecond resolution simultaneously). The price to pay is that the accuracy of such simulations is intrinsically limited by that of the underlying computational model (force-field parameters, accessible system size and timescale, simulation methodology).

During the past decades, many progresses have been made in the design of force fields for proteins, nucleic acids and lipids⁴⁸. However, the design of carbohydrate force fields is significantly more complex^{23,49}, due to the structural diversity and conformational heterogeneity of these compounds as well as to the relative scarcity of data suitable for force-field refinement. As a consequence, up to recently, the area of carbohydrate simulation has lagged somewhat behind that of the other biomolecules.

1.2 Structure and conformation of carbohydrates

Carbohydrates^{7,14,21} are organic compounds built from simple monosaccharide units with the general formula $C_n(H_2O)_n$ via polymerization (formation of glycosidic linkages) or/and functionalization (replacement of specific hydroxyl groups by other functional groups). The basic monosaccharide building blocks consist of a carbon chain where all but one carbon atoms bear hydroxyl groups, the remaining one being a carbonyl group. Depending on the location of this carbonyl group (aldehyde or ketone), the monosaccharides are classified into *aldoses* and *ketoses*. A aldose with n carbon atoms possesses $n - 2$ chiral centers, leading to 2^{n-2} stereoisomers. A ketose with n carbon atoms possesses $n - 3$ chiral centers leading to 2^{n-3} stereoisomers. These stereoisomers are divided into two *enantiomeric series* according to the position of the hydroxyl group attached to carbon C_{n-1} in the Fischer projection (*L*-series: left; *D*-series: right). Two structures differing by the orientation of their hydroxyl groups are called *epimers*. If they differ by the orientation of a single hydroxyl group at a given carbon C_n they are called C_n -epimers. Thus, for example, aldohexoses (aldoses with six carbon atoms) consist of two enantiomeric series (*L* and *D*) of eight epimers each (ketohexoses only have four epimers in each series), and the aldohexoses *D*-mannose and *D*-galactose are the C_2 - and C_4 -epimers, respectively, of *D*-glucose. The carbonyl group of a monosaccharide can react reversibly with a hydroxyl group in the same molecule, forming a heterocyclic hemiacetal or hemiketal. Because they usually provide an optimal compromise between the entropic cost of cyclization against the resulting angle and eclipsing strain, the preferred ring sizes are five-membered rings (*furanoses*) and six-membered rings (*pyranoses*). The intramolecular cyclization converts the carbonyl group to a carbon-hydroxyl group. The specific carbon is then referred to as the *anomeric center* and the attached hydroxyl group as the *lactol group*. The two possible stereochemical configurations at the anomeric center are termed the α and β *anomers* (Figure 1.1).

The lactol group of a monosaccharide may further condense with the hydroxyl group of an other molecule (*e.g.* an alcohol or another monosaccharide), resulting in the formation of an acetal or ketal. In this case, the cyclized form becomes permanent and, because the carbonyl function is no longer available, the monosaccharide is termed *non-reducing*. When the latter hydroxyl group is provided by another monosaccharide molecule, the resulting connection is called a *glycosidic linkage*. The connectivity and stereochemistry of a linkage is described by the configuration of the

anomeric center of the first (non-reducing) residue and the position of the linkage on the second (reducing in a disaccharide; non-reducing if itself involved in another glycosidic linkage) residue (Figure 1.1). In case of a possible ambiguity, the ring size of the residues is sometimes also specified. Thus, for example, $\text{Glc}\alpha(1 \rightarrow 4)\text{Glc}\beta$ indicates an α anomery at the C_1 carbon atom of the first glucopyranose ($\text{Glc}\beta$) residue and a linkage involving this atom and the C_4 hydroxyl group of the second residue.

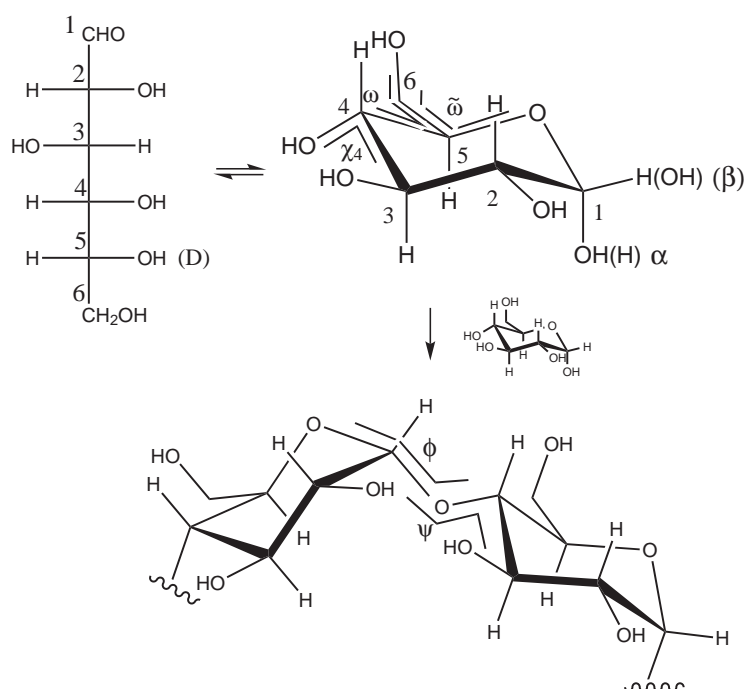


Figure 1.1: The principle of monosaccharide cyclization and glycosidic linkage formation are illustrated using D-glucopyranose and an $\alpha(1 \rightarrow 4)$ linkage as examples. The atom labeling is also indicated. The primed labels correspond to the second residue that is linked to the (non-reducing) one. According to the IUPAC nomenclature^{50,51} the dihedral angles defining the orientation of the hydroxyl groups are labeled χ_n ($\text{C}_{n-1}-\text{C}_n-\text{O}_n-\text{H}_n$; illustrated for $n = 4$; χ_n' for the second residue). The dihedral angles defining orientation of the hydroxymethyl group are labeled ω ($\text{C}_4-\text{C}_5-\text{C}_6-\text{O}_6$; ω' for the second residue) or $\tilde{\omega}$ ($\text{O}_5-\text{C}_5-\text{C}_6-\text{O}_6$; non-IUPAC, but also commonly used; $\tilde{\omega}'$ for the second residue). Finally, the dihedral angles defining the conformation of the glycosidic linkage are labeled ϕ ($\text{O}_5-\text{C}_1-\text{O}_1-\text{C}'_1$) and ψ ($\text{C}_1-\text{O}_1-\text{C}'_1-\text{C}'_{n-1}$). Note that in the special case of a three-bond ($1 \rightarrow 6$)-linkage, ω' (or $\tilde{\omega}'$) is also a glycosidic dihedral angle. Note also that ψ is equivalent to χ_n' for a ($1 \rightarrow n$)-linkage.

The conformational freedom of a mono-, oligo- or polysaccharide can be broadly defined in terms of six different classes of processes^{21,52,53} associated with different timescales, which are described below with a specific focus on aldohexo(pyrano)se-based saccharides in aqueous solution:

1. *Constitutional isomerization.* For reducing monosaccharides, at least six distinct constitu-

tional isomers may be present at equilibrium in solution^{54–62}: the α - and β -pyranose forms (six-membered rings), the α - and β -furanose forms (five-membered rings), and the (free or hydrated) acyclic forms (free or hydrated carbonyl group). Septanose forms^{63–65} (seven-membered rings) are not observed for aldohexoses in water^{65–67}. Constitutional changes (mutarotations) are typically associated with timescales^{54, 62, 68–73} of the order of 10^3 s, and the relative populations of the corresponding species may be readily determined by NMR spectroscopy^{55, 56, 58, 59, 61, 74}. For aldohexoses in aqueous solution, the α - and β -pyranose forms (mixture) are the leading constitutional isomers. The contribution of the furanose forms is below 10% for all aldohexoses⁶¹ except altrose, idose and talose (about 30%), while that of acyclic isomers is below 0.1% in all cases^{58, 59, 61, 75} except for idose (about 1%). For non-reducing monosaccharides, the constitution is defined during the (chemical or enzymatic) formation of the glycosidic linkage and cannot change anymore thereafter.

- Ring conformational isomerization.* Pyranose systems may present at least eight relevant ring conformers in solution^{55, 57, 76–79}: two chair forms (labeled 4C_1 and 1C_4) and six boat forms (labeled 1,4B , 2,5B , 3,O B, $B_{1,4}$, $B_{2,5}$ and $B_{3,O}$). Skew-boat forms may also be of relevance for some systems^{80–92}. For nearly all aldohexopyranoses of the *D*-series, the 4C_1 chair conformation appears to be the only significantly populated ring conformer in aqueous solution^{76, 93}. For aldohexopyranoses of the *L*-series, this behavior is inverted (*e.g.* the 1C_4 conformation of a β -*L*-hexopyranose is the mirror image of the 4C_1 conformation of the corresponding β -*D*-hexopyranose). Ultrasonic-relaxation^{73, 94–96} and NMR spectroscopy⁹⁷ measurements suggest that chair-chair interconversion in aldohexopyranoses takes place on the μ s timescale.
- Conformation of the exocyclic hydroxymethyl group.* The conformational properties of the exocyclic hydroxymethyl group in aldohexopyranoses may be investigated based on NMR coupling constants together with appropriate Karplus-type equations^{98–106}. The reported conformer populations around ω for glucose^{101, 103, 104, 106–112}, mannose¹¹³ and galactose^{101, 103, 104, 108, 110, 112}, either free or O_1 -methylated, show significant variations between different literature sources. However, for glucose and mannose, all studies evidence a clear predominance of the g_+ (60°) and t (180°) rotamers (with comparable populations), g_- (-60°) being nearly absent. For galactose, one set of studies^{101, 108, 110, 112} suggests that the three rotamers are significantly populated, while the two most recent determinations^{103, 104} support a predominance of the t and g_- rotamers (with a higher population for t , g_+ being nearly absent). Ultrasonic-relaxation spectroscopy^{73, 95, 114} and NMR^{97, 115, 116} measurements suggest that the hydroxymethyl group rotation takes place on the nanosecond timescale.
- Conformation of the exocyclic hydroxyl groups.* Experimentally, little is known about the orientational preferences and the torsional dynamics of the exocyclic hydroxyl groups in aqueous solution. This is mainly because: (i) hydroxyl protons are difficult to detect via NMR spectroscopy^{74, 117–120} (chemical exchange, mutual overlap, overlap with solvent signal); (ii) hydroxyl group vibrations are difficult to characterize via infrared (IR) spectroscopy^{118, 121–123} (dual hydrogen bond donor-acceptor character, mutual overlap, overlap with solvent bands). Furthermore, X-ray crystallography provides limited reference information in the solid state, because the hydroxyl protons are invisible in these exper-

iments (weak diffraction centers, orientational averaging). However, detailed (although model-dependent) information concerning rotamer distributions and intramolecular hydrogen bonding, as well as the associated timescales, may be obtained from atomistic simulations. Recent MD studies^{21,52} suggested a timescale of about 30 ps for the hydroxyl group rotation and relatively weak correlations between hydroxyl group orientations (*i.e.* weak hydrogen-bonding) around the pyranose ring in aqueous environment (except in the case of 1,3-*syn*-diaxial hydroxyl groups as *e.g.* in talose).

5. *Conformation of the glycosidic linkages (local preferences).* The local conformation of a glycosidic linkage is determined by the values of the glycosidic dihedral angles^{124–132} ϕ and ψ (along with ω' , often loosely written ω , in the case of (1→6) linkage), which determine the relative orientations of two successive monosaccharide rings. For a given linkage type, the populated values of these dihedral angles are restricted to limited ranges, resulting from an interplay between steric, stereoelectronic, electrostatic (*e.g.* interresidue hydrogen-bonding) and solvation effects. These ranges can be characterized by considering the Ramachandran free energy maps at the disaccharide level. Unfortunately, although the corresponding conformational preferences can be inferred qualitatively based on experimental (x-ray, NMR) data^{24,133}, the Ramachandran maps themselves are not accessible and the relative influences of the four above type of effects remain largely unclear (in particular concerning the role of interresidue hydrogen-bonding and solvation). Here also, detailed (although model-dependent) information may be obtained from atomistic simulations. Recent MD studies⁵² suggested the existence of a single dominant conformational basin in the Ramachandran map for all reducing glucose disaccharides, and timescales of the order of 100 ns (or longer) for rotations around the ϕ and ψ dihedral angles (about 1 ns for rotations around ω in (1→6)-linked disaccharides). Ultrasonic relaxation spectroscopy measurements^{73,96} suggest timescales of the order of 10 ns.
6. *Conformation of an oligo- or polysaccharide chain (non-local preferences).*

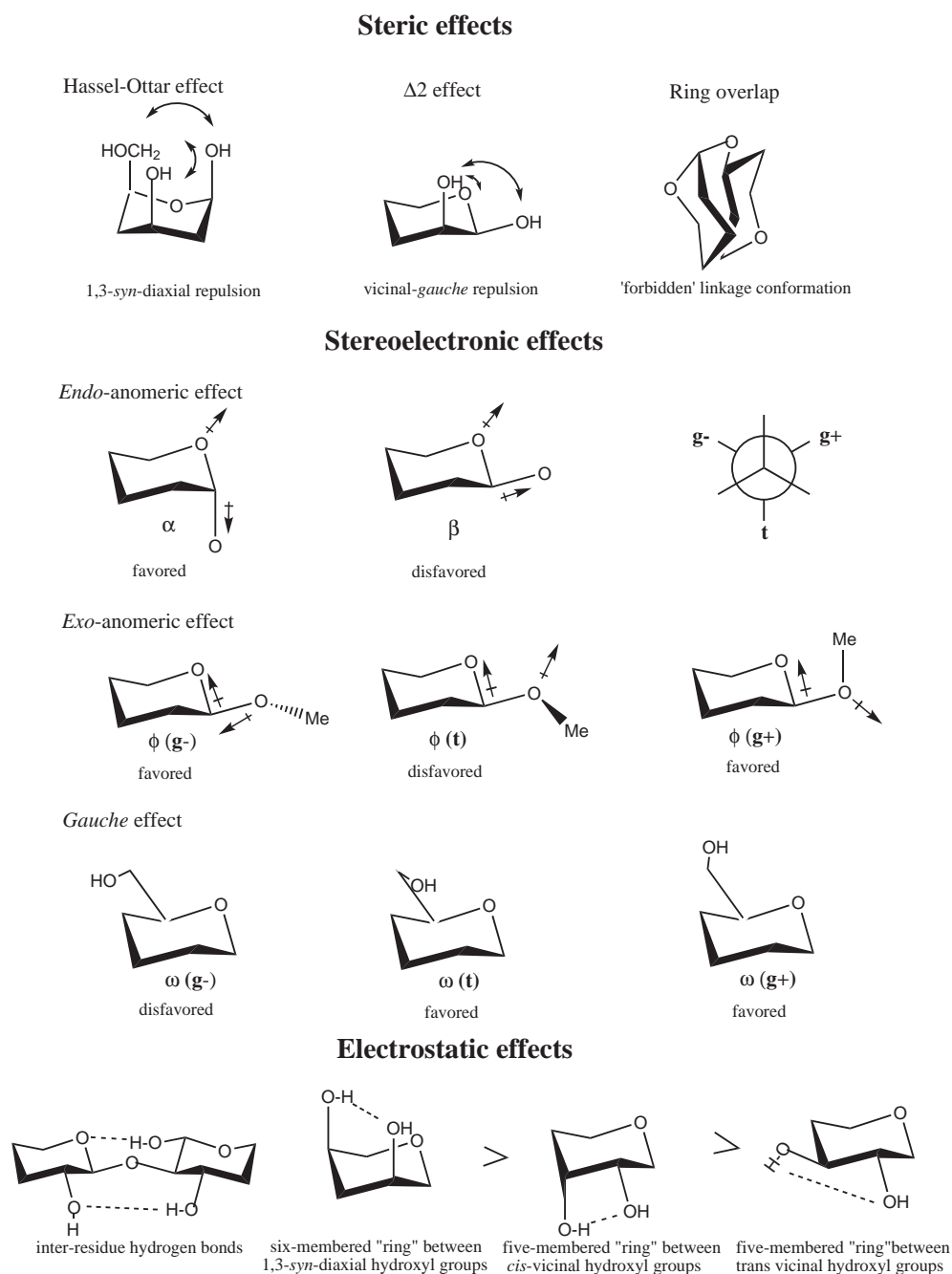
Local conformational preferences at the level of the individual glycosidic linkages are not rigid constraints, and are usually compatible with different kinds of longer-ranged structuring of the chain. Depending on the environment, oligo- and polysaccharides can thus still adopt a wide variety of shapes. In crystals or fibers, helical secondary structure (with tertiary lateral chain association) is the most common motif^{134–136}. A regular helical structure is usually characterized by the number n of residues per helix turn (often fractional; positive for right-handed and negative for left-handed helices) and the raise h per residue along the helix axis. Typical examples are the 2_1 -, 3_1 - and 3_2 -helices (with $n = \pm 2, +3$ and -3). In aqueous solutions, oligo- and polysaccharides are mostly disordered, but thought to locally adopt conformations similar to those found in the solid state¹³⁷. The characterization of polysaccharide conformations requires the use of spectroscopic methods^{22,23,26,138} such as x-ray crystallography^{25,27,28} (solid state), or methods applicable for studies of solutions like NMR spectroscopy^{28–30}, optical rotation (OR), fluorescence energy transfer (FET), Raman optical activity (ROA), as well as computer modeling^{22,28,32,33,35–42,46,139}.

For a given mono- oligo- or polysaccharide stereochemistry (chain length, sequence, linkages), the relative stabilities (and, thus, the associated equilibrium populations) of the different constitutional isomers, ring conformers, hydroxymethyl rotamers, hydroxyl rotamers, glycosidic

linkage conformers and chain conformers result from a delicate balance between many different effects.

A brief summary of these effects is provided below (with a focus on aldohexopyranose-based carbohydrates in aqueous solution), together with a few key conclusions resulting from experimental investigations :

1. *Steric effects.* Steric effects result from the (closed-shell) non-bonded repulsion between pairs of atoms in close spatial proximity, and should in principle be solvent insensitive. In the context of carbohydrates, these include²¹ (Figure 1.2): (i) the Hassel-Ottar effect^{55, 60, 76, 140–145} and other 1,3-*syn*-diaxial repulsions^{55, 76, 110, 141, 143, 144, 146–148} within a monosaccharide; (ii) the $\Delta 2$ effect^{60, 76, 149, 150} and other vicinal-*gauche* repulsions^{55, 60, 141, 143, 144, 146, 150–152} within a monosaccharide; (iii) residue-overlap (forbidden) regions of the Ramachandran map for a glycosidic linkage. Note, however, that the two former effects may not be entirely steric in nature, *i.e.* they may also involve a significant hyperconjugative (stereoelectronic)^{153–156} or dipolar^{147, 157, 158} (electrostatic) component. For example, recent articles have questioned the steric nature, magnitude, and even repulsiveness¹⁵⁵ of 1,3-*syn*-diaxial interactions involving aliphatic hydrogen atoms and axial substituents, suggesting that the general equatorial preference in substituted six-membered rings^{77, 159} is rather of predominantly hyperconjugative origin^{153–155, 160}.
2. *Stereoelectronic effects.* Stereoelectronic effects result from the particular electronic properties associated with specific molecular fragments in specific geometries. These effects may also involve a dipolar component¹⁶¹ and thus be somewhat solvent sensitive^{162, 163}, but this dependence should be limited compared to that of electrostatic effects (see below). In the context of carbohydrates, these include²¹ (Figure 1.2) : (i) the *endo*-anomeric effect^{55, 124, 144, 148, 150, 153, 161, 162, 164–169} (affecting the ratio of the α vs β anomers in reducing monosaccharides); (ii) the *gauche* effect^{165, 170–174} (affecting the populations of hydroxymethyl group rotamers); (iii) the *exo*-anomeric effect^{124, 162, 165–168, 175–177} (affecting the conformational preferences of the glycosidic dihedral angle ϕ) ; (iv) other stereoelectronic effects^{153, 154, 160, 178–181}.
3. *Electrostatic effects.* Electrostatic effects, mainly intramolecular hydrogen-bonding, are strongly sensitive to the nature and polarity of the solvent. Experimentally, intramolecular hydrogen-bonds in an aqueous environment can be detected only indirectly, *e.g.* via NMR^{74, 117–120, 123, 147, 154, 182–189} or IR^{118, 121–123, 187, 190–194} spectroscopy. However, most studies to date have investigated model compounds (*e.g.* monosaccharide analogs) and only provided qualitative information on the existence or absence of a hydrogen-bond. Based on these studies, it appears that intramolecular hydrogen-bonds in aqueous solution can be broadly classified²¹ into three groups in order of decreasing occurrences^{118, 121, 123, 195} (Figure 1.2) : (i) inter-residue hydrogen-bonds across specific glycosidic linkages (*e.g.* in all glucose-based disaccharides⁵² involving a (1 \rightarrow *n*)-linkage with *n* = 2, 3 or 4); (ii) six-membered hydrogen-bonded “rings” between 1,3-*syn*-diaxial hydroxyl groups^{118, 122, 147, 185, 187, 190, 196–199}; (iii) five-membered hydrogen-bonded “rings” between *cis*-vicinal (axial-equatorial) hydroxyl groups^{190, 196}; (iv) five-membered hydrogen-bonded “rings” between *trans*-vicinal (diequatorial) hydroxyl groups^{190, 196}. Hydrogen-bonding interactions are strong in vacuum or in solvents of low polarity^{147, 187, 196, 199, 200}, but probably



represent rather weak conformational driving forces in aqueous environment^{190,201–203}. For example, although hydrogen-bonds between 1,3-*syn*-diaxial hydroxyl groups in fixed orientations are always observed in water^{118,185,198,204}, they are not always present when this orientation is conformer-dependent (*e.g.* when the “ring” involves one rotatable C-C bond^{109,110,187,190} or can only be formed in one chair conformation of a flexible six-membered ring system^{147,194,196,197,199,200}). In disaccharides⁵² it also appears that inter-residue hydrogen-bonds are only formed when the leading linkage conformation places specific hydroxyl groups in close proximity (rather than representing a driving force on the linkage conformation that promotes this proximity). However, a number of studies have also underlined the importance of cooperativity effects in intramolecular hydrogen-bonding for carbohydrates^{123,195,205–209} (*i.e.* the mutual reinforcement of individual hydrogen-bonds within the same molecule or saccharide chain).

4. *Solvation effects.* Solvation effects influence the conformational properties of carbohydrates in a very complex way^{210–212}. They consist of a combination of the following phenomena²¹: (i) dielectric screening of the intramolecular (mainly electrostatic, *i.e.* hydrogen-bonding) interactions (evidenced by numerous investigations in solvents of various polarities^{147,179,187,192,199,200,202,213}); (ii) competition between intramolecular and solute-solvent hydrogen bonds^{200,214,215} (in protic solvents); (iii) preferential solvation of sterically more accessible hydroxyl groups^{168,200,216,217}; (iv) specific interactions between solute hydroxyl groups and tightly bound solvent molecules or solvent bridges^{21,162,163,218–220}.

1.3 Molecular modeling of carbohydrates

1.3.1 Available molecular models

Theoretical methods (including glycoinformatics³⁹, conformational analysis^{22,23,28,32–38,40,41,46}, analysis of protein-carbohydrate interactions^{35,43,44} and multiscale modeling approaches^{42,45}) play an increasingly important role in glycobiology. Due to their remarkable spatial (individual atoms) and temporal (femtosecond) resolution, as well as their firm basis in the laws of (classical) microscopic physics, explicit-solvent (MD) simulations represent a key component in many theoretical approaches. A number of force-fields have been developed to represent interatomic interactions in such simulations, which can be broadly classified in three categories: (i) force-fields with the primary goal of reproducing properties in vacuum or together with an implicit representation of the solvent, *e.g.* HSEA¹⁷⁶, MM2²²¹, MM3^{222–224}, MM4^{149,150,169,225}, CFF (PEF²²⁶), QMFF²²⁷, CHARMM (QUANTA²²⁸, CHEAT^{229,230}) and AMBER (AMBER*^{231,232}); (ii) force fields with the primary goal of reproducing properties in solution together with an explicit representation of the solvent molecules, *e.g.* CHARMM (HGFB²³³, PHLB²³⁴, CSFF²³⁵, Reiling²³⁶ and PARM22-SU01²³⁷), AMBER (Homans²³⁸, Glennon^{239,240}, Gregurick²⁴¹, SPASIBA¹⁹³, GLY-CAM^{242–245}, AMB99C^{246,247} and Simmerling²⁴⁸), OPLS (AA²⁴⁹ and AA-SEI^{250,251}) and GRO-MOS (43A1^{252,253}, Kouwijzer^{254,255}, Ott²⁵⁶, Spieser²⁵⁷, Galema^{258,259} and 45A4²⁶⁰); (iii) coarse-grained force-fields with the primary goal of performing low-resolution simulations of large-scale systems or long-timescale processes, *e.g.* M3B^{261,262} and Martini²⁶³. For completeness, it should be mentioned that *ab initio* MD simulations of carbohydrates have also been performed^{106,264,265}.

Furthermore, the most commonly used force-fields have been discussed in recent reviews^{23,38,48,49}, and a number of more or less extensive force field comparisons (against vacuum^{227,266,267}, solid-state^{268,269} or solution^{266,270,271} data) have also been undertaken.

1.3.2 Molecular model used in this work

The following sections briefly describe the modeling approach relevant for the present thesis (atomistic explicit-solvent MD simulation) in terms of the main choices, approximations and methods it involves. There are four basic choices defining a molecular model : (i) choice of the *degrees of freedom* (elementary particles of the model); (ii) choice of the *interaction* (between these particles); (iii) choice of a *configuration generation* method (to produce a sequence of configurations for these particles); (iv) choice of the *boundary conditions* (defining global properties of this configuration sequence). The resulting combination should be adequate to evaluate the property or quantity of interest. Most of the times, however, a balance must be struck between the accuracy of the selected model and its computational cost for a given application. These four components are discussed in sequence below, focusing nearly exclusively on choices relevant for the type of modeling carried out in the present thesis.

Degrees of freedom

Modeling extends over a wide range of possible levels of resolution^{47,272}, *e.g.* nuclei and electrons (described by quantum mechanics), individual atoms (described by classical mechanics), beads representing atom or molecule groups (described by coarse-grained models) or volume elements in a continuous fluid (described by means of conservation and transport equations). In order to simulate large systems and reach a sufficient configuration sampling (*e.g.*, in MD, a sufficient timescale), it is often necessary to trade resolution for computational speed, *i.e.* to explicitly include only those degrees of freedom which (presumably) have the largest influence on the system and properties under investigation. An alternative is to use hybrid approaches, where different parts of the system are treated at different levels of resolution. Two typical examples of this kind are implicit-solvent models (where the solvent particles are not included explicitly, and the mean effect of solvation is introduced by a suitable modification of the interaction between the solute particles) or quantum-mechanics/molecular-mechanics approaches (where a small part of the system is treated quantum mechanically, while its surrounding is treated classically). The level of resolution adopted throughout this thesis is atomistic, including an explicit representation of the solvent molecules, the only exception being that aliphatic hydrogen atoms are treated implicitly using united CH_{*n*} (*n* = 1, 2, 3) atoms²⁷³.

Interaction function

When the elementary particles of the model are subatomic particles, a system configuration is defined by the corresponding many-particle wavefunctions and the interaction is governed by a Hamiltonian operator (encompassing both kinetic and potential energy operators, and acting on the wavefunction). When the elementary particles are classical, a system configuration is defined by the corresponding (generalized) coordinate and momentum vectors, and the interaction is governed by a Hamiltonian function (encompassing both kinetic and potential energy functions).

In such classical simulations, the Hamiltonian is usually formulated in terms of the Cartesian many-particle coordinate vector \mathbf{r} and the corresponding Cartesian momentum vector \mathbf{p} , as

$$\mathcal{H}(\mathbf{r}, \mathbf{p}) = \mathcal{K}(\mathbf{p}) + \mathcal{V}(\mathbf{r}) \quad . \quad (1.1)$$

The kinetic energy function \mathcal{K} is given by

$$\mathcal{K}(\mathbf{p}) = \frac{1}{2} \sum_{atm\ i} m_i^{-1} \mathbf{p}_i^2 \quad , \quad (1.2)$$

where the sum involves all atoms (*atm*) in the system, while m_i and \mathbf{p}_i are the mass and Cartesian momentum vector of atom i , respectively. The potential energy function \mathcal{V} is represented as a sum of terms, each of which involves one or more internal coordinates of the system along with parameters specific to the term. A prototypical functional form (as used in the GROMOS force-field^{252,253}) reads

$$\begin{aligned} \mathcal{V}(\mathbf{r}) = & \sum_{bnd\ n} \frac{1}{4} K_{b_n} [b_n^2 - b_{0_n}^2]^2 + \sum_{ang\ n} \frac{1}{2} K_{\theta_n} [\cos(\theta_n) - \cos(\theta_{0_n})]^2 \\ & + \sum_{imp\ n} \frac{1}{2} K_{\xi_n} [\xi_n - \zeta_{0_n}]^2 + \sum_{dih\ n} K_{\phi_n} [1 + \cos(m_n \phi_n - \delta_n)] \\ & + \sum_{prs\ ij} \left(\frac{C_{12,ij}}{r_{ij}^{12}} - \frac{C_{6,ij}}{r_{ij}^6} \right) + \sum_{prs\ ij} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad . \end{aligned} \quad (1.3)$$

The successive terms describe the *bond stretching* interaction along covalent bonds (*bnd*), the *bond-angle bending* interaction at covalent bond angles (*ang*), the deformation interactions at planar or tetrahedral sites, described here by means of (harmonic) *improper dihedral-angle bending* (*imp*) terms, the torsional interaction around bonds, described here by means of (trigonometric) *dihedral-angle torsion* (*dih*) terms, the non-bonded van der Waals interactions between all atom pairs (*prs*), and the electrostatic interactions between all atom pairs. The variables b_n , θ_n , ξ_n , ϕ_n and r_{ij} indicate the corresponding internal coordinates, namely the bond lengths, bond-angles, improper dihedral-angles, torsional dihedral-angles and pairwise distances, respectively (these internal coordinates are derived from the vector \mathbf{r} in a given configuration). The parameters b_0 , θ_0 and ξ_0 are the corresponding reference or ideal values. The parameters K_b , K_θ , K_ξ and K_ϕ indicate corresponding force constants. For torsional dihedral-angle terms, the parameters m_n and δ_n indicate the multiplicity and phase-shift cosine of a given term. Finally, for non-bonded terms, the parameters $C_{12,ij}$, $C_{6,ij}$ and q_i denote, respectively, the van der Waals (Pauli) repulsion parameter, the van der Waals (London) dispersion parameter and the partial charge of an atom, while ϵ_0 is the permittivity of vacuum.

1.3.3 Configuration generation

A given configuration generation method^{47,53,274,275} will generate a sequence of configurations of the molecular system obeying a certain ensemble probability distribution and a certain sequence ordering. One may distinguish between three main classes of methods: (*i*) *search methods* (e.g. genetic algorithms, random search, homology modeling), where both ensemble and sequence are

arbitrary; (ii) *sampling* methods, where the ensemble is uniquely defined by the interaction (and boundary conditions), but the sequence is arbitrary; (iii) *simulation methods*, where the ensemble is uniquely defined by the interaction (and boundary conditions) and the sequence obeys the "natural" (*i.e.* physical) dynamics of the system (*e.g.* explicit-solvent: MD; implicit-solvent: stochastic or Brownian dynamics). When the elementary particles of the model are subatomic (*i.e.* the interaction characterized by a Hamiltonian operator), the "natural" system dynamics (simulation method) corresponds to integrating the time-dependent Schrödinger equation for propagating the system configuration (wavefunction) in time. When these elementary particles are classical (*i.e.* the interaction characterized by a Hamiltonian function) and all explicitly included in the model, the "natural" system dynamics (simulation method) corresponds to integrating the classical (generalized coordinates: Hamiltonian or Lagrangian; Cartesian coordinates: Newtonian) equations of motion for propagating the system configuration (coordinate and momentum vectors) in time. Which properties of the molecular system can be calculated depends on the selected method. Only sampling and simulation methods will permit the calculation of the *thermodynamical* properties of the system. Only simulation methods will permit the calculation of the both thermodynamical and *dynamical* properties of the system. However, the calculation will only be meaningful if a statistically representative set of configurations has been generated, which can be difficult for complex molecular systems due to the presence of numerous minima and barriers on the potential energy surface. In this case, it may be advantageous to sacrifice the dynamical information and use enhanced-sampling (rather than simulation) methods.

The configuration generation method adopted throughout this thesis is explicit-solvent MD where the classical (Newtonian) equations of motions are integrated in time for the Hamiltonian defined in Eqs (1.2) and (1.3). In this approach, one can calculate the force \mathbf{f}_i on a particle i is calculated at a given instant as the negative of the first derivative of the potential energy with respect to its position

$$\mathbf{f}_i = -\frac{\partial}{\partial \mathbf{r}_i} \mathcal{V}(\mathbf{r}). \quad (1.4)$$

Newton's equation of motion connects this force to the second derivative of the particle position as

$$\frac{d^2}{dt^2} \mathbf{r}_i = \frac{1}{m_i} \mathbf{f}_i. \quad (1.5)$$

A common numerical algorithm to integrate this equation in time is the leap-frog algorithm²⁷⁶. In this scheme, the numerical propagation of positions and velocities is based on a finite time step Δt according to

$$\mathbf{v}_i\left(t + \frac{1}{2}\Delta t\right) = \mathbf{v}_i\left(t - \frac{1}{2}\Delta t\right) + \frac{1}{m_i} \mathbf{f}_i(t) \Delta t \quad (1.6)$$

and

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i\left(t + \frac{1}{2}\Delta t\right). \quad (1.7)$$

The integration of Eqs (1.6) and (1.7), while maintaining the kinetic energy coupled to a reference temperature bath (see below), leads to a trajectory of the system configuration in the time that obeys a well defined (Boltzmann) probability distribution and represents the natural dynamics of the system (simulation method).

1.3.4 Boundary conditions

The term boundary condition refers to a global constraint imposed to the system and generated configurations during the simulation. One may distinguish between: (i) *spatial* boundary conditions (spatial construction of the simulated sample); (ii) *thermodynamical* boundary conditions (average thermodynamical properties of the simulated ensemble); (iii) *geometrical* boundary conditions (constraints or restrictions imposed to specific internal coordinates).

Spatial boundary conditions^{47,277,278} determine the shape, surroundings and boundary of the simulated sample. To overcome finite-size effects (dominated by long-range electrostatic interaction) and surface effects (caused by the possible presence of an interface with vacuum), periodic boundary conditions are most commonly used in classical MD simulation. The system of interest is then encompassed within a computational box of space filling shape, which is surrounded by identical periodic copies of itself (thereby eliminating any interface with vacuum). In this case, the non-bonded interactions can be treated either as periodic (lattice-sum methods) or truncated (to a given cutoff distance; so as to reduce both computational costs and possible periodicity artifacts). Common box shapes are the cube, rectangle, truncated octahedron, hexagonal prism and triclinic box. In all simulations reported in the present thesis, periodic boundary conditions were applied, using either a rectangular or a truncated octahedral box.

Thermodynamical boundary conditions^{279,280} determine what the system is allowed to exchange with its surrounding. A system may be closed (constant number of particles N) or grand (constant average chemical potential μ at equilibrium), isochoric (constant volume V) or isobaric (constant pressure P at equilibrium), and adiabatic (constant energy E) or isothermal (constant average temperature T at equilibrium). Nearly all simulations reported in the present thesis correspond to the (closed) isothermal-isobaric (NPT) or canonical (NVT) ensembles, since they correspond to the most common experimental conditions. By default, however, MD simulations sample the microcanonical (NVE) ensemble. The enforcement of a constant (average) pressure or/and temperature requires a suitable modification of the equations of motion, *e.g.* by means of the Berendsen barostat or/and thermostat algorithm²⁸¹.

Geometrical boundary conditions include the use of bond-length constraints²⁸² (as an alternative to harmonic bonds, providing a better approximation to the quantum-mechanical behavior of stiff harmonic oscillators) as well as experimentally-derived boundary conditions^{283,284} (used to enforce the reproduction of experimentally derived information by the simulated ensemble). Only the former type of geometrical boundary conditions (bond-length constraints) was applied in the simulations reported in the present thesis.

1.4 References

- [1] Ramesh, H.P. & Tharanathan, R.N. *Crit. Rev. Biotech.* **23** 149-173 (2003)
- [2] Tokstoguzov, V. *Origins of Life and Evolution of the Biosphere* **34** 571-597 (2004)
- [3] Stern, R. & Jedrzejak, M.J. *Chem. Rev.* **108** 5061-5085 (2008)
- [4] Nelson, D.L. & Cox, M.M. *Lehninger Principles of Biochemistry Volume . Edition 5th.* Palgrave Macmillan (2008)

- [5] Berg, J.M., Tymoczko, J.L. & Stryer, L. Biochemistry Volume . Edition 5th. W.H. Freeman, Co. Ltd. (2007)
- [6] Dwek, R.A. *Chem. Rev.* **96** 683-720 (1996)
- [7] Lehmann, J. Carbohydrates structure and biology. Georg Thieme Verlag, Stuttgart, Germany. (1998)
- [8] Varki, A., Cummings, R., Esko J., Freeze, H., Hart & G., Marth J. Essentials of glycobiology. CSHL Press, New York (1999)
- [9] Rüdiger, H., Siebert, H.-C., Solís, D., Jiménez-Barbero, J., Romero, A., von der Lieth, C.-W., Diaz-Mauriño, T. & Gabius, H.-J. *Curr. Med. Chem.* **7** 389-416 (2000)
- [10] Gabius, H.-J., André, S., Kaltner, H. & Siebert, H.-C. *Biochim. Biophys. Acta* **1572** 165-177 (2002)
- [11] Tunnacliffe, A. & Lapinski, J. *Phil. Trans. R. Soc. London B* **358** 1755-1771 (2003)
- [12] Lowe, J.B. & Marth, J.D. *Annu. Rev. Biochem.* **72** 643-691 (2003)
- [13] Gabius, H.-J., Siebert, H.-C., André, S., Jiménez-Barbero, J. & Rüdiger, H. *Chem-BioChem* **5** 740-764 (2004)
- [14] Robyt, J.F. Essentials of carbohydrate chemistry. Springer Verlag, New York, U.S.A. (1998)
- [15] Walter, R.H. Polysaccharide association structures in food. Marcel Dekker Inc., New York, U.S.A. (1998)
- [16] Davis, A.P. & Wareham, R.S. *Angew. Chem. Int. Ed.* **38** 2978-2996 (1999)
- [17] Crowe, J.H., Crowe, L.M., Wolkers, W.F., Oliver, A.E., Ma, X., Auh, J.-H., Tang, M., Zhu, S., Norris, J. & Tablin, F. *Integr. Comp. Biol.* **45** 810-820 (2005)
- [18] Reeves, R.E. *Annu. Rev. Biochem.* **27** 15-34 (1958)
- [19] Laine, R.A. *Glycobiology* **4** 759-767 (1994)
- [20] Hricovini, M. *Curr. Med. Chem.* **11** 2565-2583 (2004)
- [21] Kräutler, V., Müller, M. & Hünenberger, P.H. *Carbohydr. Res.* **342** 2097-2124 (2007)
- [22] Meyer, B. Conformational aspects of oligosaccharides. In: *Topics in current chemistry. Volume 154*. Thiem, J., Ed. Springer-Verlag, Berlin, Germany. pp 141-208 (1990)
- [23] Woods, R.J. *Curr. Opin. Struct. Biol.* **5** 591-598 (1995)
- [24] Rao, V.S.R., Qasba, P.K., Balaji, P.V. & Chandrasekaran, R. Conformation of carbohydrates. Harwood Academic Publishers, Amsterdam, The Netherlands. (1998)
- [25] Doubet, S., Bock, K., Smith, D., Darvill, A. & Albersheim, P. *TIBS* **14** 475-477 (1989)

- [26] Rice, K.G., Wu, P., Brand, L. & Lee, Y.C. *Curr. Opin. Struct. Biol.* **3** 669-674 (1993)
- [27] Pérez, S. & Kouwijzer, M. Shapes and interactions of polysaccharide chains. In: *Carbohydrates: Structures, Dynamics and Syntheses*. Finch, P., Ed. Kluwer Academic Press pp 258-293 (1999)
- [28] Wormald, M.R., Petrescu, A.J., Pao, Y.-L., Glithero, A., Elliott, T., *Chem. Rev.* **102** 371-386 (2002)
- [29] Agrawal, P.K. *Phytochemistry* **31** 3307-3330 (1992)
- [30] Jiménez-Barbero, J., Asensio, J.L., Cuevas, G., Canales, A., Fernández-Alonso, M.C. & Cañada, F.J. *Biocatalysis and Biotransformation* **24** 13-22 (2006)
- [31] Pérez, S. & Delage, M.-M. *Methods Enzymology* **203** 510-556 (1991)
- [32] French, A.D. & Brady, J.W. *ACS Symposium Series* **430** 1-19 (1990)
- [33] Madsen, L.J., Ha, S.N., Tran, V.H. & Brady, J.W. *ACS Symposium Series* **430** 69-90 (1990)
- [34] Pérez, S. & Delage, M.-M. *Carbohydr. Res.* **212** 253-259 (1991)
- [35] Pérez, S., Meyer, C. & Imberty, A. Practical tools for molecular modeling of complex carbohydrates and their interactions with proteins. In: *Modelling of biomolecular structure and mechanism*. Pullman, A., Jortner, J. & Pullman, B., Eds.; ; Kluwer Academic Publishers, The Netherlands. pp 425-454 (1995)
- [36] Pérez, S., Kouwijzer, M., Mazeau, K. & Engelsens, S.B. *J. Mol. Graphics* **14** 307-321 (1996)
- [37] Imberty, A. *Curr. Opin. Struct. Biol.* **7** 617-623 (1997)
- [38] Woods, R.J. *Glycoconjugate J.* **15** 209-216 (1998)
- [39] Pérez, S. & Mulloy, B. *Curr. Opin. Struct. Biol.* **15** 517-524 (2005)
- [40] Engelsens, S.B., Cros, S., Mackie, W. & Pérez, S. *Biopolymers* **39** 417-433 (1996)
- [41] Imberty, A. & Pérez, S. *Chem. Rev.* **100** 4567-4588 (2000)
- [42] Queyroy, S., Neyertz, S., Brown, D. & Müller-Plathe, F. *Macromolecules* **37** 7338-7350 (2004)
- [43] Laederach, A. & Reilly, P.J. *Proteins: Struct. Funct. Bioinf.* **60** 591-597 (2005)
- [44] Dyekjaer, J.D. & Woods, R.J. *ACS Symposium Series* **930** 203-219 (2006)
- [45] Limbach, H.J. & Kremer, K. *Trends Food Sci. Technol.* **17** 215-219 (2006)
- [46] Gerbst, A.G., Grachev, A.A., Shashkov, A.S. & Nifantiev, N.E. *Russian J. Bioorg. Chem.* **33** 24-37 (2007)

- [47] van Gunsteren, W.F. & Berendsen, H.J.C. *Angew. Chem. Int. Ed.* **29** 992-1023 (1990)
- [48] MacKerell, D.A., *J. Comput. Chem.* **25** 1584-1604 (2004)
- [49] Tschampel, S.M., Kirschner, K.N. & Woods, R.J. *ASC Symposium Series* **930** 235-257 (2006)
- [50] McNaught, A.D. *Pure Appl. Chem.* **68** 1919-2008 (1996)
- [51] McNaught, A.D. *J. Carbohydr. Chem.* **16** 1191-1280 (1996)
- [52] Pereira, C.S., Kony, D., Baron, R., Müller, M., van Gunsteren, W.F. & Hünenberger, P.H. *Biophys. J.* **90** 4337-4344 (2006)
- [53] Hansen, H.S. & Hünenberger, P.H. *J. Comput. Chem.*, in press (2009)
- [54] Pigman, W. & Isbell, H.S. *Adv. Carbohydr. Chem. Biochem.* **23** 11-57 (1968)
- [55] Angyal, S.J. *Angew. Chem. Int. Ed.* **8** 157-166 (1969)
- [56] Angyal, S.J. & Pickles, V.A. *Aust. J. Chem.* **25** 1695-1710 (1972)
- [57] Bentley, R. *Annu. Rev. Biochem.* **41** 953-996 (1972)
- [58] Angyal, S.J. *Adv. Carbohydr. Chem. Biochem.* **42** 15-69 (1984)
- [59] Angyal, S.J. *Adv. Carbohydr. Chem. Biochem.* **42** 19-35 (1991)
- [60] Ma, B., Schaefer III, H.F. & Allinger, N.L. *J. Am. Chem. Soc.* **120** 3411-3422 (1998)
- [61] Zhu, Y., Zajicek, J. & Serianni, A.S. *J. Org. Chem.* **66** 6244-6251 (2001)
- [62] Lewis, B.E., Choytun, N., Schramm, V.L. & Bennet, A.J. *J. Am. Chem. Soc.* **128** 5049-5058 (2006)
- [63] Danilova, V.A. & Krivdin, L.B. *Russ. J. Org. Chem.* **40** 57-62 (2004)
- [64] DeMatteo, M.P., Snyder, N.L., Morton, M., Baldisseri, D.M., Hadad, C.M. & Peczu, M.W. *J. Org. Chem.* **70** 24-38 (2004)
- [65] Pakulski, Z. *Polish J. Chem.* **80** 1293-1326 (2006)
- [66] Grindley, T.B. & Gulasekharan, V. *J. Chem. Soc. Chem. Commun.* **23** 1073-1074 (1978)
- [67] Snyder, J.R. & Serianni, A.S. *J. Org. Chem.* **51** 2694-2702 (1986)
- [68] Lee, C.Y., Acree, T.E. & Schallenberger, R.S. *Carbohydr. Res.* **9** 356-360 (1969)
- [69] Acree, T.E., Schallenberger, R.S., Lee, Y. & Einset, J.W. *Carbohydr. Res.* **10** 355-360 (1969)
- [70] Wertz, P.W., Garver, J.C. & Anderson, L. *J. Am. Chem. Soc.* **103** 3916-3922 (1981)

- [71] Serianni, A.S., Pierce, J., Huang, S.-G. & Barker, R. *J. Am. Chem. Soc.* **104** 4037-4044 (1982)
- [72] Snyder, J.R., Johnston, E.R. & Serianni, A.S. *J. Am. Chem. Soc.* **111** 2681-2687 (1989)
- [73] Behrends, R. & Kaatze, U. *ChemPhysChem* **6** 1133-1145 (2005)
- [74] Bubb, W.A. *Concepts Magn. Reson. A* **19** 1-19 (2003)
- [75] Hayward, L.D. & Angyal, S.J. *Carbohydr. Res.* **53** 13-20 (1977)
- [76] Reeves, R.E. *J. Am. Chem. Soc.* **72** 1499-1506 (1950)
- [77] Jensen, F.R. & Bushweller, C.H. *Adv. Alicyclic. Chem.* **3** 139-194 (1971)
- [78] Mäler, L., Widmalm, G. & Kowalewski, J. *J. Phys. Chem.* **100** 17103-17110 (1996)
- [79] Kleinpeter, E. *Adv. Heterocyclic Chem.* **86** 41-127 (2004)
- [80] Sanderson, P.N., Huckerby, T.N. & Nieduszynski, I.A. *Glycoconjugate J.* **2** 109-120 (1985)
- [81] Cano, F.H., Foces-Foces, C., Jimenez-Barbero, J., Bernabe, M. & Martin-Lomas, M. *Carbohydr. Res.* **145** 319-327 (1986)
- [82] Ragazzi, M., Ferro, D.R. & Provasoli, A. *J. Comput. Chem.* **7** 105-112 (1986)
- [83] Ferro, D.R., Provasoli, A., Ragazzi, M., Torri, G., Casu, B., Gatti, G., Jacquinet, J.-C., Sinay, P., Petitou, M. & Choay, J. *J. Am. Chem. Soc.* **108** 6773-6778 (1986)
- [84] Sanderson, P.N., Huckerby, T.N. & Nieduszynski, I.A. *Biochem. J.* **243** 175-181 (1987)
- [85] Giuliano, R.M., Bryan, R.F., Hartley, P., Peckler, S. & Woode, M.K. *Carbohydr. Res.* **191** 1-11 (1989)
- [86] Ferro, D.R., Provasoli, A., Ragazzi, M., Casu, B., Torri, G., Bossenec, V., Perly, B., Sinay, P., Petitou, M. & Choay, J. *Carbohydr. Res.* **195** 157-167 (1990)
- [87] Forster, M.J. & Mulloy, B. *Biopolymers* **33** 575-588 (1993)
- [88] Horita, D.A., Hadjuk, P.J. & Lerner, L.E. *Glycoconjugate J.* **14** 691-696 (1997)
- [89] Coxon, B. & Reynolds, R.C. *Carbohydr. Res.* **331** 461-467 (2001)
- [90] Uccello-Barretta, G., Sicoli, G., Balzano, F. & Salvadori, P. *Carbohydr. Res.* **338** 1103-1107 (2003)
- [91] Roslund, M.U., Klika, K.D., Lehtilä, R.L., Tähtinen, P., Silanpää, R. & Leino, R. *J. Org. Chem.* **69** 18-25 (2004)
- [92] Hakkarainen, B., Fujita, K., Immel, S., Kenne, L. & Sandström, C. *Carbohydr. Res.* **340** 1539-1545 (2005)

- [93] Franks, F., Lillford, P.J. & Robinson, G. *J. Chem. Soc. Farad. Trans. I* **85** 2417-2426 (1989)
- [94] Polacek, R., Stenger, J. & Kaatze, U. *J. Chem. Phys.* **116** 2973-2982 (2002)
- [95] Stenger, J., Cowman, M.K., Eggers, F., Eyring, E.M., Kaatze, U. & Petrucci, S. *J. Phys. Chem. B* **104** 4782-4790 (2000)
- [96] Hagen, R. & Kaatze, U. *J. Chem. Phys.* **120** 9656-9664 (2004)
- [97] Hajduk, P.J., Horita, D.A. & Lerner, L.E. *J. Am. Chem. Soc.* **115** 9196-9201 (1993)
- [98] Haasnoot, C.A.G., De Leeuw, F.A.A.M. & Altona, C. *Tetrahedron* **36** 2783-2792 (1980)
- [99] Tvaroška, I., Hricovíni, M. & Petráková, E. *Carbohydr. Res.* **189** 359-362 (1989)
- [100] Bock, K. & Duus, J.O. *J. Carbohydr. Chem.* **13** 513-543 (1994)
- [101] Tvaroška, I. & Gajdoš, J. *Carbohydr. Res.* **271** 151-162 (1995)
- [102] Bose, B., Zhao, S., Stenutz, R., Cloran, F., Bondo, P.B., Bondo, G., Hertz, B., Carmichael, I. & Serianni, A.S. *J. Am. Chem. Soc.* **120** 11158-11173 (1998)
- [103] Stenutz, R., Carmichael, I., Widmalm, G. & Serianni, A.S. *J. Org. Chem.* **67** 949-958 (2002)
- [104] Thibaudeau, C., Stenutz, R., Hertz, B., Klepach, T., Zhao, S., Wu, Q., Carmichael, I. & Serianni, A.S. *J. Am. Chem. Soc.* **126** 15668-15685 (2004)
- [105] Kraszni, M., Szakács, Z. & Noszál, B. *Anal. Bioanal. Chem.* **378** 1149-1163 (2004)
- [106] Suzuki, T., Kawashima, H. & Sota, T. *J. Phys. Chem. B* **110** 2405-2418 (2006)
- [107] Nishida, Y., Ohruí, H. & Meguro, H. *Tetrahedron Lett.* **25** 1575-1578 (1984)
- [108] Nishida, Y., Hori, H., Ohruí, H. & Meguro, H. *J. Carbohydr. Chem.* **7** 239-250 (1988)
- [109] Abraham, R.J., Chambers, E.J. & Thomas, W.A. *Magn. Reson. Chem.* **30** S60-S65 (1992)
- [110] Abraham, R.J., Chambers, E.J. & Thomas, W.A. *Magn. Reson. Chem.* **32** 248-254 (1994)
- [111] Brochier-Salon, M.-C. & Morin, C. *Magn. Reson. Chem.* **38** 1041-1042 (2000)
- [112] Tvaroška, I., Taravel, F.R., Utille, J.P. & Carver, J.P. *Carbohydr. Res.* **337** 353-367 (2002)
- [113] Hori, H., Nishida, Y., Ohruí, H. & Meguro, H. *J. Carbohydr. Chem.* **9** 601-618 (1990)
- [114] Behrends, R., Cowman, M.K., Eggers, F., Eyring, E.M., Kaatze, U., Majewski, J., Petrucci, S., Richmann, K.-H. & Riech, M. *J. Am. Chem. Soc.* **119** 2182-2186 (1997)
- [115] Poppe, L. *J. Am. Chem. Soc.* **115** 8241-8426 (1993)

- [116] Dais, P. & Marcheussault, R.H. *Macromolecules* **24** 4611-4614 (1991)
- [117] Duus, J.O., Gotfredsen, C.H. & Bock, K. *Chem. Rev.* **100** 4589-4614 (2000)
- [118] Hawley, J., Bampos, N., Aboitiz, N., Jiménez-Barbero, J., López de la Paz, M., Sanders, J.K.M., Carmona, P. & Vicent, C. *Eur. J. Org. Chem.* **12** 1925-1936 (2002)
- [119] Fabri, D., Williams, M.A.K. & Halstead, T.K. *Carbohydr. Res.* **340** 889-905 (2005)
- [120] Sandström, C. & Kenne, L. *ACS Symposium Series* **930** 114-132 (2006)
- [121] Zhbakov, R.G. *J. Mol. Struct.* **270** 523-539 (1992)
- [122] Carmona, P., Molina, M., Aboitiz, N. & Vicent, C. *Biopolymers* **67** 20-25 (2002)
- [123] López de la Paz, M., Ellis, G., Pérez, M., Perkins, J., Jiménez-Barbero, J. & Vicent, C. *Eur. J. Org. Chem.* **5** 840-855 (2002)
- [124] Lemieux, R.U. & Koto, S. *Tetrahedron* **30** 1933-1944 (1974)
- [125] Jarvis, M. C. *Carbohydr. Res.* **259** 311-318 (1994)
- [126] Abbate, S., Biancardi, L., Longhi, G. & Conti, G. *Carbohydr. Res.* **300** 59-66 (1997)
- [127] Zhang, P., Klymachyov A.N., Brown S., Ellington, J.G. & Grandinetti, P.J. *Solid State Nuclear Magnetic Resonance* **12** 221-225 (1998)
- [128] Best, R.B., Jackson, G.E. & Naidoo, K.J. *J. Phys. Chem. B* **105** 4742-4751 (2001)
- [129] Striegel, A.M. *J. Am. Chem. Soc.* **125** 4146-4148 (2003)
- [130] da Silva, C.O. & Nascimento, M.A.C. *Theor. Chem. Acc.* **112** 342-348 (2004)
- [131] Choi, Y. & Jung, S. *Carbohydr. Res.* **340** 2550-2557 (2005)
- [132] Seo, J.-A., Kwon, H.-J., Kook Kim, H. & Hwang, Y.-H. *Carboh. Res.* **343** 660-667 (2008)
- [133] Roslund, M.U., Tähtinen, P., Niemitz, M., *Carbohydr. Res.* **343** 101-112 (2007)
- [134] Pérez, S. & Vergelati, C. *Biopolymers* **24** 1809-1822 (1985)
- [135] Braccini, I., Grasso, R.P. & Pérez, S. *Carbohydr. Res.* **317** 119-130 (1999)
- [136] Peric, L., Pereira, C.S., Perez, S. & Hünenberger, P.H. *Molecular Simulation* **34** 421-446 (2008)
- [137] Rees, D.A. & Welsh, E.J. *Angew. Chem. Int. Ed.* **16** 214-224 (1977)
- [138] Dwek, R.A., Edge, C.J., Harvey, D.J. & Wormald, M.R. *Annu. Rev. Biochem.* **62** 65-100 (1993)
- [139] Li, Y., Krilov, G. & Berne, B.J. *J. Phys. Chem. B.* **110** 13256-13263 (2006)

- [140] Hassel, O. & Ottar, B. *Acta Chem. Scand.* **1** 929-943 (1947)
- [141] Angyal, S.J. & McHugh, D.J. *Chem. Ind.* **41** 1147-1148 (1956)
- [142] Eliel, E.L. & Lukach, C.A. *J. Am. Chem. Soc.* **79** 5986-5992 (1957)
- [143] Angyal, S.J., Pickles, V.A. & Ahluwalia, R. *Carbohydr. Res.* **1** 365-370 (1966)
- [144] Angyal, S.J. *Aust. J. Chem.* **21** 2737-2746 (1968)
- [145] Augé, J. & David, S. *Tetrahedron* **40** 2101-2106 (1984)
- [146] Dowd, M.K., French, A.D. & Reilly, P.J. *Aust. J. Chem.* **49** 327-335 (1996)
- [147] de Oliveira, P.R. & Rittner, R. *J. Mol. Struct.* **743** 69-72 (2005)
- [148] Weldon, A.J., Vickrey, T.L. & Tschumper, G.S. *J. Phys. Chem. A* **109** 11073-11079 (2005)
- [149] Lii, J.-H., Chen, K.-H., Grindley, T.B. & Allinger, N.L. *J. Comput. Chem.* **24** 1490-1503 (2003)
- [150] Lii, J.-H., Chen, K.-H. & Allinger, N.L. *J. Comput. Chem.* **24** 1504-1513 (2003)
- [151] Corey, E.J. & Feiner, N.F. *J. Org. Chem.* **45** 757-764 (1980)
- [152] Corey, E.J. & Feiner, N.F. *J. Org. Chem.* **45** 765-780 (1980)
- [153] Alabugin, I.V. *J. Org. Chem.* **65** 3910-3919 (2000)
- [154] Lewis, B.R. & Schramm, V.L. *J. Am. Chem. Soc.* **123** 1327-1336 (2001)
- [155] Cortés-Guzmán, F., Hernández-Trujillo, J. & Cuevas, G. *J. Phys. Chem. A* **107** 9253-9256 (2003)
- [156] Taddei, F. & Kleinpeter, E. *J. Mol. Struct.* **683** 29-41 (2004)
- [157] Jeffrey, G.A. *Acta Crystallogr. B* **46** 89-103 (1990)
- [158] Abraham, R.J., Chambers, E.J. & Thomas, W.A. *Carbohydr. Res. C* **226** 1-5 (1992)
- [159] Jensen, F.R., Bushweller, C.H. & Beck, B.H. *J. Am. Chem. Soc.* **91** 344-351 (1969)
- [160] Kleinpeter, E., Rolla, N., Koch, A. & Taddei, F. *J. Org. Chem.* **71** 4393-4399 (2006)
- [161] Petillo, P.A. & Lerner, L.A. *ACS Symposium Series* **539** 156-175 (1993)
- [162] Praly, J.-P. & Lemieux, R.U. *Can. J. Chem.* **65** 213-223 (1987)
- [163] Walkinshaw, M.D. *J. Chem. Soc. Perkin. Trans.* **2** 1903-1906 (1987)
- [164] Edward, J.T. *Chem. Ind.* **36** 1102-1104 (1955)
- [165] Inagaki, S., Iwase, K. & Mori, Y. *Chem. Lett.* **3** 417-420 (1986)

- [166] Juaristi, E. & Cuevas, G. *Tetrahedron* **48** 5019-5087 (1992)
- [167] Graczyk, P.P. & Mikolajczyk, M. *Topics Stereochem.* **21** 159-349 (1994)
- [168] Tvaroška, I. & Carver, J.P. *Carbohydr. Res.* **309** 1-9 (1998)
- [169] Lii, J.-H., Chen, K.-H., Durkin, K. & Allinger, N.L. *J. Comput. Chem.* **24** 1473-1489 (2003)
- [170] Wolfe, S. *Acc. Chem. Res.* **5** 102-111 (1970)
- [171] Epiotis, N.D., Sarkanen, S., Bjorkquist, D., Bjorkquist, L. & Yates, R. *J. Am. Chem. Soc.* **96** 4075-4084 (1974)
- [172] Subbotin, O.A. & Sergeyev, N.M. *Anal. Chem.* **48** 545-546 (1976)
- [173] Zefirov, N.S., Samoshin, V.V., Subbotin, O.A. & Baranekov, V.I. *Tetrahedron* **34** 2953-2959 (1978)
- [174] Juaristi, E. & Antúnez, S. *Tetrahedron* **48** 5941-5950 (1992)
- [175] Pérez, S. & Marchessault, R.H. *Carbohydr. Res.* **65** 114-120 (1978)
- [176] Thogersen, H., Lemieux, R.U., Bock, K. & Meyer, B. *Can. J. Chem.* **60** 44-57 (1982)
- [177] Mayato, C., Dorta, R.L. & Vázquez, J.T. *Tetrahedron: Asymmetry* **15** 2385-2397 (2004)
- [178] de Vries, N.K. & Buck, H.M. *Carbohydr. Res.* **165** 1-16 (1987)
- [179] de Vries, N.K. & Buck, H.M. *Rec. Trav. Chim. Pays-Bas* **106** 453-460 (1987)
- [180] Wilberg, K.B., Hammer, J.D., Castejon, H., Bailey, W.F., DeLeon, E.L. & Jarret, R.M. *J. Org. Chem.* **64** 2085-2095 (1999)
- [181] Alabugin, I.V., Manoharan, M. & Zeidan, T.A. *J. Am. Chem. Soc.* **125** 14014-14031 (2003)
- [182] Harvey, J.M., Symons, M.C.R. & Naftalin, R.J. *Nature* **261** 435-436 (1976)
- [183] Harvey, J.M. & Symons, M.C.R. *J. Solut. Chem.* **7** 571-586 (1978)
- [184] Pfeffer, P.E., Valentine, K.M. & Parrish, F.W. *J. Am. Chem. Soc.* **101** 1265-1274 (1979)
- [185] Reuben, J. *J. Am. Chem. Soc.* **106** 6180-6186 (1984)
- [186] Adams, B. & Lerner, L. *J. Am. Chem. Soc.* **114** 4827-4829 (1992)
- [187] Beeson, C., Pham, N., Shipp Jr, G. & Dix, T.A. *J. Am. Chem. Soc.* **115** 6803-6812 (1993)
- [188] Bolvig, S. & Hansen, P.E. *Curr. Org. Chem.* **4** 19-54 (2000)
- [189] Bosco, M., Picotti, F., Radoicovich, A. & Rizzo, R. *Biopolymers* **53** 272-280 (2000)
- [190] Kuhn, L.P. *J. Am. Chem. Soc.* **74** 2492-2499 (1952)

- [191] Casu, B., Reggiani, M., Gallo, G.G. & Vigevani, A. *Tetrahedron* **22** 3061-3083 (1966)
- [192] Lemieux, R.U. & Brewer, J.T. *Adv. Chem.* **117** 121-146 (1973)
- [193] Durier, V., Tristram, F. & Vergoten, G. *Carbohydr. Res.* **395** 81-90 (1997)
- [194] de Oliveira, P.R. & Rittner, R. *Spectrochimica Acta A* **61** 1737-1745 (2005)
- [195] Simperler, A., Watt, S.W., Bonnet, P.A., Jones, W. & Motherwell, W.D.S. *Cryst. Eng. Commun.* **8** 589-600 (2006)
- [196] Bacon, J.F., van der Maas, J.H., Dixon, J.R., George, W.O. & McIntyre, P.S. *Spectrochimica Acta A* **45** 1313-1318 (1989)
- [197] Abraham, R.J., Chambers, E.J. & Thomas, W.A. *J. Chem. Soc. Perkin Trans.* **2** 1061-1066 (1993)
- [198] Luque, F.J., López, J.M., López de la Paz, M., Vicent, C. & Orozco, M. *J. Phys. Chem. A* **102** 6690-6696 (1998)
- [199] de Oliveira, P.R. & Rittner, R. *Spectrochimica Acta A* **62** 30-37 (2005)
- [200] Tominaga, T., Tenma, S. & Watanabe, H. *J. Chem. Soc. Farad. Trans* **92** 1863-1867 (1996)
- [201] Roën, A., Padrón, J.I. & Vásquez, J.T. *J. Org. Chem.* **68** 4615-4630 (2002)
- [202] Reuben, J. *J. Am. Chem. Soc.* **107** 5867-5870 (1985)
- [203] Bernet, B. & Vasella, A. *Helv. Chim. Acta* **83** 995-1021 (2000)
- [204] Abraham, R.J., Byrne, J.J., Griffiths, L. & Koniotou, R. *Magn. Reson. Chem.* **43** 611-624 (2005)
- [205] Tse, Y.-C. & Newton, M.D. *J. Am. Chem. Soc.* **99** 611-613 (1977)
- [206] Jeffrey, G.A. & Mitra, J. *Acta Crystallogr.* **B39** 469-480 (1983)
- [207] Carcabal, P., Jockusch, R.A., Hünig, I., Snoek, L.C., Kroemer, R.T., Davis, B.G., Gambin, D.P., Compagnin, I., Oomens, J. & Simons, J.P. *J. Am. Chem. Soc.* **127** 11414-11425 (2005)
- [208] Simons, J.P., Jockusch, R.A., Carcabal, P., Hünig, I., Kroemer, R.T., McLeod, N.A. & Snoek, L.C. *Int. Rev. Phys. Chem.* **24** 489-531 (2005)
- [209] Dashnau, J.L., Sharp, K.A. & Vanderkooi, J.M. *J. Phys. Chem. B* **109** 24152-24159 (2005)
- [210] Franks, F. *Cryobiology* **20** 335-345 (1983)
- [211] Franks, F. *Pure Appl. Chem.* **59** 1189-1202 (1987)
- [212] Pérez, S. Polysaccharide interactions with water. In: *Water and biological macromolecules*. Westhof, E., Ed. Macmillan Press pp 295-320 (1993)

- [213] Onodera, K., Hirano, S. & Masuda, F. *Carbohydr. Res.* **7** 27-37 (1968)
- [214] Liu, Q. & Brady, J.W. *J. Am. Chem. Soc.* **118** 12276-12286 (1996)
- [215] Costa, F.S., Eusébio, M.E., Redinha, J.S. & Leitao, M.L.P. *J. Chem. Thermodyn.* **32** 311-317 (2000)
- [216] Kabayama, M.A. & Patterson, D. *Can. J. Chem.* **36** 563-573 (1958)
- [217] Bonnet, A., Chisholm, J., Motherwell, W.D.S. & Jones, W. *Cryst. Eng. Commun.* **7** 71-75 (2005)
- [218] Chalikian, T.V. *J. Phys. Chem. B* **102** 6921-6926 (1998)
- [219] Engelsen, S.B., Monteiro, C., Hervé de Penhoat, C. & Pérez, S. *Biophys. Chem.* **93** 103-127 (2001)
- [220] Almond, A. *Carbohydr. Res.* **340** 907-920 (2005)
- [221] Allinger, N.L. *J. Am. Chem. Soc.* **25** 8127-8134 (1977)
- [222] Allinger, N.L. *J. Am. Chem. Soc.* **111** 8551-8566 (1989)
- [223] Allinger, N.L. *J. Am. Chem. Soc.* **111** 8566-8575 (1989)
- [224] Allinger, N.L. *J. Am. Chem. Soc.* **111** 8576-8582 (1989)
- [225] Allinger, N.L., Chen, K.-H., Lii, J.-H., *J. Comput. Chem.* **24** 1447-1472 (2003)
- [226] Rasmussen, K. *J. Mol. Struct.* **395** 91-106 (1997)
- [227] Hwang, M.-J., Ni, X., Waldman, M., Ewig, C.S. & Hagler, A.T. *Biopolymers* **45** 435-468 (1998)
- [228] Momany, F.A. & Rone, R. *J. Comput. Chem.* **13** 888-900 (1992)
- [229] Grootenhuis, P.D.J. & Haasnoot, C.A.G. *Mol. Simul.* **10** 75-95 (1993)
- [230] Kouwijzer, M.L.C.E. & Grootenhuis, P.D.J. *J. Phys. Chem.* **99** 13426-13436 (1995)
- [231] Senderowitz, H., Parish, C. & Still, W.C. *J. Am. Chem. Soc.* **118** 2078-2086 (1996)
- [232] Senderowitz, H. & Still, W.C. *J. Org. Chem.* **62** 1427-1438 (1997)
- [233] Ha, S.N., Giammona, A., Field, M. & Brady, J.W. *Carbohydr. Res.* **180** 207-221 (1988)
- [234] Palma, R., Zuccato, P., Himmel, M.E., Liang, G. & Brady, J.W. Molecular mechanics studies of cellulases. In: *Glycosyl hydrolases in biomass conversions*. Himmel, M.E., Ed. American Chemical Society, Washington DC, U.S.A. pp 112-130 (2000)
- [235] Kuttel, M., Brady, J.W. & Naidoo, K.J. *J. Comput. Chem.* **23** 1236-1243 (2002)

- [236] Reiling, S., Schlenkrich, M. & Brickmann, J. *J. Comput. Chem.* **17** 450-568 (1995)
- [237] Eklund, R. & Widmalm, G. *Carbohydr. Res.* **338** 393-398 (2003)
- [238] Homans, S.W. *Biochemistry* **29** 9110-9118 (1990)
- [239] Glennon, T.M., Zheng, Y.-J., Le Grand, S.M., Schutzberg, B.A. & Merz Jr, K.M. *J. Comput. Chem.* **15** 1019-1040 (1994)
- [240] Glennon, T.M. & Merz Jr, K.M. *J. Mol. Struct.* **395** 157-171 (1997)
- [241] Gregurick, S.K., Liu, J.H.-Y., Brant, D.A. & gerber, R.G. *J. Phys. Chem. B* **103** 3476-3488 (1999)
- [242] Woods, R.J., Dwek, R.A., Edge, C.J. & Fraser-Reid, B. *J. Phys. Chem.* **99** 3832-3846 (1995)
- [243] Woods, R.J. & Chappelle, R. *J. Mol. Struct.* **527** 149-156 (2000)
- [244] Basma, M., Sundara, S., Calgan, D., Vernali, T. & Woods, R.J. *J. Comput. Chem.* **22** 1125-1137 (2001)
- [245] Kirschner, K.N. & Woods, R.J. *Proc. Natl. Acad. Sci. USA* **98** 10541-10545 (2001)
- [246] Momany, F.A. & Willett, J.L. *Carbohydr. Res.* **326** 194-209 (2000)
- [247] Momany, F.A. & Willett, J.L. *Carbohydr. Res.* **326** 210-226 (2000)
- [248] Simmerling, C., Fox, T. & Kollman, P.A. *J. Am. Chem. Soc.* **120** 5771-5782 (1998)
- [249] Damm, W., Frontera, A., Tirado-Rives, J. & Jorgensen, W. *J. Comput. Chem.* **18** 1955-1970 (1997)
- [250] Kony, D., Damm, W., Stoll, S. & van Gunsteren, W.F. *J. Comput. Chem.* **23** 1416-1429 (2002)
- [251] Kony, D., Damm, W., Stoll, S. & Hünenberger, P.H. *J. Phys. Chem. B* **108** 5815-5826 (2004)
- [252] van Gunsteren, W.F., Billeter, S.R., Eising, A.A., Hünenberger, P.H., Krüger, P., Mark, A.E., Scott, W.R.P. & Tironi, I.G. Biomolecular simulation: The GROMOS96 manual and user guide. Verlag der Fachvereine, Zürich, Switzerland. (1996)
- [253] Scott, W.R.P., Hünenberger, P.H., Tironi, I.G., Mark, A.E., Billeter, S.R., Fennen, J., Torda, A.E., Huber, T., Krüger, P. & van Gunsteren, W.F. *J. Phys. Chem. A* **103** 3596-3607 (1999)
- [254] Kouwijzer, M.L.C.E., van Eijck, B.P., Kooijman, H. & Kroon, J. *Acta Crystallogr. B* **51** 209-220 (1995)
- [255] Klewinghaus, P., van Eijck, B.P., Kouwijzer, M.L.C.E. & Kroon, J. *J. Mol. Struct.* **395** 289-295 (1997)

- [256] Ott, K.-H. & Meyer, B. *J. Comput. Chem.* **17** 1068-1084 (1996)
- [257] Spiesser, S.A.H., van Kuik, J.A., Kroon-Batenburg, L.M.J. & Kroon, J. *Carbohydr. Res.* **322** 264-273 (1999)
- [258] Galema, S.A., Howard, E., Engberts, J.B.F.N. & Grigera, J.R. *Carbohydr. Res.* **265** 215-225 (1994)
- [259] Howard, E.I. & Grigera, J.R. *Carbohydr. Res.* **282** 25-40 (1996)
- [260] Lins, R.D. & Hünenberger, P.H. *J. Comput. Chem.* **26** 1400-1412 (2005)
- [261] Molinero, V. & Goddard III, W.A. *J. Phys. Chem. B* **108** 1414-1427 (2004)
- [262] Molinero, V. & Goddard III, W.A. *ACS Symposium Series* **930** 271-284 (2006)
- [263] Lopez, C.A., Rzepiela, A., de Vries, A.H., Dijkhuizen, L., Hünenberger, P.H. & Marrink, S.J. *J. Chem. Theor. Comput.*, submitted (2009)
- [264] Molteni, C. & Parrinello, M. *J. Am. Chem. Soc.* **120** 2168-2171 (1998)
- [265] Ionescu, A.R., Bérces, A., Zgierski, M.Z., Whitfield, D.M. & Nukada, T. *J. Phys. Chem. A* **109** 8096-8105 (2005)
- [266] Pérez, S., Imberty, A., Engelsen, S.B., Gruza, J., Mazeau, K., Jimenez-Barbero, J., Poveda, A., Espinosa, J.-F., van Eyck, B.P., Johnson, G., French, A.D., Kouwijzer, M.L.C.E., Grootenuis, P.D.J., Bernardi, A., Raimondi, L., Senderowitz, H., Durier, V., Vergoten, G. & Rasmussen, K. *Carbohydr. Chem.* **314** 141-155 (1998)
- [267] Hemmingsen, L., Madsen, D.E., Esbensen, A.L., Olsen, L. & Engelsen, S.B. *Carbohydr. Res.* **339** 937-948 (2004)
- [268] Kouwijzer, M.L.C.E., van Eijck, B.P. & Kroon, J. *J. Comput. Chem.* **14** 1281-1289 (1993)
- [269] van Eijck, B.P. & Kroon, E.J. *J. Comput. Chem.* **20** 799-812 (1999)
- [270] Behler, J., Price, D.W. & Drew, M.G.B. *Phys. Chem. Chem. Phys.* **3** 588-601 (2001)
- [271] Corzana, F., Motawia, M.S., Hervé du Penhoat, C., Perez, S., Tschampel, S.M., Woods, R.J. & Engelsen, S.B. *J. Comput. Chem.* **25** 573-586 (2004)
- [272] Berendsen, H.J.C. *Simulating the physical world*. Cambridge University Press, Cambridge, U.K. (2007)
- [273] Schuler, L.D. & van Gunsteren, W.F. *Mol. Simul.* **25** 301-319 (2000)
- [274] van Gunsteren, W.F., Huber, T. & Torda, A.E. *AIP Conf. Proc.* **330** 253-268 (1995)
- [275] Christen, M. & van Gunsteren, W.F. *J. Comput. Chem.* **29** 157-166 (2008)
- [276] Hockney, R.W. *Methods Comput. Phys.* **9** 136-211 (1970)

- [277] Allen, M.P. & Tildesley, D.J. *Computer simulation of liquids*. Oxford University Press, New York, U.S.A. (1987)
- [278] Bekker, H. *J. Comput. Chem.* **18** 1930-1942 (1997)
- [279] Hünenberger, P.H. *J. Chem. Phys.* **116** 6880-6897 (2002)
- [280] Hünenberger, P.H. *Adv. Polym. Sci.* **173** 105-149 (2005)
- [281] Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., Di Nola, A. & Haak, J.R. *J. Chem. Phys.* **81** 3684-3690 (1984)
- [282] Ryckaert, J.-P., Ciccotti, G. & Berendsen, H.J.C. *J. Comput. Phys.* **23** 327-341 (1977)
- [283] van Gunsteren, W.F., Brunne, R.M., Gros, P., van Schaik, R.C., Schiffer, C.A. & Torda, A.E. Accounting for molecular mobility in structure determination based on nuclear magnetic resonance spectroscopic and X-ray diffraction data. In: *Methods in enzymology: nuclear magnetic resonance. Volume 239*. James, T.L. & Oppenheimer, N.J., Eds.; Academic Press, New York, U.S.A. pp 619-654 (1994)
- [284] van Gunsteren, W.F., Nanzer, A.P. & Torda, A.E. Molecular simulation methods for generating or trajectories consistent with experimental data. In: *Monte Carlo and molecular dynamics of condensed matter systems, Proceedings of the Euroconference, 3-28 July 1995, Como, Italy. Volume 49*. Binder, K. & Ciccotti, G., Eds.; SIF, Bologna, Italy. pp 777-788 (1996)

Chapter 2

Conformational properties of glucose-based disaccharides investigated using molecular dynamics simulations with local elevation umbrella sampling

2.1 Summary

Explicit-solvent molecular dynamics (MD) simulations of the 11 glucose-based disaccharides trehalose, isotrehalose, neotrehalose, kojibiose, sophorose, nigerose, laminarabiose, maltose, cellobiose, isomaltose and gentiobiose in water at 1 bar and 300K are reported. The simulations were carried out using the GROMOS 45A4 force-field and the sampling along the glycosidic dihedral angles ϕ and ψ was artificially enhanced using the local elevation umbrella sampling (LEUS) method. The trajectories are analyzed in terms of free-energy maps in the space of the glycosidic dihedral angles, stable and metastable conformational states (relative free-energies and estimated transition timescales), intramolecular hydrogen-bonding (H-bonding), solute configurational entropies and agreement with experimental data. All disaccharides considered are found to be characterized either by a single stable (overwhelmingly populated) state ($(1 \rightarrow n)$ -linked disaccharides with $n = 1, 2, 3$ or 4) or by two stable (comparably populated) states with a low interconversion barrier ($(1 \rightarrow 6)$ -linked disaccharides). Metastable states are also identified with relative free energies in the range 8-22 kJ mol⁻¹. The 11 compounds can be classified into four families (which correlate with approximate topological symmetries): (i) the $\alpha(1 \rightarrow 1)\alpha$ -linked disaccharide trehalose (axial-axial linkage) presents no metastable state, the lowest configurational entropy and no intramolecular H-bonds; (ii) the four $\alpha(1 \rightarrow n)$ -linked disaccharides ($n = 1, 2, 3$ or 4 ; axial-equatorial linkage) present one metastable state, an intermediate configurational entropy and two alternative intramolecular H-bonds; (iii) the four $\beta(1 \rightarrow n)$ -linked disaccharides ($n = 1, 2, 3$ or 4 ; equatorial-equatorial linkage) present two metastable states, an intermediate configurational entropy and one intramolecular H-bond; (iv) the two $(1 \rightarrow 6)$ -linked disaccharides (additional glycosidic dihedral angle) present no (isomaltose) or a pair of (gentiobiose) metastable states, the highest configurational entropy and no intramolecular H-bonds. The results suggest that the observed conformational preferences are dictated by four main driving forces (conformational preferences

of the residue rings, *exo*-anomeric effect, steric constraints, possible presence of a third glycosidic dihedral angle), leaving a secondary role to intramolecular H-bonding and specific solvation effects. In spite of the weak conformational driving force attributed to solvent-exposed H-bonds in water (highly polar protic solvent), intramolecular H-bonds may still have a significant influence on the physico-chemical properties of the disaccharide by decreasing its hydrophilicity. Along with previous work, the present results also complete the suggestion of a spectrum of approximate transition timescales for carbohydrates up to the disaccharide level, namely : ~ 30 ps (hydroxyl groups), ~ 1 ns (free lactol group, free hydroxymethyl groups, glycosidic dihedral angle $\tilde{\omega}$ in (1 \rightarrow 6)-linked disaccharides), ~ 10 ns to 2μ s (ring conformation, glycosidic dihedral angles ϕ and ψ). The calculated average values of the glycosidic torsional angles agree well with available experimental data, providing validation for the force-field and simulation methodology employed. Finally, the possible implications of the above findings in terms of the biological functions of the considered linkage types in carbohydrates, as well as carbohydrates in general, are discussed.

2.2 Introduction

Carbohydrates play a key role in numerous aspects of biology, biochemistry and human technology¹⁻⁷ including, for example : structure preservation and cell protection in various organisms⁸⁻¹⁶ (*e.g.* cellulose and polyuronates in plants, capsular polysaccharides in bacteria), energy storage and interconversion in all living cells^{11,12,17,18} (*e.g.* glucose, sucrose, glycogen and amylose), intra- and intercellular recognition processes^{3-6,19-27} (*e.g.* glycoproteins and glycolipids involved in cell signaling and immunological response), cell preservation under extreme environmental conditions²⁸⁻³¹ (*e.g.* trehalose), source of building blocks for cofactors or biomolecules^{11,12} (*e.g.* nucleotides and nucleic acids) and technological use in the design of new (biocompatible and biodegradable) materials³²⁻³⁸ (*e.g.* for the food, paper, textile, pharmaceutical or cosmetic industries). Some theories also attribute to these compounds a primordial role in the origins of life on earth^{39,40}. The diverse and complex functions of carbohydrates in living cells may be related to the high density of functional (predominantly hydroxyl) groups present in these compounds, in a virtually infinite variety of possible stereoisomers^{1,2,41} (at the monomer level) and chain-formation patterns^{1,2,41,42} (length, sequence, linkage type and branching).

Disaccharides are the simplest carbohydrate molecules presenting all the rotational degrees of freedom characteristic of more complex oligo- and polysaccharides^{1,2,43-45}. For this reason, characterizing the conformational properties of these molecules represents a key step towards the understanding of the conformation, flexibility and dynamics of more complex carbohydrates. In the context of molecular simulations, disaccharides also represent important benchmark systems for validating simulation methodologies and force-field parameters to be subsequently used for larger systems⁴³⁻⁴⁸. Finally, disaccharides themselves play an essential biological and technological role in terms of, for example, polysaccharide biosynthesis and degradation (*e.g.* maltose, isomaltose and cellobiose), energy storage and interconversion (*e.g.* sucrose, lactose and glucose-based disaccharides, all of which can be found in honey⁴⁹), bioprotection phenomena^{28-31,50} (*e.g.* trehalose in anhydrobiosis) and industrial food processing^{51,52}.

In a previous study^{53,54} (Pereira *et al.*, 2006), we reported the results of explicit-solvent molecular dynamics (MD) simulations investigating the conformational properties of the 8 reducing disaccharides of D-glucopyranose in water on the 50 ns timescale using the GROMOS 45A4

force-field^{55–59} (including a recently reoptimized parameter set for hexopyranose-based carbohydrates^{41,53–55,60}). The absolute configurational entropies of these disaccharides were also estimated by means of a quasi-harmonic (QH) analysis^{61–65} (including corrections for mode anharmonicities and pairwise supralinear correlations⁶⁵). One of the main observations made in this investigation was that the probability distributions associated with the glycosidic dihedral angles ϕ and ψ were essentially unimodal for all considered disaccharides, and that well-to-well (*i.e.* about 120°) transitions around these angles occurred at most once or twice for ϕ (never for ψ) on the 50 ns timescale. As a consequence, these simulations mainly (or only) probed the conformational regions of the corresponding Ramachandran (ϕ, ψ) maps closest to the starting structures employed in the simulations, which were taken in that study to be the experimentally derived crystallographic (x-ray) structures. The average values of the glycosidic dihedral angles calculated from the simulations agreed well with available experimental data from x-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy, suggesting that the sampled regions indeed corresponded to the most stable conformational regions for the considered disaccharides. However, in the absence of a more extensive sampling of the Ramachandran maps, the following restrictions apply to the conclusions of the article by Pereira *et al.*^{53,54}: (i) the relative populations of alternative conformers (which, even if less stable at the disaccharide level, may still be relevant in the context of oligo- and polysaccharides, either at equilibrium or along conformational transitions) could not be evaluated; (ii) the timescales associated with torsional transitions around the glycosidic dihedral angles or, equivalently, with state transitions between the alternative (meta)stable regions of the Ramachandran maps (which are important quantities for characterizing the conformational dynamics of more complex oligo- and polysaccharides) could not be evaluated; (iii) the solute configurational entropies estimated based on these simulations (QH analysis plus correction terms) only characterized the regions of the Ramachandran maps neighboring the starting conformations (*i.e.* these estimates were not totally converged); (iv) the force-field validation in the context of disaccharides was not entirely complete (because these simulations could not show in a strict sense that the conformational regions surrounding the experimental structures are indeed the most stable regions within the employed force-field).

The goal of the present work is to revisit (and extend) the conclusions of the above study^{53,54}, using MD simulations of similar durations in water, but that are biased so as to afford an extensive (*i.e.* nearly exhaustive in terms of the sterically allowed regions) coverage of the Ramachandran maps for the considered disaccharides. These new simulations rely on the application of the local elevation umbrella sampling (LEUS) method⁶⁰. This approach combines the advantages of the local elevation⁶⁶ (LE) conformational searching method and of the umbrella sampling^{67,68} (US) conformational sampling method. More specifically, the LEUS scheme relies on two steps: (i) a LE build-up (searching) phase, that is used to construct an optimized biasing potential within a subspace of conformationally relevant degrees of freedom; (ii) an US sampling phase, that is used to generate a biased ensemble with extensive coverage of the selected conformational subspace. The expression “biasing potential” refers here to an unphysical potential energy term that is added to the physical potential energy function (force-field) during the MD simulation. The LE build-up phase consists of a MD simulation involving the progressive construction of a memory-based (*i.e.* time-dependent) biasing potential that penalizes the resampling of previously visited regions within the considered conformational subspace^{60,66}. This potential is defined as a sum of local (short-ranged) repulsive functions centered at regularly spaced grid points within this subspace, the weights of which are made proportional to the number of visits of the corresponding grid

cells. A successful build-up phase will produce a biasing potential that is approximately equal to the negative of the free-energy hypersurface within the considered conformational subspace. The US sampling phase consists of a MD simulation involving a “frozen” (*i.e.* time-independent) biasing potential, which is set equal to the biasing potential reached at the end of the LE build-up phase. Because this biasing potential is close to the negative of the free-energy hypersurface, a long enough sampling phase will result in a nearly homogeneous coverage of the considered conformational subspace. In addition, because the biasing potential in this second phase is time-independent, thermodynamical information (*e.g.* relative conformer populations, free energies, enthalpies and entropies; free-energy barriers for conformer interconversions; structural properties) relevant for the physical (unbiased) ensemble can be recovered from the simulated data by means of a simple reweighting procedure^{60,67,68}. Although the dynamical information that would in principle be available from a (very long) free (unbiased) MD simulation is lost in the LEUS approach, it can still be evaluated in an approximate fashion by application of transition-state theory⁶⁹ (TST). Note that the LEUS approach bears some analogies with the adaptive umbrella sampling^{70–72}, adaptive biasing force^{73–75} and metadynamics^{76,77} methods, but within a very distinct two-steps implementation. As detailed in a previous article⁶⁰, this specific implementation leads to a number of advantages in terms of efficiency, accuracy, robustness and versatility.

The LEUS approach was previously applied to the calculation of the relative free energies and isomerization barriers of β -D-glucopyranose ring conformers in water (based on a three-dimensional conformational subspace defined by the out-of-plane dihedral angles characterizing the ring conformation according to Pickett and Strauss⁷⁸), and found to dramatically enhance the sampling compared to free (unbiased) MD simulations⁶⁰. In the present work, this approach is applied to disaccharides (based on a two-dimensional conformational subspace defined by the glycosidic dihedral angles ϕ and ψ ; possibly also including the third glycosidic dihedral angle $\tilde{\omega}$ in the case of (1 \rightarrow 6)-linked disaccharides). This choice of subspace for the LEUS enhancement is justified by the previous observation^{53,54} that torsional transitions around all other exocyclic dihedral angles occur in the \sim 30 ps (rotation of the hydroxyl groups) to ns (rotation of the free lactol and hydroxymethyl groups, and around the glycosidic dihedral angle $\tilde{\omega}$ in (1 \rightarrow 6)-linked disaccharides) range. This suggests that these degrees of freedom are appropriately sampled on the 50 ns timescale, even in the absence of sampling enhancement. In contrast, conformational transitions of the glucopyranose rings (from the most stable 4C_1 chair to the inverted 1C_4 chair, or to boat or skew-boat conformations) are much slower, with timescales of the order of 15-1000 ns (within the present force field⁶⁰) or 50-1000 ns (ultrasonic-relaxation spectroscopy^{79–82} and NMR⁸³ measurements on glucopyranose; estimate for penta-acetylated idopyranose⁸⁴). However, the corresponding reverse transitions are fast, with timescales in the ps to ns range (within the present force-field⁶⁰). Thus, the sampling of alternative ring conformations in (unbiased as well as LEUS) MD simulations is unlikely (with the possible exception of a small 0.1-0.2% population of 3,0B boat conformation within the present force-field^{41,60}), and in any case expected to have a negligible influence on the results (considering the very small equilibrium populations of these conformers and the rapidity of their back conversion to 4C_1).

The present simulations are also extended (compared to the previous work of Pereira *et al.*^{53,54}) to include the 3 non-reducing disaccharides of D-glucopyranose (trehaloses), in addition to the 8 reducing ones, for a total of 11 possible disaccharides in water (Figure 2.1). More specifically, the present simulations consider the non-reducing disaccharides trehalose ($\mathbf{T}_{\alpha\alpha}$), isotrehalose ($\mathbf{T}_{\alpha\beta}$) and neotrehalose ($\mathbf{T}_{\beta\beta}$), as well as the reducing disaccharides kojibiose (\mathbf{K}), sophorose

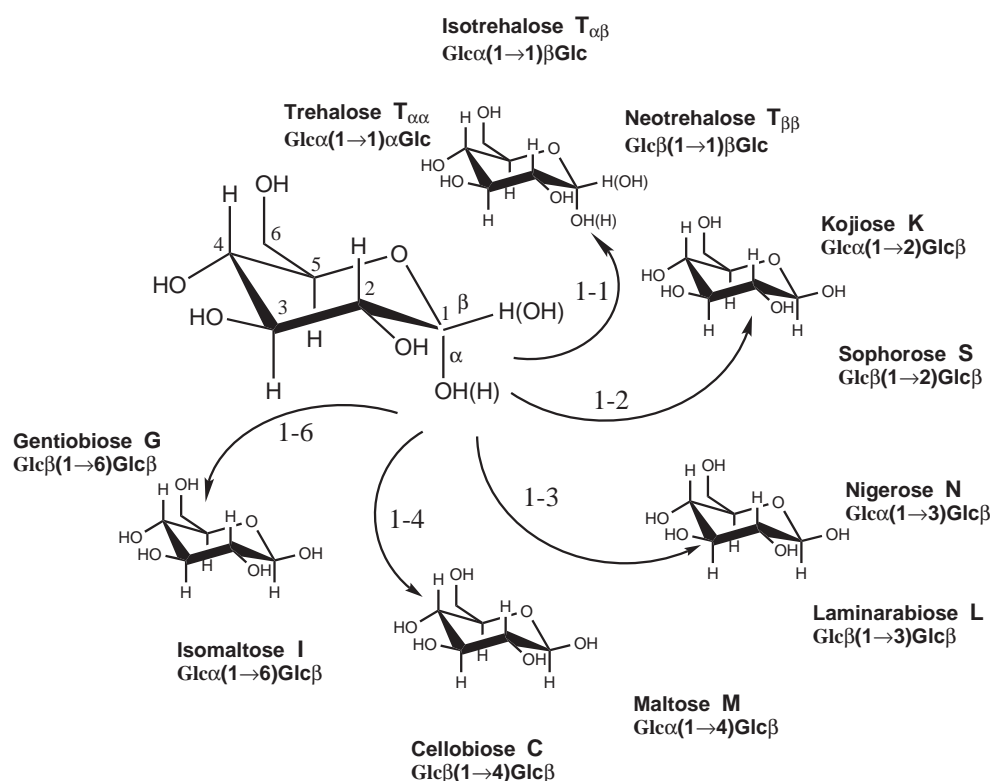


Figure 2.1: Schematic representation of the 11 disaccharides considered in the present study, i.e. all possible D-glucopyranose (Glc) dimers. Arrows indicate the (1 \rightarrow n)-linkage of a first (non-reducing) Glc residue (large drawing; 4C_1 chair conformation; atom numbering indicated) with a second non-reducing ($n = 1$; for the 3 non-reducing trehaloses) or reducing ($n = 2, 3, 4$ or 6 ; for the 8 reducing disaccharides) Glc residue. The first Glc residue can adopt an α (1-OH group axial; $\alpha(1 \rightarrow n)$ -linkage) or a β (1-OH group equatorial; $\beta(1 \rightarrow n)$ -linkage) anomery within the linkage. For the trehaloses, the second residue can also adopt an α or a β anomery within the linkage, resulting in three possible isomers. For the reducing disaccharides, the 8 isomers are only simulated with a β anomeric configuration at the reducing residue. The disaccharides considered are thus : trehalose ($T_{\alpha\alpha}$; $Glc\alpha(1 \rightarrow 1)\alpha Glc$), isotrehalose ($T_{\alpha\beta}$; $Glc\alpha(1 \rightarrow 1)\beta Glc$), neotrehalose ($T_{\beta\beta}$; $Glc\beta(1 \rightarrow 1)\beta Glc$), β -kojibiose (**K**; $Glc\alpha(1 \rightarrow 2)Glc\beta$), β -sophorose (**S**; $Glc\beta(1 \rightarrow 2)Glc\beta$), β -nigerose (**N**; $Glc\alpha(1 \rightarrow 3)Glc\beta$), β -laminarabiose (**L**; $Glc\beta(1 \rightarrow 3)Glc\beta$), β -maltose (**M**; $Glc\alpha(1 \rightarrow 4)Glc\beta$), β -cellobiose (**C**; $Glc\beta(1 \rightarrow 4)Glc\beta$), β -isomaltose (**I**; $Glc\alpha(1 \rightarrow 6)Glc\beta$) and β -gentiobiose (**G**; $Glc\beta(1 \rightarrow 6)Glc\beta$).

(**S**), nigerose (**N**), laminarabiose (**L**), maltose (**M**), cellobiose (**C**), isomaltose (**I**) and gentiobiose (**G**). For simplicity, however, the 8 reducing disaccharides are only simulated in their β -anomeric form at the reducing end (expected to account for about 62% of the total anomeric population⁸⁵). All of these disaccharides have been subject to previous investigations, both experimentally by x-ray crystallography^{2, 86–92}, NMR spectroscopy^{93–104} or circular dichroism¹⁰⁵ (CD), as well as theoretically by quantum mechanics^{106–111} (QM), molecular mechanics^{112–120} (MM), or MD simulation^{53, 54, 97, 99, 102, 104, 121–137}.

After optimization of the LEUS protocol and evaluation of the performance of this approach (compared to plain MD simulations), the 11 new simulations with LEUS sampling enhancement

are analyzed in terms of : (i) calculated free-energy maps $G(\phi, \psi)$, as well as $G(\phi, \psi, \tilde{\omega})$ for the (1 \rightarrow 6)-linked disaccharides, in the subspace of the glycosidic dihedral angles; (ii) relative free energies of the identified conformational states (free-energy basins); (iii) interconversion barriers between these states (and transition timescales estimated using the TST⁶⁹ approximation); (iv) occurrences of intramolecular hydrogen bonds (H-bonds) in these states; (v) estimates for the solute configurational entropies (QH analysis plus correction terms⁶⁵); (vi) agreement with available experimental and theoretical data. These results complement those previously reported by Lins & Hünenberger⁵⁵, Pereira *et al.*^{53,54}, Kräutler *et al.*⁴¹ and Hansen & Hünenberger⁶⁰ in providing a detailed characterization of the (thermodynamical and dynamical) properties of mono- and disaccharides within the GROMOS 45A4 force-field⁵⁵⁻⁵⁹. Good agreement with available experimental and theoretical data also represents a validation for the latter force-field. In turn, these simulation studies provide insight into a number of properties of mono- and disaccharides that are not directly (or not unambiguously) accessible by experimental measurements.

2.3 Computational details

2.3.1 Molecular dynamics simulations

All MD simulations were carried out using the GROMOS96 program^{138,139}, together with the GROMOS 45A4 force-field⁵⁵⁻⁵⁹ (including a recently reoptimized parameter set for hexopyranose-based carbohydrates^{41,53-55,60}) and the SPC water model¹⁴⁰. Note that the 45A4 force-field is equivalent to the more recent 53A6 force-field^{141,142} in the context of pure carbohydrate systems. Periodic boundary conditions were employed based on truncated-octahedral computational boxes (derived from cubes of 4 nm initial edge lengths) containing one disaccharide molecule surrounded by about 1000 water molecules ($\mathbf{T}_{\alpha\alpha}$: 1030; $\mathbf{T}_{\alpha\beta}$: 1055; $\mathbf{T}_{\beta\beta}$: 1055; \mathbf{K} : 1012; \mathbf{S} : 988; \mathbf{N} : 972; \mathbf{L} : 1109; \mathbf{M} : 1002; \mathbf{C} : 1048; \mathbf{I} : 924; \mathbf{G} : 1070). Newton's equations of motion were integrated using the leap-frog scheme¹⁴³ with a timestep of 2 fs. The SHAKE algorithm¹⁴⁴ was applied to constrain all bond lengths with a relative geometric tolerance of 10^{-4} . The center of mass motion was removed every 100 ps. The simulations were carried out in the isothermal-isobaric (NPT) ensemble at a temperature of 300 K and a pressure of 1 bar. The temperature was maintained by weakly coupling the solute and solvent degrees of freedom separately to a heat bath¹⁴⁵, with a relaxation time of 0.1 ps. The pressure was maintained by weakly coupling the particle coordinates and box dimensions (isotropic coordinate scaling) to a pressure bath¹⁴⁵, with a relaxation time of 0.5 ps and an isothermal compressibility of $0.4575 \cdot 10^{-3} \text{ kJ}^{-1} \text{ mol nm}^3$ as appropriate for water¹³⁸. Nonbonded interactions were computed using a twin-range scheme^{138,146}, with short- and long-range cutoff distances set to 0.8 and 1.4 nm, respectively, and a frequency of 5 timesteps for the update of the short-range pairlist and intermediate-range interactions. A reaction-field correction^{147,148} was applied to account for the mean effect of omitted electrostatic interactions beyond the long-range cutoff distance, using a relative dielectric permittivity of 61 as appropriate for the SPC water model¹⁴⁹.

Whenever applied, the LEUS method relied on a biasing potential \mathcal{U}_b defined as a sum of local

(grid-based) Gaussian functions, *i.e.*⁶⁰

$$\mathcal{U}_b(\mathbf{Q}; \{n_k\}) \doteq c \sum_{k=1, n_k \neq 0}^{N_g^{N_l}} n_k \prod_{i=1}^{N_l} g(\text{MI}[Q_i - Q_{k,i}^*]; \sigma) \quad (2.1)$$

with

$$g(\Delta Q; \sigma) \doteq (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{(\Delta Q)^2}{2\sigma^2}\right]. \quad (2.2)$$

Here, \mathbf{Q} is the point representative of a system configuration in the N_l -dimensional subspace where the LEUS enhancement is to be applied, N_g is the number of grid points used to discretize this subspace along each of its N_l dimensions (chosen here to be identical along all dimensions, for simplicity), \mathbf{Q}_k^* the coordinate of a grid point k in this subspace, n_k the number of visits to the corresponding grid cell, MI the minimum-image function and g a normalized one-dimensional Gaussian of specified width σ . In the LE build-up phase, the quantity n_k encompasses all visits of a grid cell counted up to the current time point of the simulation (*i.e.* it is time-dependent). In the US sampling phase, the quantity n_k only encompasses visits summed up over the preceding build-up phase (*i.e.* it is now time-independent). The parameters σ and c in Eqs. 2.1 represent the width and force constant increment per visit associated with the local Gaussian functions (chosen here to have common values at all grid points, for simplicity). Eqs. 2.1 and 2.2 were implemented in a slightly modified version⁶⁰ of the LE routine⁶⁶ in the GROMOS96 program¹³⁸. The alterations involve : (i) the introduction of the minimum-image convention for the action of the local Gaussian functions on periodic degrees of freedom (here, dihedral angles), which was not included in the original implementation^{66,138}; (ii) the removal of the truncation to a single grid cell for the action of the local Gaussian functions, which was applied in the original implementation^{66,138}. As a result, the range of action of the local Gaussian functions is only restricted here to a single minimum-image period for periodic degrees of freedom.

The free parameters of the LEUS protocol are : (i) the dimensionality N_l of the conformational subspace in which the LEUS sampling enhancement is applied; (ii) the number N_g of grid points used to discretize this subspace along each of its N_l dimensions; (iii) the width σ of the local Gaussian functions used to represent the biasing potential; (iv) the force constant increment c per visit associated with these functions; (v) the duration t_{LE} of the LE build-up phase; (vi) the duration t_{US} of the US sampling phase. Unless otherwise specified, the LEUS subspace is the two-dimensional subspace ($N_l = 2$) defined by the glycosidic dihedral angles ϕ and ψ (*i.e.* this subspace is the one represented in the Ramachandran map). For the (1 \rightarrow 6)-linked disaccharides (**I** and **G**), additional LEUS simulations were also performed in a three-dimensional LEUS subspace ($N_l = 3$) defined by the glycosidic dihedral angles ϕ , ψ and $\tilde{\omega}$. The definitions of the dihedral angles ϕ , ψ and $\tilde{\omega}$ are provided in Section 2.3.2.

A number of possible combinations of the free parameters N_g , σ , c , t_{LE} and t_{US} were tested and compared in the specific context of cellobiose (**C**). After a few attempts (data not shown), the choices $N_g = 32$ (resulting in a grid spacing of 11.25° along each of the N_l dimensions) and $\sigma = 360^\circ / (2N_g) = 5.625^\circ$ (Gaussian width set to half this grid spacing) were adopted as providing an appropriate basis set for the representation of the biasing potential, while keeping the memory and computational requirements relatively low. However, further tests were required to determine reasonable values for the parameters c , t_{LE} and t_{US} . Illustrative results for **C** are reported in

Section 2.4.2. The choices $c = 10^{-4}$ kJ mol $^{-1}$ deg 2 ($N_l=2$) or $c = 10^{-4}$ kJ mol $^{-1}$ deg 3 ($N_l=3$; **I** and **G** only), $t_{LE} = 50$ ns and $t_{US} = 50$ ns were finally adopted for all production simulations.

The simulations analyzed include one plain (unbiased) MD simulation and one (two for **G** and **I**, either excluding or including $\tilde{\omega}$ in the LEUS subspace) LEUS simulation(s) for each of the 11 disaccharides (Figure 2.1). The 11 plain MD simulations were carried out for a duration $t_{MD} = 50$ ns (after 0.1 ns equilibration), and correspond to the 8 simulations previously reported by Pereira *et al.*^{53,54} (reducing disaccharides) complemented by 3 additional simulations for the trehaloses (**T** $_{\alpha\alpha}$, **T** $_{\alpha\beta}$ and **T** $_{\beta\beta}$). The latter three simulations were performed based on the same protocol⁵³ and initiated from structures generated using the program InsightII¹⁵⁰ based on the glycosidic dihedral angle values reported for the corresponding x-ray structures⁸⁹⁻⁹¹. The 11+2 LEUS simulations consisted of a LE build-up phase of duration $t_{LE} = 50$ ns followed by an US sampling phase of duration $t_{US} = 50$ ns. They relied on the parameters (see above) : $N_g = 32$, $\sigma = 5.625^\circ$ and $c = 10^{-4}$ kJ mol $^{-1}$ deg 2 ($N_l=2$) or $c = 10^{-4}$ kJ mol $^{-1}$ deg 3 ($N_l=3$; **I** and **G** only). Atomic coordinates were saved to file every 0.5 ps (plain MD simulations) or 0.05 ps (sampling phase of the LEUS simulations) for subsequent analysis (Section 2.3.2).

2.3.2 Analysis

The analysis of simulations was performed in terms of : (i) pyranose ring conformations; (ii) free-energy maps $G(\phi, \psi)$, as well as $G(\phi, \psi, \tilde{\omega})$ in the case of the (1 \rightarrow 6)-linked disaccharides; (iii) free-energy differences (minimum-to-minimum $\Delta G_{X \rightarrow Y}$ or state-to-state $\Delta \tilde{G}_{X \rightarrow Y}$), as well as free-energy barriers (minimum-to-saddle point $\Delta G_{X \rightarrow Y}^\ddagger$), between conformational states identified on these maps; (iv) estimated rates $k_{X \rightarrow Y}$ and timescales (half-times) $\tau_{X \rightarrow Y}$ associated with the transitions between these states; (v) numbers of transitions $n_{X \rightarrow Y}$ actually observed in the simulations between the different states; (vi) occurrence of intramolecular H-bonds in the different states; (vii) estimated solute configurational entropies S_{sol} .

In the present work, the following conventions were adopted. The atom numbering around the D-glucopyranose ring, the anomery, the linkage types and the disaccharide codes are as described in Figure 2.1. For the reducing disaccharides, the atom labels of the reducing residue are marked with a prime. For isotrehalose (**T** $_{\alpha\beta}$), the atom labels of the residue with a β anomery are marked with a prime. For trehalose (**T** $_{\alpha\alpha}$) and neotrehalose (**T** $_{\beta\beta}$), the two residues are topologically equivalent and the choice of the primed residue is arbitrary. The definitions of the glycosidic dihedral angles follow the IUPAC convention^{151,152}, except for $\tilde{\omega}$ (see below). The glycosidic dihedral angles ϕ and ψ around a (1 \rightarrow n)-linkage ($n = 2, 3, 4$ or 6; reducing disaccharide) are defined by the atom sequences O $_5$ -C $_1$ -O $_1$ -C' $_n$ and C $_1$ -O $_1$ -C' $_n$ -C' $_{n-1}$, respectively. The glycosidic dihedral angles ϕ and ψ around a (1 \rightarrow 1)-linkage (non-reducing trehaloses) are defined by the atom sequences O $_5$ -C $_1$ -O $_1$ -C' $_1$ and C $_1$ -O $_1$ -C' $_1$ -O' $_5$, respectively. Note that the definition of ψ differs from the case of the non-reducing disaccharides and that the distinction between ϕ and ψ depends on an arbitrary choice of the primed residue for **T** $_{\alpha\alpha}$ and **T** $_{\beta\beta}$ (see above; for this reason, the two dihedral angles are also sometimes noted as ϕ and ϕ' instead of ϕ and ψ). The third glycosidic dihedral angle $\tilde{\omega}$ involved in a (1 \rightarrow 6)-linkage (**I** and **G**) is defined by the atom sequence O $_1$ -C' $_6$ -C' $_5$ -O' $_5$. The “tilde” distinguishes this angle from the dihedral angle ω (IUPAC definition^{151,152}), defined by the atom sequence O $_1$ -C' $_6$ -C' $_5$ -C' $_4$. Note that $\tilde{\omega}$ (or ω) should formally rather be written^{53,54} $\tilde{\omega}'$ (or ω') since it pertains to the reducing residue (however, because the orientation of the free hydroxymethyl group in the non-reducing residue was not monitored in the present study, the

symbol $\tilde{\omega}$ was retained here for simplicity). The signs of all dihedral angles were also assigned according to the IUPAC convention^{151,152} and, unless otherwise specified, corresponding values were selected within the period $[0^\circ;360^\circ[$.

The pyranose ring conformation of the two residues was assigned based on three out-of-plane dihedral angles α_1 , α_2 and α_3 defined according to Pickett and Strauss⁷⁸ by the values of the dihedral angles $C_4-O_5-C_2-C_1$, $O_5-C_2-C_4-C_3$ and $C_2-C_4-O_5-C_5$, respectively, decreased by 180° . Configurations with α_1 , α_2 , $\alpha_3 > 0$ were attributed to the 4C_1 chair conformation. Since this conformation is the only significantly populated one in the present simulations (Sections 2.4.1 and 2.4.3), minor occurrences of alternative ring conformers were not investigated in more details.

Free-energy maps $G(\phi, \psi)$ were evaluated from the plain MD and from the LEUS simulations as detailed elsewhere⁶⁰ (see Eqs. 7 and 8 therein). Because the free energy is only defined within an unknown constant, these maps were anchored to $G = 0 \text{ kJ mol}^{-1}$ at the location of their global minimum. In addition, the value of G at grid points that were never visited during a (plain MD or LEUS) simulation (which is formally infinite) was arbitrarily set to the maximal value G_{max} of G over all grid points that were visited at least once. All free energy maps were calculated using a grid spacing of 6° .

For the ease of analysis, the free-energy maps were divided into four distinct conformational regions (*A*, *B*, *C* and *D*) using cutoff values ϕ_c and ψ_c for the $(1 \rightarrow n)$ -linked disaccharides ($n = 1, 2, 3$ or 4), or ϕ_c and $\tilde{\omega}_c$ for the $(1 \rightarrow 6)$ -linked disaccharides. In the former case, the four regions are defined by (ϕ, ψ) values in the ranges: *A* $[(0^\circ, 0^\circ);(\phi_c, \psi_c)]$, *B* $[(0^\circ, \psi_c);(\phi_c, 360^\circ)]$, *C* $[(\phi_c, 0^\circ);(360^\circ, \psi_c)]$ and *D* $[(\phi_c, \psi_c);(360^\circ, 360^\circ)]$. In the latter case, the four regions are defined by $(\phi, \tilde{\omega})$ values in the ranges: *A* $[(0^\circ, 0^\circ);(\phi_c, \tilde{\omega}_c)]$, *B* $[(0^\circ, \tilde{\omega}_c);(\phi_c, 360^\circ)]$, *C* $[(\phi_c, 0^\circ);(360^\circ, \tilde{\omega}_c)]$ and *D* $[(\phi_c, \tilde{\omega}_c);(360^\circ, 360^\circ)]$. The cutoff values selected for the different disaccharides are reported in Table 2.1. These values were chosen by visual inspection of the free-energy maps (LEUS simulations; Section 2.4.3) so as to approximately divide the maps into distinct free-energy basins and inaccessible (high free-energy) regions. Whenever a free-energy basin is present in a given region of the map, the corresponding conformational state is noted by the letter assigned to the defining region (*i.e.* the corresponding state is also labeled $X = A, B, C$ or D).

The free-energy difference between two of these states can be defined in either of two ways. The minimum-to-minimum free-energy difference $\Delta G_{X \rightarrow Y} \doteq G_{min,Y} - G_{min,X}$ between states *X* and *Y* corresponds to the difference between the free energies $G_{min,X}$ and $G_{min,Y}$ at the minimum points of states *X* and *Y*. The corresponding state-to-state free-energy difference $\Delta \tilde{G}_{X \rightarrow Y} \doteq -\beta^{-1} \ln(P_Y/P_X)$, where $\beta \doteq (k_B T)^{-1}$, k_B being Boltzmann's constant and T the absolute temperature (300 K), is defined based on the ratio of the overall populations P_X and P_Y of the conformational regions *X* and *Y* (number of visits of these two regions during the simulation; for the LEUS simulations, these values are reweighted so as to remove the effect of the biasing potential).

The free-energy barriers associated with transitions between two of these conformational states were evaluated as detailed elsewhere⁶⁰ (see Section "assignment schemes" therein). The procedure employed relies on defining, for each basin *X* on the free-energy map, a set $S_X(h_X)$ collecting grid points according to the following rules: (i) S_X contains the point corresponding to the free-energy minimum of basin *X*; (ii) all points of S_X are neighbor of at least one other point of S_X ; (iii) all points of S_X that are above $G_{min,X}$ in terms of free energy lie within a free-energy cutoff h_X from $G_{min,X}$. The (minimum-to-saddle point) free-energy barrier $\Delta G_{X \rightarrow Y}^\ddagger$ associated with a transition from *X* to *Y* is then estimated as the minimal value of h_X for which $S_X(h_X)$ contains the point corresponding to the free-energy minimum of basin *Y*. Based on this definition, the free-energy

barriers $\Delta G_{X \rightarrow Y}^\ddagger$ and $\Delta G_{Y \rightarrow X}^\ddagger$ associated with forward and backward transitions between two states are related to the corresponding minimum-to-minimum free-energy difference by

$$\Delta G_{X \rightarrow Y} = \Delta G_{X \rightarrow Y}^\ddagger - \Delta G_{Y \rightarrow X}^\ddagger . \quad (2.3)$$

The calculated barriers can in turn be used to estimate unimolecular reaction rates $k_{X \rightarrow Y}$ and associated half-life times $\tau_{X \rightarrow Y}$ according to the TST⁶⁹ approximation, as

$$k_{X \rightarrow Y} = \frac{\ln 2}{\tau_{X \rightarrow Y}} = \frac{\kappa_{X \rightarrow Y}}{\beta h} \exp \left[-\beta \Delta G_{X \rightarrow Y}^\ddagger \right] , \quad (2.4)$$

where h is Planck's constant and $\kappa_{X \rightarrow Y}$ is the corresponding transmission coefficient. In the absence of kinetic information on transition trajectories (which cannot be obtained from LEUS simulations since the dynamics of the system is unphysical), the latter coefficients were set to one as a first approximation.

For comparison purposes between plain MD and LEUS simulations (sampling phase), the real numbers of transitions $n_{X \rightarrow Y}$ from state X to state Y observed on the 50 ns timescale were also monitored. A transition from state X to state Y is assumed to have occurred whenever the system spent at least 5 ps in region X followed by at least 5 ps in region Y .

The calculation of thermodynamical and structural observables relevant for the real (unbiased) system based on the (biased) LEUS simulations relies on a reweighting procedure as detailed elsewhere⁶⁰ (see Eqs. 3-6 therein). The statistical efficiency of this reweighting may be estimated by the parameter

$$F_s \doteq N_f^{-1} \exp \left[-\sum_{i=1}^{N_f} p_i \ln p_i \right] , \quad (2.5)$$

where N_f is the number of stored trajectory frames (10^6 for the 50 ns sampling phases of the present simulations with coordinates stored every 0.05 ps) and p_i the statistical (unbiasing) weight associated with frame i , defined as

$$p_i \doteq \left[\sum_{i=1}^{N_f} \exp[\beta u_b(\mathbf{Q}_i)] \right]^{-1} \exp[\beta u_b(\mathbf{Q}_i)] . \quad (2.6)$$

The limiting case of an unbiased simulation corresponds $p_i = N_f^{-1}$ and one has $F_s = 1$ (maximum statistical efficiency). The limiting case of a biased simulation where a single configuration j entirely dominates the reweighted probability distribution corresponds to $p_i = 0$ for $i \neq j$ along with $p_j = 1$, leading to $F_s = N_f^{-1}$ (minimum statistical efficiency). The definition of F_s was inspired from considerations relative to the calculation of QH entropies based on reweighted trajectories (see below). The quantity $F_s N_f$ can be viewed as a "effective" number of frames of the biased trajectory contributing to the statistics in terms of the properties of the unbiased ensemble. Although one usually has $F_s \ll 1$ in a biased simulation, the sampling efficiency may still be greatly enhanced in practice when this "effective" number of configurations spans a much wider (*i.e.* more representative) portion of the configurational space accessible to the unbiased system (as is the case in the LEUS simulations). Note that a number of other measures for this statistical efficiency have been proposed previously¹⁵³⁻¹⁵⁷.

The occurrence of intramolecular H-bonds in the LEUS simulations (sampling phase) was also analyzed (separately for each of the conformational states), considering all hydroxyl groups

as potential H-bond donors, and all hydroxyl or ring oxygen atoms as potential acceptors. The presence of a H-bond was defined by a maximal hydrogen-oxygen distance of 0.25 nm and a minimal oxygen-hydrogen-oxygen angle of 135° . The calculation of the corresponding occurrences from the LEUS (biased) simulations involved the appropriate reweighting of the configurations so as to remove the effect of the biasing potential (see above).

Solute (absolute) configurational entropies were estimated from the LEUS simulations (sampling phase) based on a QH analysis^{61–65} by calculating the solute all-atom mass-weighted covariance matrix in Cartesian coordinates⁶², after least-square fit superposition⁶³ of the successive trajectory configurations onto a reference structure (so as to eliminate the overall translational and rotational motion of the disaccharide). The latter reference structure was selected as the configuration taken from the corresponding unbiased simulation that presented ϕ and ψ (as well as $\tilde{\omega}$ for **I** and **G**) values within 5° of the global minimum of the corresponding free-energy map and that led to the lowest configurational entropy estimate. The QH entropy estimate (S_{qm}^h) was evaluated as the entropy of a multidimensional quantum-mechanical harmonic oscillator with the same mass-weighted covariance matrix as that determined from the simulation. The six (nearly zero) eigenvalues corresponding to the suppressed rigid-body motion were left out of the analysis. The convergence of the estimated entropies with time was assessed by repeating the analysis for increasingly long time periods along the simulations (differing in length by 50 ps). Entropy corrections for anharmonicities in the QH modes (ΔS_{cl}^{ah}) and for pairwise supralinear correlations among the modes (ΔS_{cl}^{pc}) were evaluated at the classical level as detailed elsewhere⁶⁵ (see Eqs. 46–49 therein), resulting in the final estimate

$$S_{sol} = S_{qm}^h + \Delta S_{cl}^{ah} + \Delta S_{cl}^{pc} \quad (2.7)$$

for the solute configurational entropy. This approach was previously applied to the 8 non-reducing disaccharides of D-glucopyranose based on plain MD simulations on the 50 ns timescale^{53,54}, but the accuracy of the results was probably impaired by insufficient convergence. The application of this method to calculate solute configurational entropies from the LEUS (biased) simulations is straightforward, but requires the appropriate reweighting of the configurations so as to remove the effect of the biasing potential (see above). This reweighting was applied in the determination of the average structure and mass-weighted covariance matrix, as well as in the estimation of the corrections for mode anharmonicities and pairwise supralinear correlations⁶⁵. In the latter case, it is important to note that the optimized ratio κ° of the bin width to the effective Gaussian width along a given mode⁶⁵ (Eqs. C5 and C6 therein) must also be adjusted for the reduced statistical efficiency ($F_s < 1$ in Eqs. 2.5). It is easily shown that this is achieved by replacing N_f by $F_s N_f$ in the corresponding equations. The optimized ratio κ_1° for one-dimensional integrals was used in the evaluation of ΔS_{cl}^{ah} and the optimized ratio κ_2° for two-dimensional integrals was used in the evaluation of ΔS_{cl}^{pc} (for both the one- and two-dimensional integrals involved in this term, so as to maximize error cancellation).

2.4 Results

2.4.1 Unbiased MD simulations

As a preliminary verification, the ring conformations of the residues along the 50 ns plain (unbiased) MD simulations of the 11 considered disaccharides were monitored as a function of time (data not shown). It was found that the ring dihedral angles in both residues are consistent with a leading 4C_1 chair conformation, accounting for 96.5-99.9% of the sampled configurations for the different disaccharides.

The sampling of (a significant fraction of) alternative ring conformations in unbiased MD simulations is unlikely considering previous results for D-glucopyranose (with the possible exception of a small 0.1-0.2% population of 3_0B boat conformation within the present force-field^{41,60}). However, it was thought necessary to verify this statement also in the context of the disaccharides, since the presence of alternative ring conformations (if it happened) could significantly bias the results of subsequent analysis (in the absence of sampling enhancement on the ring conformations⁶⁰).

The free-energy maps $G(\phi, \psi)$ in the subspace of the glycosidic dihedral angles ϕ and ψ as obtained from these 11 plain (unbiased) MD simulations are displayed in Figure 2.2.

During the 50 ns simulations, well-to-well (*i.e.* about 120°) transitions around the two glycosidic dihedral angles are scarce, forward-backward transitions occurring only once or twice for ϕ in **S**, **C** and **G**, and six times for ψ in $\mathbf{T}_{\alpha\beta}$. For **I** and **G**, the free-energy profile along ψ presents a single broad well centered at $\psi \approx 180^\circ$ and the concept of transition makes little sense in this case. For most disaccharides, the simulation samples a single free-energy basin in the neighborhood of the starting structure, which was chosen here to be the experimental (x-ray) structure. Only in four simulations ($\mathbf{T}_{\alpha\beta}$, **S**, **C** and **G**; those for which ϕ or ψ transitions were observed) is an alternative basin visited. However, because the number of transitions between the two regions is extremely low (see above), the free energies of these alternative conformational states (relative to the main state) cannot be determined accurately.

Because alternative (metastable) conformational states are either not sampled or associated with high (apparent) relative free energies, the resulting one-dimensional probability distributions along the dihedral angles ϕ and ψ are essentially unimodal, presenting average values in excellent agreement with the available experimental (x-ray or NMR) data, as already stated by Pereira *et al.*^{53,54} for the reducing disaccharides (this observation remains valid for the trehaloses; data not shown). For the topologically symmetrical trehaloses ($\mathbf{T}_{\alpha\alpha}$ and $\mathbf{T}_{\beta\beta}$, where ϕ and ψ are interchangeable), the free-energy maps also approximately satisfy the expected mirror symmetry across the diagonal. However, the preservation of this symmetry property is largely fortuitous, and arises from the fact that the free-energy basins associated with the experimental structures are precisely on the diagonal for these compounds.

In spite of the apparent agreement with experimental data^{53,54}, it is important to stress that these 11 simulations only sample a very limited fraction of the conformational space accessible to the disaccharides. Thus, strictly speaking, they are insufficiently converged to : (i) rule out the existence of other (meta)stable conformational states; (ii) evaluate the relative populations and free-energies of the sampled alternative states (for $\mathbf{T}_{\alpha\beta}$, **S**, **C** and **G**) accurately; (iii) evaluate the timescales associated with transitions in the ϕ and ψ glycosidic dihedral angles; (iv) provide unambiguous validation data for the employed force-field. These goals could only be achieved

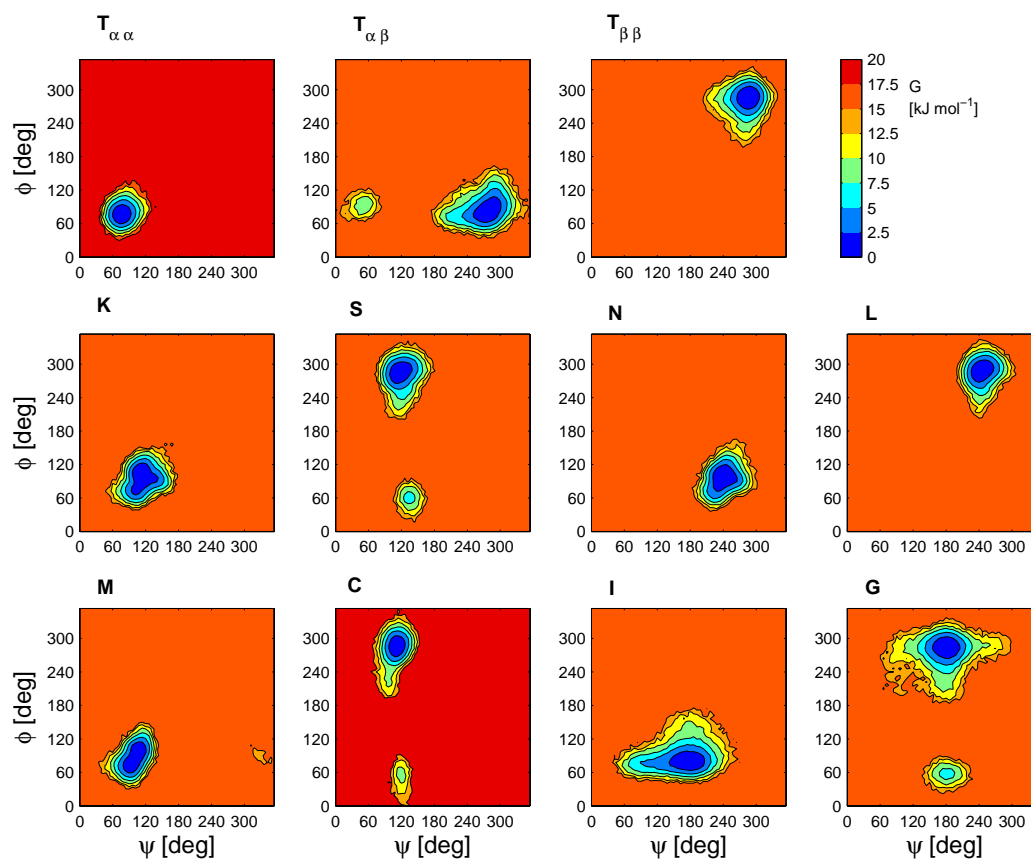


Figure 2.2: Free-energy maps $G(\phi, \psi)$ in the subspace of the glycosidic dihedral angles ϕ and ψ as obtained from 50 ns plain (unbiased) MD simulations of the 11 considered disaccharides in water at 300 K and 1 bar. The maps are anchored to $G = 0 \text{ kJ mol}^{-1}$ at the location of their global minimum, and the value of G at grid points that were never visited during the simulations is arbitrarily set to the maximal value G_{max} of G over all grid points that were visited at least once. The maps were calculated using a grid spacing of 6° . Contours are drawn at successive 2.5 kJ mol^{-1} intervals starting from the global minimum. The codes employed for the different disaccharides are provided in the legend of Figure 2.1.

by undertaking (significantly) longer simulations, or, as described in the following sections, by applying a sampling enhancement method such as the LEUS scheme⁶⁰.

2.4.2 LEUS protocol

Application of the LEUS approach requires the specification of five parameters (Section 2.3.1) : N_g , σ , c , t_{LE} and t_{US} . A number of possible combinations of these free parameters were tested and compared in the context of **G**. After a few tests (data not shown), the choices $N_g=32$ (resulting in a spacing of 11.25° between grid points along each of the N_l dimensions) and $\sigma = 360^\circ/(2N_g) = 5.625^\circ$ (Gaussian width set to half this spacing) were adopted as providing an appropriate basis set for the representation of the biasing potential, while keeping the memory and computational

requirements relatively low. Illustrative results for further tests aiming at the determination of reasonable values for the parameters c , t_{LE} and t_{US} in the specific context of cellobiose (C) are presented in Figure 2.3.

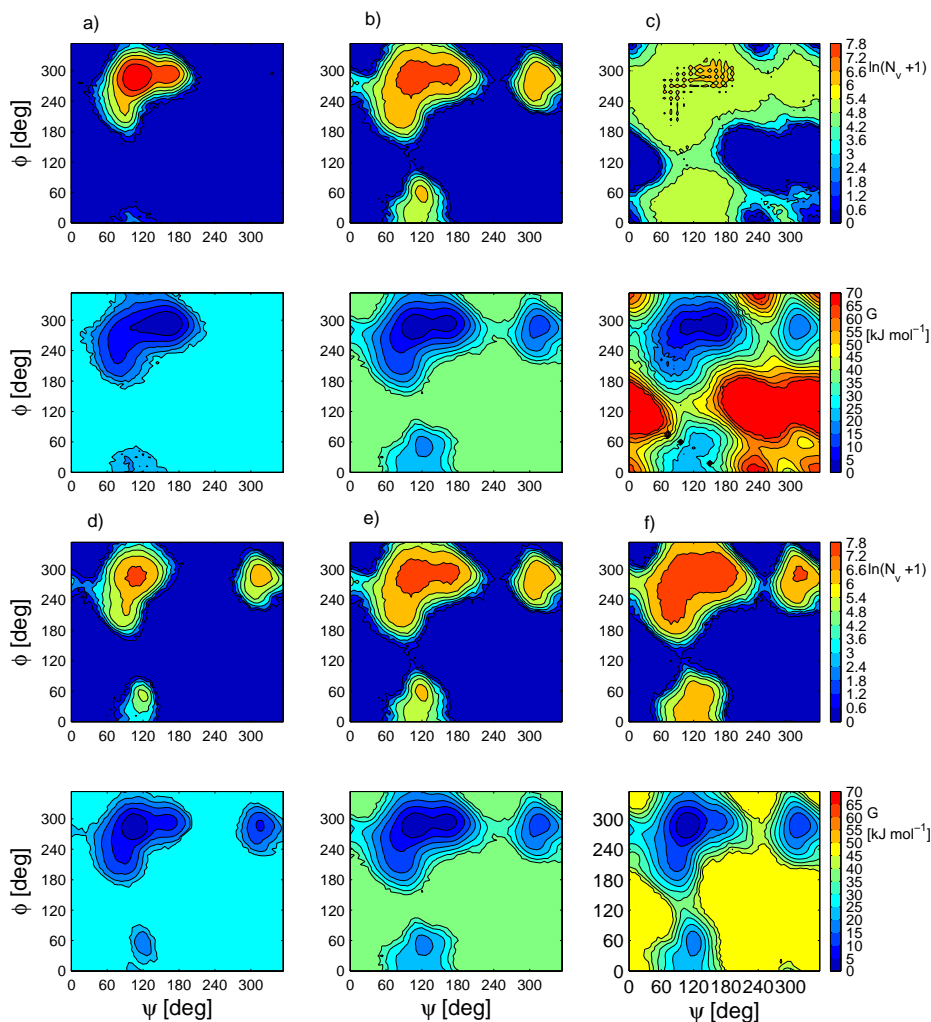


Figure 2.3: Number of visits N_v per grid point at the end of the LE build-up phase (top graph in each panel; displayed as $\ln(N_v + 1)$) and free-energy maps $G(\phi, \psi)$ calculated from the US sampling phase (bottom graph in each panel) in the subspace of the glycosidic dihedral angles ϕ and ψ , as obtained from LEUS simulations of the disaccharide cellobiose (C; Figure 2.1) in water at 300 K and 1 bar. The LEUS parameters are $N_l = 2$, $N_g = 32$, $\sigma = 360^\circ / (2N_g) = 5.625^\circ$, together with : $t_{LE} = t_{US} = 25$ ns (a-c) along with $c = 10^{-5}$ (a), 10^{-4} (b) or 10^{-3} (c) $\text{kJ mol}^{-1} \text{deg}^2$; $c = 10^{-4}$ $\text{kJ mol}^{-1} \text{deg}^2$ (d-f) along with $t_{LE} = t_{US} = 10$ (d), 25 (e) or 50 (f) ns. See legend of Figure 2.2 for the conventions used in drawing the free-energy maps. The maps were calculated using a grid spacing of 6° . Contours are drawn at successive 5 kJ mol^{-1} intervals starting from the global minimum.

The influence of the force constant increment c (10^{-5} , 10^{-4} or 10^{-3} $\text{kJ mol}^{-1} \text{deg}^2$) is illus-

trated in Figure 2.3a-c, using $t_{LE} = t_{US} = 25$ ns. Considering the number of visits N_v per grid point at the end of the LE build-up phase (displayed as $\ln(N_v + 1)$), increasing the value of c increases the extent of coverage of the Ramachandran map during this phase (fraction of the map for which $N_v > 0$, *i.e.* $\ln(N_v + 1) > 0$). This is expected, because each visit to a grid point is associated with a contribution to the corresponding (repulsive) local Gaussian function that is proportional to c . However, it also increases the roughness of the biasing potential, which represents less accurately the negative of the free-energy hypersurface over the sampled regions. This feature is particularly visible in Figure 2.3c, where the grid structure upon which the biasing potential is constructed becomes apparent in the lowest free-energy basin. Considering the resulting free energy maps $G(\phi, \psi)$ evaluated based on the US sampling phase (after unbiasing), increasing c is also found to increase the extent of the map that is explored during this phase (or, equivalently, the maximal value G_{max} of G above the global minimum up to which the map is sampled). However, the accuracy of the map in the lowest free-energy regions is simultaneously reduced due to the increased roughness of the biasing potential. For example, the position (ϕ, ψ) of the global minimum shifts from $(300^\circ, 168^\circ)$ to $(294^\circ, 114^\circ)$ and $(294^\circ, 162^\circ)$ upon increasing c . The value $c = 10^{-4}$ kJ mol $^{-1}$ deg 2 was found to provide a reasonable balance between sampling extent (up to about 35 kJ mol $^{-1}$ above the global minimum for **C** with $t_{LE} = t_{US} = 25$ ns) and accuracy in the low free-energy regions, and was adopted for all subsequent calculations.

The influence of the build-up and sampling times t_{LE} and t_{US} , respectively, chosen identical here for simplicity ($t_{LE} = t_{US} = 10, 20$ or 50 ns), is illustrated in Figure 2.3d-f. Obviously, increasing the simulation time increases both the extent of the map that is sampled (or, equivalently, the maximal value G_{max} of G above the global minimum up to which the map is sampled), and the accuracy of the map in the lowest free-energy regions. The value $t_{LE} = t_{US} = 50$ ns was found to provide a reasonable balance between sampling extent (up to about 45 kJ mol $^{-1}$ above the global minimum for **C**) and computational cost, and was adopted for all subsequent calculations.

2.4.3 Free-energy maps

As a preliminary verification, the ring conformations of the residues along LEUS simulations of the 11 considered disaccharides (50 ns sampling phase, after unbiasing) were monitored (data not shown). It was found that the ring dihedral angles in both residues are consistent with a leading 4C_1 chair conformation, accounting for 97.5-99.9% of the sampled configurations for the different disaccharides.

As stated previously (Section 2.4.1), the sampling of (a significant fraction of) alternative ring conformations is unlikely. However, it was found necessary to verify this statement also in the context of the LEUS simulations of the disaccharides, since the forced sampling of configurations potentially presenting steric clashes could possibly have enhanced ring transitions.

The free-energy maps $G(\phi, \psi)$ in the subspace of the glycosidic dihedral angles ϕ and ψ as obtained from the 11 LEUS simulations (50 ns sampling phase, after unbiasing) are displayed in Figure 2.4.

These maps can be directly compared with the corresponding maps issued from the 50 ns plain (unbiased) MD simulations in Figure 2.2. This comparison reveals the dramatic sampling enhancement brought by application of the LEUS protocol (at identical sampling times). While the plain MD simulations explored conformations up to a free energy of about 15-20 kJ mol $^{-1}$ above the minimum corresponding to the starting structure, this threshold is extended to about

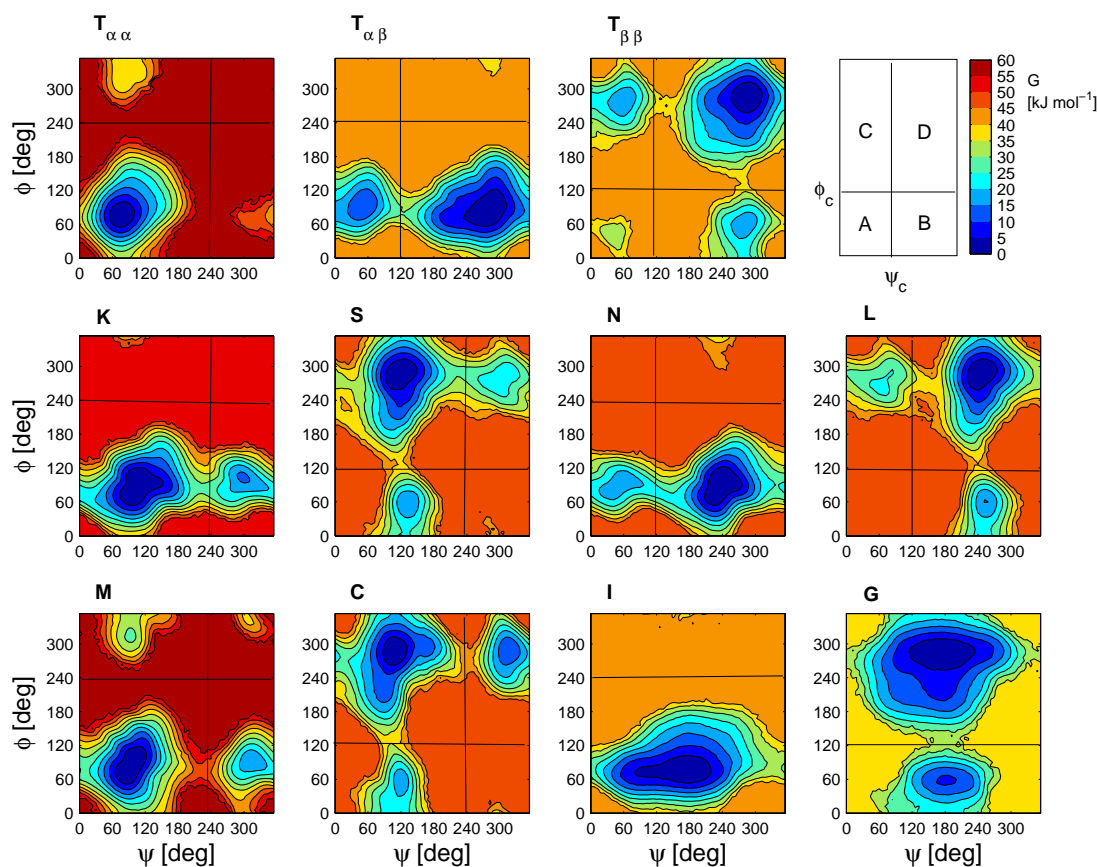


Figure 2.4: Free-energy maps $G(\phi, \psi)$ in the subspace of the glycosidic dihedral angles ϕ and ψ as obtained from LEUS simulations of the 11 considered disaccharides in water at 300 K and 1 bar. The LEUS parameters employed are : $N_l = 2$, $N_g = 32$, $\sigma = 360^\circ / (2N_g) = 5.625^\circ$, $c = 10^{-4} \text{ kJ mol}^{-1} \text{ deg}^2$, $t_{LE} = 50 \text{ ns}$ and $t_{US} = 50 \text{ ns}$. Note that no LEUS enhancement is applied here for the $\bar{\omega}$ dihedral angle in the case of **I** and **G**. See legend of Figure 2.2 for the conventions used in drawing the free-energy maps. The maps were calculated using a grid spacing of 6° . Contours are drawn at successive 5 kJ mol^{-1} intervals starting from the global minimum. The free-energy maps for the $(1 \rightarrow n)$ -linked disaccharides ($n = 1, 2, 3$ or 4) are divided into four distinct conformational regions (A, B, C and D; see top right panel) using cutoff values ϕ_c and ψ_c (Section 2.3.2 and Table 2.1). The analogous division is shown in Figure 2.5 for the $(1 \rightarrow 6)$ -linked disaccharides. The codes employed for the different disaccharides are provided in the legend of Figure 2.1.

$35\text{--}60 \text{ kJ mol}^{-1}$ in the LEUS simulations. This maximum free energy value varies significantly between the considered disaccharides because the LEUS build-up procedure (for a given choice of N_l , N_g , σ , c and t_{LE}) produces a biasing potential that maps out a fixed "volume" of low free energy regions. For disaccharides where these regions are few, narrow or/and shallow, the resulting map will be sampled up to a higher free-energy threshold (for a given choice of t_{US}).

Although the maps issued from the plain MD and LEUS simulations closely resemble each other in the free-energy basins corresponding to the starting structures (and the location of the alternative basins explored in the plain MD simulations of **T** $_{\alpha\beta}$, **S**, **C** and **G**), the latter simulations

reveal a number of additional basins that had remained unvisited in the former ones. However, none of these alternative regions is lower in free energy than the one associated with the experimental structure, *i.e.* these alternative conformational states are all metastable. This observation explains why the plain MD simulations of the non-reducing disaccharides (initiated from the experimental structure) already evidenced excellent agreement with experimental data^{53,54}.

For the topologically symmetrical trehaloses ($\mathbf{T}_{\alpha\alpha}$ and $\mathbf{T}_{\beta\beta}$, where ϕ and ψ are interchangeable), the maps also approximately satisfy the expected mirror symmetry across the diagonal. However, unlike for the plain MD simulations (Section 2.4.1), this (approximate) symmetry now also encompasses two symmetry-equivalent off-diagonal metastable state for $\mathbf{T}_{\alpha\alpha}$ and $\mathbf{T}_{\beta\beta}$ (as well as one additional diagonal metastable state for $\mathbf{T}_{\beta\beta}$).

The (1 \rightarrow 6)-linked disaccharides (\mathbf{I} and \mathbf{G}) possess one additional glycosidic dihedral angle ($\tilde{\omega}$), and corresponding three-dimensional free-energy maps $G(\phi, \psi, \tilde{\omega})$ as obtained from the two additional LEUS simulations (50 ns sampling phase, after unbiasing) where the dihedral angle $\tilde{\omega}$ was also included in the LEUS subspace are displayed in Figure 2.5 (along with the corresponding two-dimensional projections).

The two-dimensional ϕ, ψ -projections issued from the latter simulations are closely similar to the ones displayed in Figure 2.4 for the LEUS simulations excluding $\tilde{\omega}$ from the LEUS subspace, at least in the low free-energy regions. This is expected, since the timescale of about 1 ns associated with the rotation around the $\tilde{\omega}$ dihedral angle^{53,54} is significantly shorter than for ϕ and ψ , so that this rotation should be adequately sampled on the 50 ns timescale, even in the absence of sampling enhancement. For the same reason, the inclusion of this additional dihedral angle in the LEUS protocol does not significantly alter the free-energy threshold up to which the free energy map is sampled (about 40 and 30-35 kJ mol⁻¹ for \mathbf{I} and \mathbf{G} , respectively).

Considering Figures 2.4 and 2.5, it should be stressed that the sampling of the Ramachandran map in the LEUS simulations is extensive but by no means exhaustive, *i.e.* large areas remain unexplored. On the one hand, this is a drawback of the present approach (compared to *e.g.* the systematic scanning of the ϕ and ψ dihedral-angle space), because these unexplored regions could in principle hide further (meta)stable states. This is, however, unlikely in the present case, because these regions correspond to configurations involving sterical clashes between the atoms of the two residues. One might thus say in the present case that the coverage is exhaustive in terms of the sterically allowed regions. On the other hand, the incomplete coverage is also an advantage of the LEUS approach, which keeps the sampling focused on the lowest free energy regions. Although this advantage is moderately significant in the present case (small system size, LEUS subspace of low dimensionality; *i.e.* where an approach by systematic scanning would be a feasible alternative), it becomes absolutely essential in more complex situations.

Visual inspection of the (two- and three-dimensional) free-energy maps reveals the following main features. Except for the (1 \rightarrow 6)-linked disaccharides (\mathbf{I} and \mathbf{G}), all maps are characterized by a lowest free-energy basin corresponding to the experimental structure, along with up to three alternative metastable states. For \mathbf{I} and \mathbf{G} , the (three-dimensional) maps present two low free-energy basins with comparable free energies and differing in their $\tilde{\omega}$ values. For \mathbf{G} only, two metastable states are also observed (at a distinct ϕ value), that also present comparable free energies relative to each other and differ in their $\tilde{\omega}$ values. The $\alpha(1 \rightarrow 1)\alpha$ -linked disaccharide with an axial-axial linkage ($\mathbf{T}_{\alpha\alpha}$) presents a single (significant) basin for $\phi \approx \psi \approx 60^\circ$ (g^+). The $\alpha(1 \rightarrow n)$ -linked disaccharides ($n = 1, 2, 3, 4$ or 6) with an axial-equatorial linkage ($\mathbf{T}_{\alpha\beta}$, \mathbf{K} , \mathbf{N} , \mathbf{M} and \mathbf{I}) only present (significant; stable and, possibly, metastable) minima for $\phi \approx 60^\circ$ (g^+).

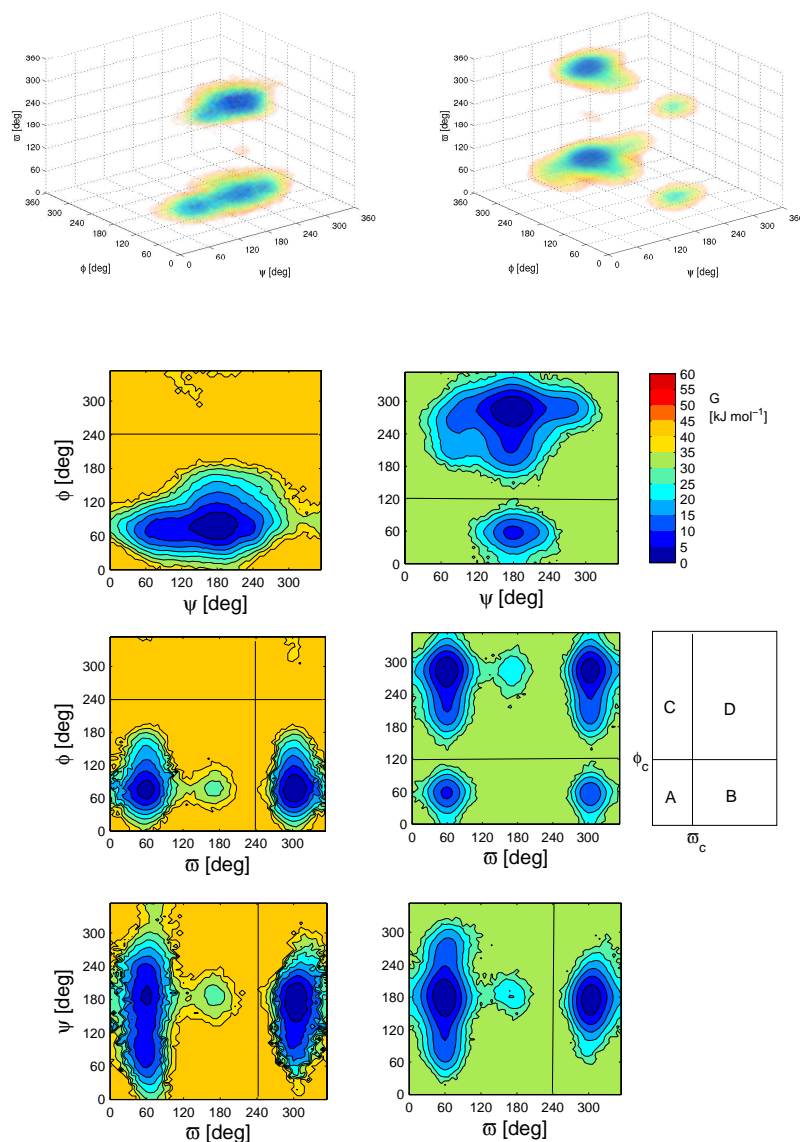


Figure 2.5: Free-energy maps $G(\phi, \psi, \tilde{\omega})$ (top) in the subspace of the glycosidic dihedral angles ϕ , ψ and $\tilde{\omega}$ as obtained from LEUS simulations of the (1 \rightarrow 6)-linked disaccharides **I** (left) and **G** (right) in water at 300 K and 1 bar, as well as two-dimensional projections of these maps (bottom). The LEUS parameters employed are : $N_l = 3$, $N_g = 32$, $\sigma = 360^\circ / (2N_g) = 5.625^\circ$, $c = 10^{-4} \text{ kJ mol}^{-1} \text{ deg}^3$, $t_{LE} = 50 \text{ ns}$ and $t_{US} = 50 \text{ ns}$. Note that a LEUS enhancement is also applied here for the $\tilde{\omega}$ dihedral angle. See legend of Figure 2.2 for the conventions used in drawing the two-dimensional free-energy maps. The maps were calculated using a grid spacing of 6° . Contours are drawn at successive 5 kJ mol^{-1} intervals starting from the global minimum. The free-energy maps for the (1 \rightarrow 6)-linked disaccharides are divided into four distinct conformational regions (A, B, C and D; see panel at the right of the $\tilde{\omega}, \phi$ -projection) using cutoff values ϕ_c and $\tilde{\omega}_c$ (Section 2.3.2 and Table 2.1). The analogous division is shown in Figure 2.4 for the (1 \rightarrow n)-linked disaccharides ($n = 1, 2, 3$ or 4). The codes employed for the different disaccharides are provided in the legend of Figure 2.1.

In contrast, the $\beta(1 \rightarrow n)$ -linked disaccharides ($n = 1, 2, 3, 4$ or 6) with an equatorial-equatorial linkage ($\mathbf{T}_{\beta\beta}$, \mathbf{S} , \mathbf{L} , \mathbf{C} and \mathbf{G}) present (stable and, possibly, metastable) minima for both $\phi \approx 300^\circ$ (g^-) and $\phi \approx 60^\circ$ (g^+), the most stable state being always associated with the former value.

The above observations are compatible with the *exo*-anomeric effect^{1,2,158–165} along with the presence of steric constraints. The *exo*-anomeric effect implies a stereoelectronic preference for $\phi \approx 60^\circ$ (g^+) and $\phi \approx 300^\circ$ (g^-). For a non-reducing residue involved in an α -linkage, the latter value of the dihedral angle is, however, strongly disfavored by steric constraints (similar considerations apply for ψ in $\mathbf{T}_{\alpha\alpha}$). For a non-reducing residue involved in a β -linkage, it is the former value of the dihedral angle that is slightly disfavored by steric constraints (similar considerations apply for ψ in $\mathbf{T}_{\alpha\beta}$ and $\mathbf{T}_{\beta\beta}$). A particular consequence of these effects is that the diaxially-linked disaccharide trehalose ($\mathbf{T}_{\alpha\alpha}$), for which the most restrictive constraint (α -anomer) applies to both ϕ and ψ , appears to be particularly rigid within the disaccharide series. For the same reason, considering the other disaccharides, the α -linked (axial-equatorial) disaccharides ($\mathbf{T}_{\alpha\beta}$, \mathbf{K} , \mathbf{N} , \mathbf{M} and \mathbf{I}) appear to be systematically more rigid compared to the corresponding β -linked (equatorial-equatorial) disaccharides ($\mathbf{T}_{\beta\beta}$, \mathbf{S} , \mathbf{L} , \mathbf{C} and \mathbf{G}).

Another interesting observation is that the the maps of the $(1 \rightarrow 1)$ - and $(1 \rightarrow 3)$ -linked disaccharides with matching linkage anomery (α : $\mathbf{T}_{\alpha\beta}$ vs \mathbf{N} ; β : $\mathbf{T}_{\beta\beta}$ vs \mathbf{L}) are similar, as well as the maps of the the $(1 \rightarrow 2)$ - and $(1 \rightarrow 4)$ -linked disaccharides with matching linkage anomery (α : \mathbf{K} vs \mathbf{M} ; β : \mathbf{S} vs \mathbf{C}). Furthermore, the corresponding maps of the two sets for a given linkage anomery (α : $\mathbf{T}_{\alpha\beta} \sim \mathbf{N}$ vs $\mathbf{K} \sim \mathbf{M}$; β : $\mathbf{T}_{\beta\beta} \sim \mathbf{L}$ vs $\mathbf{S} \sim \mathbf{C}$) are also approximately symmetrical with respect to a sign change of the ψ dihedral angle.

As illustrated in Figure 2.6, the above observations can be easily understood when considering the approximate topological symmetries existing between the 8 $(1 \rightarrow n)$ -linked disaccharides ($n = 1, 2, 3$ or 4 ; excluding $\mathbf{T}_{\alpha\alpha}$). The simplifications used in the approximate topological representation evidencing these symmetries are : (i) all ring atoms (carbon, oxygen; irrespective of their substitution) are considered equivalent; (ii) all free exocyclic hydroxyl and hydroxymethyl groups are omitted; (iii) both rings are in a 4C_1 chair conformation.

The observation of these approximate symmetries is interesting, because it suggests that the detailed positioning of the free exocyclic hydroxyl and hydroxymethyl groups as well as of the ring oxygen atoms relative to the glycosidic linkage only exerts a minor influence on the conformational properties of these disaccharides (provided that the sign of ψ is inverted for $(1 \rightarrow n)$ -linked disaccharides with n even or n odd and excepting the *exo*-anomeric influence of a ring oxygen atom next to the linkage). The determinant factors for the overall shape of the free-energy map are thus the ring conformation (4C_1) and the axial or equatorial nature of the C_1 - O_1 bond (α versus β -linkage). In other words, short-range through-bond (stereoelectronic) and through-space (steric) effects (the latter considering the relative positioning of the rings and omitting the details of their local substitution patterns) appear to be much more significant than intramolecular H-bonding and specific solvation effects. Of course, these considerations are restricted in the present case to disaccharides with residues presenting a leading 4C_1 chair conformation and involving exocyclic groups that are all equatorial (except the C_1 - O_1 bond in the case of the α -linkages).

The maps for the $(1 \rightarrow 6)$ -linked disaccharides (\mathbf{I} and \mathbf{G}) only present (stable and, possibly, metastable) minima for $\tilde{\omega} \approx 60^\circ$ (g^+) and $\tilde{\omega} \approx 300^\circ$ (g^-). These correspond to the conformations usually referred to as *gt* and *gg*, respectively (the first letter referring to $\tilde{\omega}$ and the second to ω ; Section 2.3.2). The (near) absence of the *tg* conformer, also evidenced by NMR^{166–175}, CD^{105,176}, and x-ray crystallography^{177,178} experiments, is a well known feature of D-hexopyranoses and

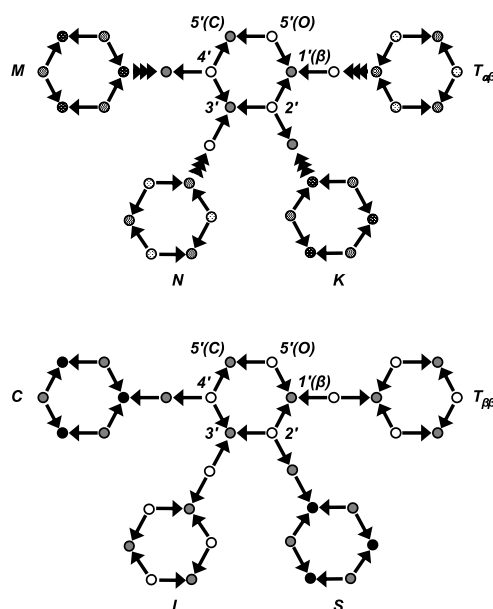


Figure 2.6: Approximate topological symmetries existing between the 8 ($1 \rightarrow n$)-linked ($n = 1, 2, 3$ or 4) disaccharides (excluding $\mathbf{T}_{\alpha\alpha}$). The top drawing corresponds to α -linked and the bottom drawing to β -linked disaccharides. In each drawing, the four possible connectivities of the first (non-reducing) residue onto the second (reducing except for $\mathbf{T}_{\alpha\beta}$ and $\mathbf{T}_{\beta\beta}$) residue are displayed. The simplifications used in the approximate topological representation are : (i) all ring atoms are considered equivalent; (ii) all free exocyclic hydroxyl and hydroxymethyl groups are omitted; (iii) both rings are in a 4C_1 chair conformation. The rings are viewed perpendicularly to their midplanes from the side where atom 1 is in the back and atom 4 in the front. The bonds between ring atoms and from these to the glycosidic oxygen atom are marked with a simple arrow, except axial bonds to the glycosidic oxygen atom that are marked by a triple-arrow (arrows point from the atom that is at the back to the atom that is at the front; the latter atom is then also represented in a darker color compared to the former one). In this representation, approximate topological symmetries connect the pairs $\mathbf{T}_{\alpha\beta} \leftrightarrow \mathbf{N}$, $\mathbf{K} \leftrightarrow \mathbf{M}$, $\mathbf{T}_{\beta\beta} \leftrightarrow \mathbf{L}$ and $\mathbf{S} \leftrightarrow \mathbf{C}$. Furthermore, a mirroring of the second residue (i.e. the one drawn at the center) with respect to its midplane (which implies a sign change of the ψ dihedral angle), connects the pairs of pairs ($\mathbf{T}_{\alpha\beta}, \mathbf{N}$) \leftrightarrow (\mathbf{K}, \mathbf{M}) and ($\mathbf{T}_{\beta\beta}, \mathbf{L}$) \leftrightarrow (\mathbf{S}, \mathbf{C}). The codes employed for the different disaccharides are provided in the legend of Figure 2.1.

D-hexopyranosides presenting the *gluco* configuration (or C_2 or/and C_3 epimer thereof). As discussed in details elsewhere⁴¹, however, the exact origin of this conformational preference is still matter of discussion, different authors invoking distinct effects as a main cause (e.g. the *gauche* effect^{169, 179–182}, 1,3-*syn*-diaxial repulsion with the C_4 hydroxyl group^{169, 179, 180}, or specific solvation effects^{172, 183–189}). Note, finally, that the $\beta(1 \rightarrow 6)$ -linked disaccharides (**I** and **G**) are expected to be the most flexible ones along the disaccharide series, due to the presence of an extra glyco-

sidic dihedral angle ($\tilde{\omega}$) within this linkage. The flexibility of **G** is also expected to be somewhat larger than that of **I** (equatorial-equatorial vs axial-equatorial linkages, respectively; see above).

2.4.4 Conformational states: definition and sampling

For the ease of analysis, the free-energy maps displayed in Figures 2.4 and 2.5 were divided into four distinct conformational regions ($X = A, B, C$ and D) using cutoff values ϕ_c and ψ_c for the $(1 \rightarrow n)$ -linked disaccharides ($n = 1, 2, 3$ or 4), or ϕ_c and $\tilde{\omega}_c$ for the $(1 \rightarrow 6)$ -linked disaccharides, as described in Section 2.3.2 (the cutoff values selected for the different disaccharides are reported in Table 2.1). In the following, those conformational regions that correspond to a free energy basin are further referred to as conformational states and labeled using the letter associated to the corresponding region.

The relative populations of the different states, along with the locations of the associated free-energy minima and the corresponding average values of the glycosidic dihedral angles, are reported in Table 2.1. Illustrative structures corresponding to the free-energy minima within the most stable states are also represented in Figure 2.7.

All $(1 \rightarrow n)$ -linked disaccharides ($n = 1, 2, 3$ or 4) are characterized by a single overwhelmingly populated state (A for **T $_{\alpha\alpha}$** , **K** and **M**; B for **T $_{\alpha\beta}$** and **N**; C for **S** and **C**; D for **T $_{\beta\beta}$** and **L**; relative population above 98.5% in all cases), along with the possible presence of up to three alternative states with marginal populations. The $(1 \rightarrow 6)$ -linked disaccharides are characterized by two significantly populated states (B and A for **I**; C and D for **G**; the population of the former one being in both cases about twice that of the latter one), along with two alternative states with marginal populations for **G** only.

As an illustration of the sampling enhancement achieved by the LEUS approach, the numbers of transitions $n_{X \rightarrow Y}$ observed between these states during the LEUS simulations are reported in Table 2.2 (the corresponding numbers for the unbiased MD simulations are also indicated between parentheses for comparison, whenever differing from zero). Comparison with Table 2.1 shows that nearly all (forward and backward) transitions between states with non-negligible populations ($\geq 0.2\%$) occur about hundred times or more during the sampling phase of the LEUS simulations, while they occurred a few times at best during the plain MD simulations (Section 2.4.1). The only exceptions are the transitions involving states A and B of **G** which are less frequent, probably because the dihedral angle $\tilde{\omega}$ is not included in the LEUS subspace in these simulations. Note that the $n_{X \rightarrow Y}$ values obtained from the LEUS simulations cannot be used to evaluate corresponding transition timescales for the real system, since the dynamics of the system is entirely unphysical in this case.

There is a price to pay in terms of statistical efficiency when using a biased ensemble approach such as LEUS to evaluate ensemble averages of observables appropriate for the real (unbiased) system. This efficiency is estimated here by the parameter F_s (Eqs. 2.5), found to be in the range 0.05-0.20 for the different LEUS simulations. Thus, in broad terms, only an ‘‘effective’’ 5-20% fraction of the N_f trajectory frames stored along these simulations (10^6 for the 50 ns sampling phases of the present simulations with coordinates stored every 0.05 ps) actually contributes to the statistics on the ensemble averages of observables appropriate for the real (unbiased) system.

This means that at identical N_f , plain MD simulations would in principle provide 5-20 times more statistics. On the other hand, the sampling enhancement permits : (i) to overcome higher free-energy barriers and include the statistics on alternative metastable states into the averaging;

| System | ϕ_c | ψ_c | $\tilde{\omega}_c$ | P_A | P_B | P_C | P_D | $(\phi_m, \psi_m, \tilde{\omega}_m)_A$ $(\phi_a, \psi_a, \tilde{\omega}_a)_A$ | $(\phi_m, \psi_m, \tilde{\omega}_m)_B$ $(\phi_a, \psi_a, \tilde{\omega}_a)_B$ | $(\phi_m, \psi_m, \tilde{\omega}_m)_C$ $(\phi_a, \psi_a, \tilde{\omega}_a)_C$ | $(\phi_m, \psi_m, \tilde{\omega}_m)_D$ $(\phi_a, \psi_a, \tilde{\omega}_a)_D$ |
|--------------------------------------|----------|----------|--------------------|---------------|--------------|--------------|--------------|--|--|--|--|
| | [deg] | [deg] | [deg] | [%] | [%] | [%] | [%] | [deg] | [deg] | [deg] | [deg] |
| T$_{\alpha\alpha}$ | 240 | 240 | - | 100.00 | 0.00 | 0.00 | - | (78, 78) (77.7, 77.7) | - | - | - |
| T$_{\alpha\beta}$ | 240 | 120 | - | 0.93 | 99.07 | - | 0.00 | (96, 54) (84.8, -86.2) (94.3, 50.2) | (84, -78) (84.8, -86.2) | - | - |
| T$_{\beta\beta}$ | 120 | 120 | - | 0.00 | 0.15 | 0.18 | 99.67 | - | (60, -78) (57.4, -81.2) | (-84, 60) (-80.8, 57.6) | (-78, -72) (-75.6, -75.6) |
| K | 240 | 240 | - | 99.75 | 0.25 | 0.00 | - | (102, 102) (87.0, 109.9) | (102, -60) (93.2, -65.6) | - | - |
| S | 120 | 240 | - | 0.17 | 0.00 | 99.80 | 0.03 | (60, 132) (57.0, 134.0) | - | (-72, 114) (-75.2, 118.9) | (-72, -48) (-82.3, -59.7) |
| N | 240 | 120 | - | 0.10 | 99.90 | - | 0.00 | (96, 60) (88.7, 70.2) | (96, -120) (89.0, -121.5) | - | - |
| L | 120 | 120 | - | 0.00 | 0.13 | 0.03 | 99.84 | - | (60, -108) (55.3, -105.6) | (-96, 66) (-87.1, 64.8) | (-72, -114) (-73.6, -114.1) |
| M | 240 | 240 | - | 99.87 | 0.13 | 0.00 | 0.00 | (72, 90) (81.7, 92.5) | (90, -42) (87.9, -51.2) | - | - |
| C | 120 | 240 | - | 0.12 | 0.00 | 98.57 | 1.31 | (60, 120) (49.4, 117.3) | - | (-72, 108) (-73.2, 108.3) | (-72, -42) (-73.7, -44.0) |
| I | 240 | - | 240 | 31.58 | 68.42 | - | - | (72, 180, 60) (78.0, 178.5, 66.5) | (78, 174, -54) (78.4, 174.8, -56.1) | - | - |
| G | 120 | - | 240 | 1.55 | 0.85 | 61.33 | 36.27 | (66, -174, 66) (57.5, -176.0, 64.8) | (60, 180, -48) (56.2, 171.6, -56.8) | (-66, -162, 66) (-80.1, -178.8, 65.4) | (-72, -174, -48) (-79.4, 168.9, -57.1) |

Table 2.1: Conformational states and characteristic parameters identified in LEUS simulations of the 11 considered disaccharides in water at 300 K and 1 bar. The reported quantities are the cutoff values ϕ_c , ψ_c and $\tilde{\omega}_c$ used to define the conformational states ($X = A, B, C$ and D ; Section 2.3.2) based on the free-energy maps (Figures 2.4 and 2.5), the corresponding fractional populations P_X (after unbiasing), the locations $(\phi_m, \psi_m, \tilde{\omega}_m)$ of the associated free-energy minima and the corresponding average values $(\phi_a, \psi_a, \tilde{\omega}_a)$ of the glycosidic dihedral angles (over the state). Empty entries for P_X indicates regions that were never visited. Empty entries for $(\phi_m, \psi_m, \tilde{\omega}_m)$ and $(\phi_a, \psi_a, \tilde{\omega}_a)$ indicate regions that were too poorly sampled ($P_X < 0.01\%$) for a reliable evaluation. Entries corresponding to the most populated states ($P_X > 30\%$) are printed in boldface for readability. The dihedral-angle values were selected here within the period $]-180^\circ, 180^\circ]$. The codes employed for the different disaccharides are provided in the legend of Figure 2.1.

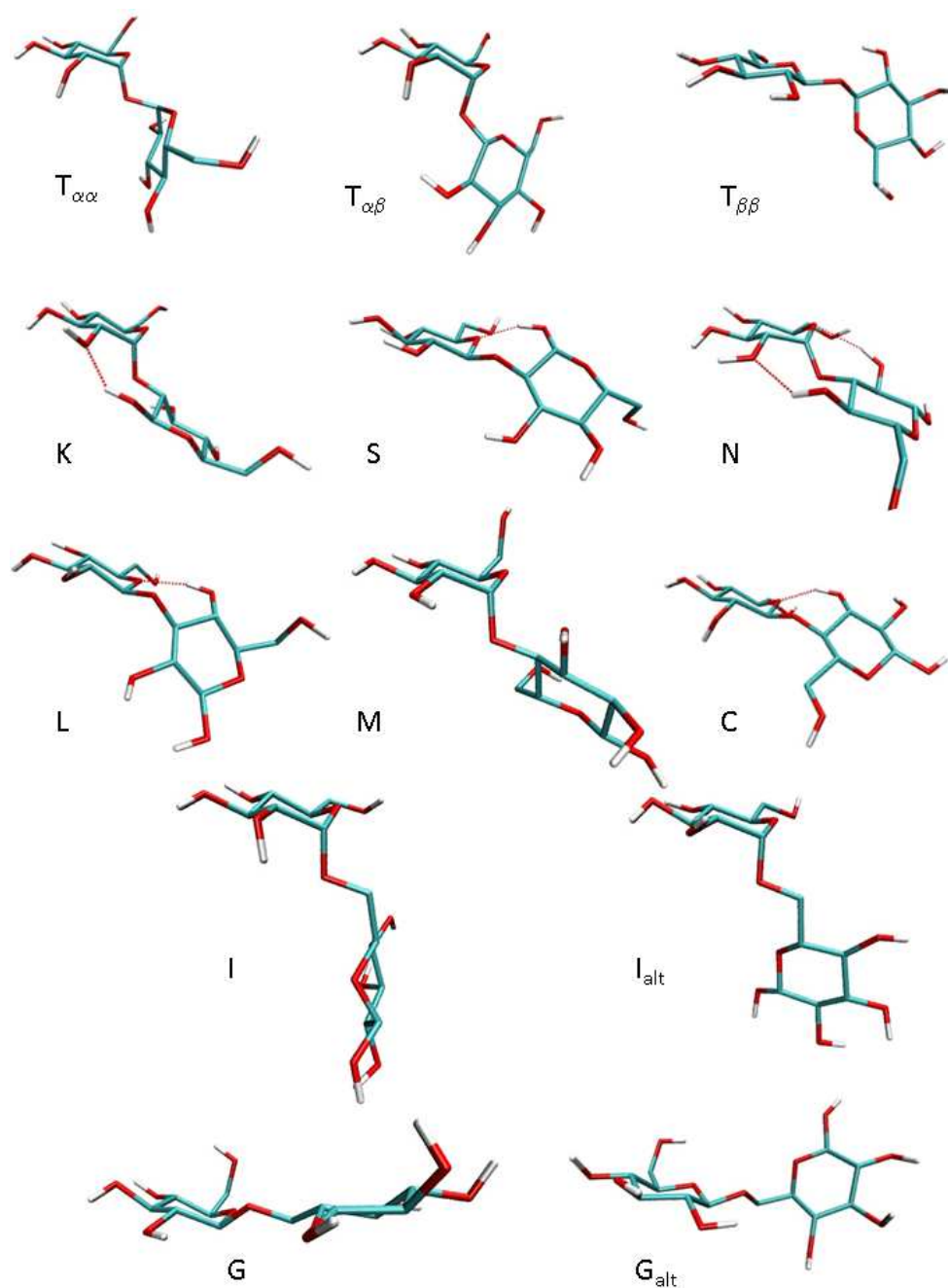


Figure 2.7: Illustrative structures corresponding to the minimum free-energy points of the most stable conformational states identified in LEUS simulations of the 11 considered disaccharides in water at 300 K and 1 bar. The represented structures are taken from the simulations and characterized by glycosidic dihedral angle values within 5° of the values $(\Phi_m, \Psi_m, \tilde{\omega}_m)$ corresponding to the minimum free-energy point of the state (Table 2.1). They were also selected to illustrate the main intramolecular H-bonds present in this state (Table 2.5; except for **M** where the dominant H-bonds were incompatible with this specific structure). For the (1 \rightarrow 6)-linked disaccharide, both the structures corresponding to the states with the highest (**I** and **G**) and second-highest (**I_{alt}** and **G_{alt}**) populations are displayed. The codes employed for the different disaccharides are provided in the legend of Figure 2.1.

| System | $n_{A \rightarrow B}$ | $n_{A \rightarrow C}$ | $n_{A \rightarrow D}$ | $n_{B \rightarrow C}$ | $n_{B \rightarrow D}$ | $n_{C \rightarrow D}$ |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | $n_{B \rightarrow A}$ | $n_{C \rightarrow A}$ | $n_{D \rightarrow A}$ | $n_{C \rightarrow B}$ | $n_{D \rightarrow B}$ | $n_{D \rightarrow C}$ |
| T_{αα} | - | - | - | - | - | - |
| T_{αβ} | 153(7) | - | - | - | - | - |
| | 155(6) | - | - | - | - | - |
| T_{ββ} | - | - | - | 29 | 63 | 75 |
| | - | - | - | 31 | 63 | 77 |
| K | 109 | - | - | - | - | - |
| | 109 | - | - | - | - | - |
| S | - | 73(1) | 14 | - | - | 67 |
| | - | 69(1) | 20 | - | - | 63 |
| N | 95 | - | - | - | - | - |
| | 95 | - | - | - | - | - |
| L | - | - | - | 29 | 46 | 79 |
| | - | - | - | 23 | 51 | 73 |
| M | 111 | - | - | - | - | - |
| | 110 | - | - | 2 | - | - |
| C | - | 44(2) | 3 | - | - | 92 |
| | - | 43(2) | 3 | - | - | 91 |
| I | 197(18) | - | - | - | - | - |
| | 197(19) | - | - | - | - | - |
| G | 7 | 14(2) | 1 | 10 | 5 | 160 |
| | 6 | 16(2) | 1 | 11 | 3 | 163 |

Table 2.2: Numbers of transitions $n_{X \rightarrow Y}$ between the conformational states ($X, Y = A, B, C$ and D ; Section 2.3.2) as observed in LEUS simulations of the 11 considered disaccharides in water at 300 K and 1 bar. The corresponding number of transitions in the 50 ns plain (unbiased) MD simulations are reported between parentheses for comparison (if differing from zero). A transition from state X to state Y is assumed to have occurred whenever the system spent at least 5 ps in region X followed by at least 5 ps in region Y . Empty entries indicate transitions that were never observed. The codes employed for the different disaccharides are provided in the legend of Figure 2.1 and the definitions of the conformational states in Section 2.3.2 and Table 2.1 (see also Figures 2.4 and 2.5).

(*ii*) to explore a much broader volume of the conformational space accessible to the system. The reduced statistical efficiency may be (at least in part) compensated by the storage a larger number of frames along the trajectories. This was done in the present simulations (Section 2.3.1), where the storage rate of trajectory frames was increased by a factor 10 for the LEUS simulations compared to the plain MD simulations (it was also verified that the value of F_s is essentially invariant for filtered trajectories with N_f ranging from 10^6 down to about 10^3).

2.4.5 Comparison with experimental data

The glycosidic linkage conformations associated with the most stable states (value of the glycosidic dihedral angles at the free-energy minimum and corresponding average values over the state) identified in the LEUS simulations of the 11 considered disaccharides (50 ns sampling phase, after unbiasing) are compared with available experimental and theoretical data in Table 2.3. The corresponding probability distributions around the individual glycosidic dihedral angles were also monitored (data not shown). For the non-reducing disaccharides, they do not differ significantly from those reported previously based on unbiased MD simulations⁵³ (see Figure 2 therein). The rotational properties of the unfunctionalized hydroxymethyl groups are not discussed, since they were investigated in details previously^{41,53}.

In general, the LEUS simulation results are in good qualitative agreement with experimental and theoretical data from the literature, as were the previously reported plain MD simulation results for the non-reducing disaccharides^{53,54}. The best agreement is met for the ϕ glycosidic dihedral angle, the preferences around which are largely dictated by the *exo*-anomeric effect (Section 2.4.3).

For the α -linked disaccharides ($\mathbf{T}_{\alpha\alpha}$, $\mathbf{T}_{\alpha\beta}$, \mathbf{K} , \mathbf{N} , \mathbf{M} and \mathbf{I}) the average values of the ϕ dihedral angle (as well as those of the ψ dihedral angle for $\mathbf{T}_{\alpha\alpha}$) are in the range $[60^\circ; 100^\circ]$ for nearly all entries of Table 2.3, while for the β -linked disaccharides ($\mathbf{T}_{\beta\beta}$, \mathbf{S} , \mathbf{L} , \mathbf{C} and \mathbf{G}) they are nearly all in the range $[-90^\circ; -60^\circ]$ (note, however, the exception of \mathbf{C} based on one set of calculation¹¹⁰). The variability is more important for the average values of the ψ dihedral angle. Considering the entries of Table 2.3 corresponding to the $(1 \rightarrow n)$ -linked disaccharides ($n = 1, 2, 3$ or 4), all reducing disaccharides with a *R*-configuration at C'_n (\mathbf{K} , \mathbf{S} , \mathbf{M} and \mathbf{C}), as well as $\mathbf{T}_{\alpha\alpha}$, typically present values in the range $[60^\circ; 120^\circ]$. In contrast, the reducing disaccharides with a *S*-configuration at C'_n (\mathbf{N} and \mathbf{L}), as well as $\mathbf{T}_{\alpha\beta}$ and $\mathbf{T}_{\beta\beta}$, typically present values in the range $[-140^\circ; -70^\circ]$. Finally, for the $(1 \rightarrow 6)$ -linked disaccharides (\mathbf{I} and \mathbf{G}), the ψ values are generally found in the range $[160^\circ; 180^\circ]$. The existence of two alternative conformations at equilibrium (differing in the value of the $\tilde{\omega}$ dihedral angle) for the $(1 \rightarrow 6)$ -linked disaccharides (\mathbf{I} and \mathbf{G}) also agrees with the results from NMR measurements^{93,96} and independent simulations^{53,131}. The population ratio between the *tg* ($\tilde{\omega} = 180^\circ$), *gg* ($\tilde{\omega} = -60^\circ$) and *gt* ($\tilde{\omega} = +60^\circ$) conformers in the LEUS simulations is $[0:68:32]$ for \mathbf{I} and $[0:37:63]$ for \mathbf{G} . These distributions are in remarkable agreement with the corresponding ratios of $[1:70:29]$ and $[0:34:66]$ determined based on NMR measurements^{93,96}.

Finally, the Ramachandran free energy maps (Figures 2.4 and 2.5) are also found to present a good qualitative agreement with corresponding MM3 adiabatic maps for the different disaccharides^{112–117}.

| System | $(\phi_m, \psi_m, \tilde{\omega}_m)_X$ [deg] | $(\phi_a, \psi_a, \tilde{\omega}_a)_X$ [deg] | $(\phi, \psi, \tilde{\omega})_{NMR}$ [deg] | $(\phi, \psi, \tilde{\omega})_{x-ray}$ [deg] | $(\phi, \psi, \tilde{\omega})_{MM/MD}$ [deg] | $(\phi, \psi, \tilde{\omega})_{QM}$ [deg] |
|-----------------------|---|---|---|--|--|--|
| T_{αα} | (78,78) _A | (77.7, 77.7) _A | (76,76) ¹⁰⁰ (77.8, 77.8) ⁹⁹ | (60.8,60.1) ⁹¹ , (74.7, 75.5) ⁸⁸ (61.7, 74.8) ⁹² (68.6,-93.4) ⁹⁰ | (65,65) ¹¹⁹ ,(76,76) ^{120,137} (61,64) ¹²⁶ , (70,69) ¹³⁶ (75,-80) ¹²⁰ | (75.61,75.61) ¹⁰⁷ |
| T_{αβ} | (84,-78) _B | (84.8, -86.2) _B | | | | (65.82,-96.57) ¹⁰⁷ |
| T_{ββ} | (-78,-72) _D | (-75.6, -75.6) _D | | (-75.6,-75.6) ⁸⁹ | (-77,-77) ¹²⁰ | (-71.37,-71.37) ¹⁰⁷ |
| K | (102,102) _A | (87.0, 109.9) _A | | | (95.8,114.7) ⁵³ ,(80.6,78.8) ¹¹³ | |
| S | (-72,114) _C | (-75.2, 118.9) _C | (-80,110) ¹⁰¹ | (-81.2,128.5) ² | (-83.8,119.8) ⁵³ ,(-71.2,111.7) ¹¹² | |
| N | (96,-120) _B | (89.0, -121.5) _B | (81.2,-142.6) ⁹⁹ | (99.9, -135.8) ² | (95.7,-119.4) ⁵³ ,(85.2,-72.6) ¹¹³ (97.3, -119.0) ¹³⁵ , (84.4, -129.6) ⁹⁹ | (73.23,-153) ¹⁰⁷ |
| L | (-72,-114) _D | (-73.6, -114.1) _D | (-85.8,-143.8) ⁹⁹ | (-69.1,-109.1) ² (-93.6, -161.0) ⁸⁷ (96.8,105.2) ² | (-72.6,-114.2) ⁵³ ,(-85.0,-162.3) ¹¹² (-58.6, -94.5) ¹³¹ | (-65.67,-134.4) ¹⁰⁷ |
| M | (72,90) _A | (81.7, 92.5) _A | (88.5,95.5) ⁹⁹ | | (87.2,100.5) ⁵³ ,(97.3, 98.3) ¹¹³ (94.7,106.1) ⁹⁹ , (80, 89) ¹²⁴ | (73.0,87.2) ¹⁰⁷ |
| C | (-72,108) _C | (-73.2, 108.3) _C | (-87.8,98.9) ⁹⁹ | (-75.6,108.2) ² (-77.7, 102.1) ⁸⁶ | (-77.5,111.4) ⁵³ ,(-84.0, 73.8) ¹¹² (-75.9,119.8) ⁹⁹ | (-66.2,106.4) ¹⁰⁷ (59.4, 119.4) ¹¹⁰ |
| I | (78,174,-54) _B (72,180,60) _A | (78.4, 174.8, -56.1) _B (gg) (78.0, 178.5, 66.5) _A (gt) [0:68:32] | [1:70:29] ⁹³ | (70.8,167.3, 74.7) ² | (82.7,167.5)[0:44:56] ⁵³ (75.8,-173.6, 69.5) ¹¹⁴ ,(71, 196, 74) ¹⁰⁴ | (116.3, 163.2, -45.7) ¹⁰⁶ [39:12:49] ¹¹¹ |
| G | (-66,-162,66) _C (-72,-174,-48) _D | (-80.1, -178.8, 65.4) _C (gt) (-79.4, 168.9, -57.1) _D (gg) [0:37:63] | (-71.4,-129.0) ⁹⁹ [0:34:66] ⁹⁶ | (-58.3,-156.3,-61.5) ² | (-81.6,173.1)[0:46:54] ⁵³ (-77.5,178.4, 68.9) ¹¹⁴ (-60.8,-177.0) ¹³¹ [2:36:62] ¹³¹ | |

Table 2.3: Comparison of the most stable glycosidic linkage conformations identified in LEUS simulations of the 11 considered disaccharides in water at 300 K and 1 bar with available experimental and theoretical data. The locations $(\phi_m, \psi_m, \tilde{\omega}_m)$ of the identified (most stable) free-energy minima and the corresponding average values $(\phi_a, \psi_a, \tilde{\omega}_a)$ of the glycosidic dihedral angles (Table 2.1) are compared with (average) results $(\phi, \psi, \tilde{\omega})$ from experiment (NMR, x-ray) or theory (MM, MD, QM). For the (1 → 6)-linked disaccharides (**I** and **G**), $\tilde{\omega}$ may either be reported explicitly (single value) or in the form of relative percentages of the tg:gg:gt conformers (between square brackets). These conformers are defined in terms of the dihedral angles ω ($O_1-C'_6-C'_5-C'_4$) and $\tilde{\omega}$ ($O_1-C'_6-C'_5-O'_5$; used in this study), the first letter referring to $\tilde{\omega}$ and the second to ω (tg: $\tilde{\omega} \approx 180^\circ$; gg: $\tilde{\omega} \approx -60^\circ$; gt: $\tilde{\omega} \approx 60^\circ$). Values from the literature reported in terms of the dihedral angles ϕ_H ($HC_1-C_1-O_1-C'_n$) and ψ_H ($C_1-O_1-C'_n-HC'_n$) were transformed to ϕ ($O_5-C_1-O_1-C'_n$) and ψ ($C_1-O_1-C'_n-C'_{n-1}$) values (used in this study) so as to permit a direct comparison. This transformation reads²: (i) $\phi = \phi_H \pm 120^\circ$, where "+" applies to the α -linked disaccharides (**T_{αα}**, **T_{αβ}**, **K**, **N**, **M**, **I**) and "-" applies to the β -linked disaccharides (**T_{ββ}**, **S**, **L**, **C**, **G**); (ii) $\psi = \psi_H \pm 120^\circ$ (except for the trehaloses), where "+" applies to the disaccharides with a R-configuration at C'_n (**K**, **S**, **M**, **C**) and "-" to those with a S-configuration at C'_n (**N**, **L**); (iii) $\phi = \phi_H \pm 120^\circ$ (for the trehaloses), where "+" applies to **T_{αα}** and "-" applies to **T_{αβ}** and **T_{ββ}**. Note that ψ_H is not defined for **I** and **G** because C'_n corresponds to a methylene group in this case. The dihedral-angle values were selected here within the period $]-180^\circ, 180^\circ]$. The codes employed for the different disaccharides are provided in the legend of Figure 2.1 and the definitions of the conformational states in Section 2.3.2 and Table 2.1 (see also Figures 2.4 and 2.5).

2.4.6 Conformational states: thermodynamics and kinetics

The parameters characterizing the relative free energies and interconversion barriers between the conformational states identified in the LEUS simulations of the 11 considered disaccharides (50 ns sampling phase, after unbiasing) are reported in Table 2.4. Only data concerning significantly populated conformational states ($P_X \geq 0.01\%$; Table 2.1) is reported. Note also that only transitions along direct pathways (*i.e.* not involving an intermediate conformational state) are considered. These exclude the transitions $C \leftrightarrow B$ for $\mathbf{T}_{\beta\beta}$ ($C \leftrightarrow D \leftrightarrow B$), $A \leftrightarrow D$ for \mathbf{S} ($A \leftrightarrow C \leftrightarrow D$), $B \leftrightarrow C$ for \mathbf{L} ($B \leftrightarrow D \leftrightarrow C$), $D \leftrightarrow A$ for \mathbf{C} ($D \leftrightarrow C \leftrightarrow A$), $C \leftrightarrow B$ for \mathbf{G} ($C \leftrightarrow A \leftrightarrow B$ or $C \leftrightarrow D \leftrightarrow B$) and $D \leftrightarrow A$ for \mathbf{G} ($D \leftrightarrow B \leftrightarrow A$ or $D \leftrightarrow C \leftrightarrow A$). The parameters (free energy differences and barriers) corresponding to these indirect transitions can easily be deduced from the data in the table. Since the state-to-state free energy differences ($\Delta\tilde{G}_{x \rightarrow y}$) do not differ significantly from the corresponding minimum-to-minimum differences ($\Delta G_{x \rightarrow y}$) the following discussion will consider only the latter quantities.

For the $(1 \rightarrow n)$ -linked disaccharides ($n = 1, 2, 3$ or 4), the one (axial-equatorial linkage; $\mathbf{T}_{\alpha\beta}$, \mathbf{K} , \mathbf{N} and \mathbf{M}) or two (equatorial-equatorial linkage; $\mathbf{T}_{\beta\beta}$, \mathbf{S} , \mathbf{L} and \mathbf{C}) significantly populated metastable states lie within 10 - 22 kJ mol^{-1} of the most stable one (there is no significantly populated metastable state for the single disaccharide $\mathbf{T}_{\alpha\alpha}$ with an axial-axial linkage). The estimated transition timescales (half-times) for the forward transitions to these metastable states (from the most stable state) range between about 100 ps (600 ps, omitting $\mathbf{T}_{\alpha\beta}$) and 20 ns.

The estimated timescales for the corresponding backward transitions range between about 1 and 50 ps. In the absence of kinetic information on transition trajectories (which cannot be obtained from LEUS simulations since the dynamics of the system is unphysical), the transmission coefficients $\kappa_{X \rightarrow Y}$ used to evaluate these timescales *via* the TST⁶⁹ approximation (Eqs. 2.4) were set to one as a first approximation. However, assuming that the $\kappa_{X \rightarrow Y}$ values are of similar magnitudes for all the forward transitions to metastable states in these 8 disaccharides, comparison with the number of transitions $n_{B \rightarrow A}$ (6) actually observed in the 50 ns plain (unbiased) MD simulation of $\mathbf{T}_{\alpha\beta}$ (Table 2.2) suggests that the appropriate value is actually much lower, *i.e.* of the order of 0.01 (the corresponding estimates are reported between parentheses in the Table 2.4). This second approximation leads to estimated forward transition timescales ranging from about 10 ns (50 ns, omitting $\mathbf{T}_{\alpha\beta}$) up to about 2 μs . The latter values are more realistic considering the near absence of such transitions in the plain (unbiased) MD simulations of these disaccharides (except for $\mathbf{T}_{\alpha\beta}$; Section 2.4.1). On the other hand, the backward transition timescales estimated with a transmission coefficient of one are probably more realistic in view of the much lower transition barriers involved. For these disaccharides, the comparison of the free-energy differences and barriers associated with transitions involving the ϕ and ψ dihedral angles does not reveal any clear systematic trend. In particular, for the disaccharides involving an equatorial-equatorial linkage, the metastable state involving a transition around ϕ may be lower (\mathbf{S} , \mathbf{L}), equally high ($\mathbf{T}_{\beta,\beta}$) or higher \mathbf{C} in free energy compared to the one involving a transition around ψ . Experimentally, ultrasonic relaxation experiments suggest a timescale of about 10 ns for the relative rotation of the two pyranose rings in disaccharides^{81,82,190}, which is about one order of magnitude faster than the timescales indicated above. The cause of this discrepancy probably resides in the force-field employed in the present simulations. However, it may also partly be due to inaccuracies in the experimentally suggested timescale (which also relies on a model for its derivation based on the raw absorption spectrum).

| System | $X \rightarrow Y$ | Angle [deg] | $\Delta\tilde{G}_{X \rightarrow Y}$ [kJ mol ⁻¹] | $\Delta G_{X \rightarrow Y}$ [kJ mol ⁻¹] | $\Delta G_{X \rightarrow Y}^\ddagger$ [kJ mol ⁻¹] | $\Delta G_{Y \rightarrow X}^\ddagger$ [kJ mol ⁻¹] | $\tau_{X \rightarrow Y}$ [ps] | $\tau_{Y \rightarrow X}$ [ps] |
|-----------------------|-------------------|------------------|--|---|--|--|-----------------------------------|----------------------------------|
| T_{αα} | - | - | - | - | - | - | - | - |
| T_{αβ} | $B \rightarrow A$ | ψ | 11.7 | 10.1 | 17.4 | 7.3 | $1.1 \cdot 10^2 (1.1 \cdot 10^4)$ | 2.1 |
| T_{ββ} | $D \rightarrow C$ | (ψ) | 15.8 | 16.2 | 23.1 | 6.9 | $1.1 \cdot 10^3 (1.1 \cdot 10^5)$ | 1.8 |
| | $D \rightarrow B$ | (ϕ) | 16.2 | 16.8 | 24.0 | 7.2 | $1.6 \cdot 10^3 (1.6 \cdot 10^5)$ | 1.9 |
| K | $A \rightarrow B$ | ψ | 14.9 | 13.7 | 21.4 | 7.7 | $5.9 \cdot 10^2 (5.9 \cdot 10^4)$ | 2.4 |
| S | $C \rightarrow A$ | ϕ | 15.9 | 15.2 | 24.5 | 9.3 | $2.0 \cdot 10^3 (2.0 \cdot 10^5)$ | 4.6 |
| | $C \rightarrow D$ | ψ | 20.2 | 21.7 | 29.4 | 7.7 | $1.5 \cdot 10^4 (1.5 \cdot 10^6)$ | 2.4 |
| N | $B \rightarrow A$ | ψ | 17.3 | 17.7 | 24.0 | 6.3 | $1.6 \cdot 10^3 (1.6 \cdot 10^5)$ | 1.4 |
| L | $D \rightarrow B$ | ϕ | 16.5 | 15.6 | 23.3 | 7.7 | $1.2 \cdot 10^3 (1.2 \cdot 10^5)$ | 2.4 |
| | $D \rightarrow C$ | ψ | 20.9 | 21.8 | 30.5 | 8.7 | $2.2 \cdot 10^4 (2.2 \cdot 10^6)$ | 3.6 |
| M | $A \rightarrow B$ | ψ | 16.5 | 15.1 | 25.7 | 10.6 | $3.3 \cdot 10^3 (3.3 \cdot 10^5)$ | 7.7 |
| C | $C \rightarrow D$ | ψ | 10.8 | 10.7 | 25.6 | 14.9 | $3.2 \cdot 10^3 (3.2 \cdot 10^5)$ | $4.4 \cdot 10^1$ |
| | $C \rightarrow A$ | ϕ | 16.6 | 16.2 | 21.3 | 5.1 | $5.7 \cdot 10^2 (5.7 \cdot 10^4)$ | $8.0 \cdot 10^{-1}$ |
| I | $B \rightarrow A$ | $\tilde{\omega}$ | 2.1 | 2.5 | 19.4 | 16.9 | $2.6 \cdot 10^2$ | $9.7 \cdot 10^2$ |
| G | $C \rightarrow D$ | $\tilde{\omega}$ | 1.2 | 1.5 | 20.3 | 18.8 | $3.8 \cdot 10^2$ | $2.1 \cdot 10^2$ |
| | $A \rightarrow B$ | $\tilde{\omega}$ | 1.4 | 1.7 | 23.4 | 21.7 | $1.3 \cdot 10^3$ | $6.7 \cdot 10^2$ |
| | $C \rightarrow A$ | ϕ | 9.2 | 8.6 | 23.1 | 14.5 | $1.2 \cdot 10^3 (1.2 \cdot 10^5)$ | $3.7 \cdot 10^1$ |
| | $D \rightarrow B$ | ϕ | 9.4 | 8.8 | 22.2 | 13.4 | $8.1 \cdot 10^2 (8.1 \cdot 10^4)$ | $2.4 \cdot 10^1$ |

Table 2.4: Parameters characterizing the relative free energies and interconversion barriers between the conformational states identified in LEUS simulations of the 11 considered disaccharides in water at 300 K and 1 bar. The reported quantities for each transition ($X \rightarrow Y$, involving a rotation around the indicated glycosidic dihedral angle; for **T_{ββ}**, ϕ and ψ are interchangeable by symmetry) are the state-to-state free-energy difference $\Delta\tilde{G}_{X \rightarrow Y}$ (based on the relative state populations P_X reported in Table 2.1), the minimum-to-minimum free-energy difference $\Delta G_{X \rightarrow Y}$, the (forward and backward) minimum-to-saddle point free-energy barriers $\Delta G_{X \rightarrow Y}^\ddagger$ or $\Delta G_{Y \rightarrow X}^\ddagger$, and the corresponding transition timescales (half-times) $\tau_{X \rightarrow Y}$ or $\tau_{Y \rightarrow X}$. The transition timescales were estimated using transmission coefficients set to one in Eqs. 2.4 (values between parentheses: transmission coefficients set to 0.01). Only pairs of significantly populated conformational states ($P_X, P_Y \geq 0.01\%$; Table 2.1) and transitions along direct pathways (i.e. not involving an intermediate conformational state) are reported. The codes employed for the different disaccharides are provided in the legend of Figure 2.1 and the definitions of the conformational states in Section 2.3.2 and Table 2.1 (see also Figures 2.4 and 2.5).

For the (1 \rightarrow 6)-linked disaccharides (**I** and **G**), the free energy differences between the (most stable and, in the case of **G** metastable) states differing in the $\tilde{\omega}$ values are in the range 1.5–2.5 kJ mol⁻¹. The estimated transition timescales for the corresponding (forward or backward) transitions are of the order of 200 ps to 1.3 ns. Comparison with the corresponding numbers of transitions $n_{X \rightarrow Y}$ (19 for $B \rightarrow A$ in **I** and 135 for $C \rightarrow D$ in **G**) actually observed in the 50 ns plain (unbiased) MD simulation of these two disaccharides (Table 2.2) suggests that the choice of a transmission factor of one is probably a reasonable approximation in this case. In the case of **G**, the pair of metastable states differing in the ϕ values from the pair of stable states are about 9 kJ

mol^{-1} higher in free energy. The forward transition timescales are of the order of 1 ns (or 100 ns assuming a transmission factor of 0.01, which is reasonable considering that these transitions are observed once each in the unbiased 50 ns simulation; Table 2.2).

2.4.7 Hydrogen bonding

The occurrences of intramolecular H-bonds in the different conformational states observed in the LEUS simulations of the 11 considered disaccharides (50 ns sampling phase, after unbiasing) are reported in Table 2.5. Note that these occurrences are given relative to the overall population of the state (which may in some cases be very small; Table 2.1) and that only H-bonds with occurrences larger than 5% (in a given state) are reported. The occurring H-bonds are also indicated in the illustrative structures of Figure 2.7 (whenever possible).

All observed intramolecular H-bonds are formed between the two residues. The complete absence of intraresidue H-bonds is probably related to the rather restrictive geometrical criteria used here to define the presence of an H-bond, namely a hydrogen-acceptor distance shorter than 0.25 nm and donor-hydrogen-acceptor angle larger than 135° . When the angle criterion is relaxed (*e.g.* from 135° to 100°), intraresidue H-bonds become observable⁴¹.

For the $(1 \rightarrow n)$ -linked disaccharides ($n = 2, 3$ or 4), the dominant (significantly populated) states all present occurrences of intramolecular H-bonds. The nature of these H-bonds and the corresponding occurrences are very similar to those reported previously⁵³ (see Figure 1 and Table 2 therein). The three α -linked disaccharides (axial-equatorial linkage; **K**, **N** and **M**) are characterized by two alternative H-bonds in this state (occurrences ranging between about 10 and 40%), involving hydroxyl groups of both residues vicinal to the glycosidic linkage or/and the ring oxygen atom of the non-reducing residue. The three β -linked disaccharides (equatorial-equatorial linkage; **S**, **L** and **C**) are characterized by one possible H-bond in this state (occurrence ranging between about 40 and 80%), involving a hydroxyl group of the reducing residue vicinal to the glycosidic linkage and the ring oxygen atom of the non-reducing residue. In contrast, the $(1 \rightarrow 1)$ - and $(1 \rightarrow 6)$ -linked disaccharides (**T** _{$\alpha\alpha$} , **T** _{$\alpha\beta$} , **T** _{$\beta\beta$} , **I** and **G**) present no intramolecular H-bonding in their dominant (significantly populated) states. In a number of cases (**T** _{$\beta\beta$} , **K** and **M**), intramolecular H-bonds are also observed in metastable states. For **T** _{$\beta\beta$} one alternative state (**B**) involves a $\text{HO}'_2 \rightarrow \text{O}_2$ H-bond (occurrence of about 6-7%), while the corresponding most stable state involves none. For **K** and **M**, alternative states may involve a different H-bonding pattern compared to the most stable state, with occurrences up to about 60%. Note that with one single exception ($\text{HO}_2 \rightarrow \text{O}'_3$ in **M**; occurrence 9.6%), all the observed H-bonds (in the most stable and alternative states) involve a hydrogen donor from the reducing residue and an acceptor from the non-reducing one.

The above observations can be easily rationalized if one assumes that the formation of a solvent-exposed intramolecular H-bond in aqueous environment is an “opportunistic” consequence of the close proximity of two H-bonding groups in a given molecular conformation, rather than a major conformational driving force promoting this proximity⁴¹. Accepting this assumption, the most stable conformation of a disaccharide would be almost exclusively determined by stereo-electronic and steric effects (Section 2.4.3), while intramolecular H-bonding would play a minor role (*i.e.* only slightly modulate the basic conformational preferences of the specific linkage). This interpretation would explain why: (*i*) the $(1 \rightarrow 1)$ -linked trehaloses present no intramolecular H-bonding (no favorably positioned H-bonding groups in the dominant conformation; Figure 2.7);

| System | A [%] | B [%] | C [%] | D [%] | H-bonds solute-solvent |
|-----------------------|--|--|---|---|---------------------------|
| T_{αα} | × | - | - | - | 15.91 |
| T_{αβ} | - | × | - | - | 15.77 |
| T_{ββ} | - | HO' ₂ → O ₂ (5.7) | - | × | 15.41 |
| K | HO' ₁ → O ₂ (33.5) HO' ₃ → O ₅ (15.1) | HO' ₃ → O ₂ (33.6) HO' ₁ → O ₅ (21.2) | - | - | 15.26 |
| S | - | - | HO' ₁ → O ₅ (41.6) | - | 15.39 |
| N | - | HO' ₄ → O ₂ (39.4) HO' ₂ → O ₅ (17.7) | - | - | 15.64 |
| L | - | - | - | HO' ₄ → O ₅ (43.9) | 15.44 |
| M | HO' ₃ → O ₂ (29.0) HO ₂ → O' ₃ (9.6) | HO' ₃ → O ₅ (60.4) | HO' ₃ → O ₅ (6.4) | - | 16.07 |
| C | - | - | HO' ₃ → O ₅ (77.5) | - | 15.18 |
| I | × | × | - | - | 16.07 |
| G | - | - | × | × | 16.19 |

Table 2.5: Occurrences of intramolecular H-bonds in the different conformational states observed in LEUS simulations of the 11 considered disaccharides in water at 300 K and 1 bar. The reported data indicate the identity and occurrence (relative to the overall population P_X of the state, which may in some cases be very small; Table 2.1) of a given H-bond. The average number of solute-solvent H-bonds during the simulation (total number, after unbiasing) is also reported. The presence of an H-bond was defined by a maximal hydrogen-oxygen distance of 0.25 nm and a minimal oxygen-hydrogen-oxygen angle of 135°. Only H-bonds with occurrences larger than 5% are reported. Entries printed in boldface (or “×”) correspond to the most stable state(s) of the disaccharide ($P_X > 30\%$; Table 2.1). The codes employed for the different disaccharides are provided in the legend of Figure 2.1 and the definitions of the conformational states in Section 2.3.2 and Table 2.1 (see also Figures 2.4 and 2.5).

(ii) the (1 → 6)-linked disaccharides present no intramolecular H-bonding (high flexibility of the three-bond glycosidic linkage and absence of significant driving force for H-bonding); (iii) some metastable states presenting a significant extent of intramolecular H-bonding are nevertheless not significantly populated. This view is also supported by the approximate topological symmetries observed in the free-energy maps of the disaccharides (Section 2.4.3), which suggest that the detailed positioning of the hydroxyl groups, hydroxymethyl groups, and ring oxygen atoms (*i.e.* of the potentially H-bonding groups) relative to the glycosidic linkage only exerts a minor influence on the conformational properties of the glucose-based disaccharides. Note, however, even if the formation of intramolecular H-bonds does not represent a major conformational driving force, it may still affect the physico-chemical properties of a sugar. The reason is that this formation is expected to reduce the propensity of the molecule to interact with the solvent molecules (hydrophilicity), with itself (at finite concentration) or with other potentially H-bonding solute molecules (in the presence of other solutes). As a possible indicator of hydrophilicity, the average number of solute-solvent H-bonds is also reported in Table 2.5. The disaccharides lacking

intramolecular H-bonds in their most stable state indeed tend to form more H-bonds with the solvent. However, the trends are not entirely systematic, suggesting that this indicator might be somewhat too simplistic. For example, $\mathbf{T}_{\beta\beta}$ appears to be poorly hydrated in spite of the lack of intramolecular H-bonds, while \mathbf{M} appears to be strongly hydrated in spite of the high occurrence of intramolecular H-bonds.

The absence of intramolecular H-bonds in \mathbf{I} and the presence of interresidue H-bonds between the hydroxyl groups at positions 2 and 3' in \mathbf{M} were previously suggested by a combined NMR/MD study¹⁹¹. The H-bonds present in \mathbf{K} and \mathbf{S} were also observed in the corresponding global minima found in MM3 calculations¹¹⁶. Finally, the H-bond observed for \mathbf{C} was reported in a number of previous experimental and theoretical studies^{97, 109, 110, 115, 122, 123}. In the present simulations, this specific H-bond is the one with the highest occurrence (nearly 80%). It is known to persist in the polysaccharide chains (cellulose) formed by this disaccharide in the crystalline state^{8, 192}, and to be an important determinant of the physico-chemical and mechanical properties of this material (stability, rigidity and insolubility in water¹⁹³).

2.4.8 Configurational entropy

The build-up curves of the QH estimates S_{qm}^h for the solute configurational entropies observed in the LEUS simulations of the 11 considered disaccharides (50 ns sampling phase, after unbiasing) are displayed as a function of the sampling time in Figure 2.8. The corresponding estimated solute configurational entropies (including classical correction terms⁶⁵ for mode anharmonicities ΔS_{cl}^{ah} and pairwise supralinear correlations ΔS_{cl}^{pc}) are reported in Table 2.6. These entropy values are directly comparable because all molecules contain an identical number of atoms (31).

The build-up curves of S_{qm}^h show a pronounced stepwise evolution in the first 10 ns of the simulations, which suggests that even in the presence of the LEUS sampling enhancement, simulations of at least this length are required for a disaccharide in water to sample a significant fraction of its accessible conformational space. After this time period, the curves are characterized by long plateaus and convergence appears to be essentially reached after about 30-40 ns. The values of S_{qm}^h calculated over the entire 50 ns simulations agree reasonably well with the corresponding (less well converged) values reported previously for the non-reducing disaccharides using plain (unbiased) simulations of identical durations^{53, 54} (with deviations of at most $25 \text{ J K}^{-1} \text{ mol}^{-1}$ *i.e.* about 3.5 %).

The correction terms ΔS_{cl}^{ah} for anharmonicities of the QH modes are negative and relatively small (about 2-4 % of S_{qm}^h), which agrees with previous results in the context of small solute molecules^{65, 194, 195} (but does not necessarily apply to larger macromolecules). These corrections are of similar magnitudes for all compounds. The correction terms ΔS_{cl}^{pc} for pairwise correlations between the QH modes are also negative, but of significantly larger magnitudes (about 40-70 % of S_{qm}^h), which also agrees with previous results in the context of β -peptides in methanol⁶⁵. Because this correction is very large, omitting it would lead to a significant overestimation of the configurational entropy. Note that the values of ΔS_{cl}^{ah} and ΔS_{cl}^{pc} calculated here are significantly larger than the corresponding values reported previously for the non-reducing disaccharides using plain (unbiased) simulations of identical durations^{53, 54}. This is due in part to the improved convergence afforded by the LEUS sampling enhancement, but also to a mistake in the previously reported data (the values reported previously⁵⁴ for ΔS_{cl}^{ah} and ΔS_{cl}^{pc} mistakenly omitted the multiplicative factor

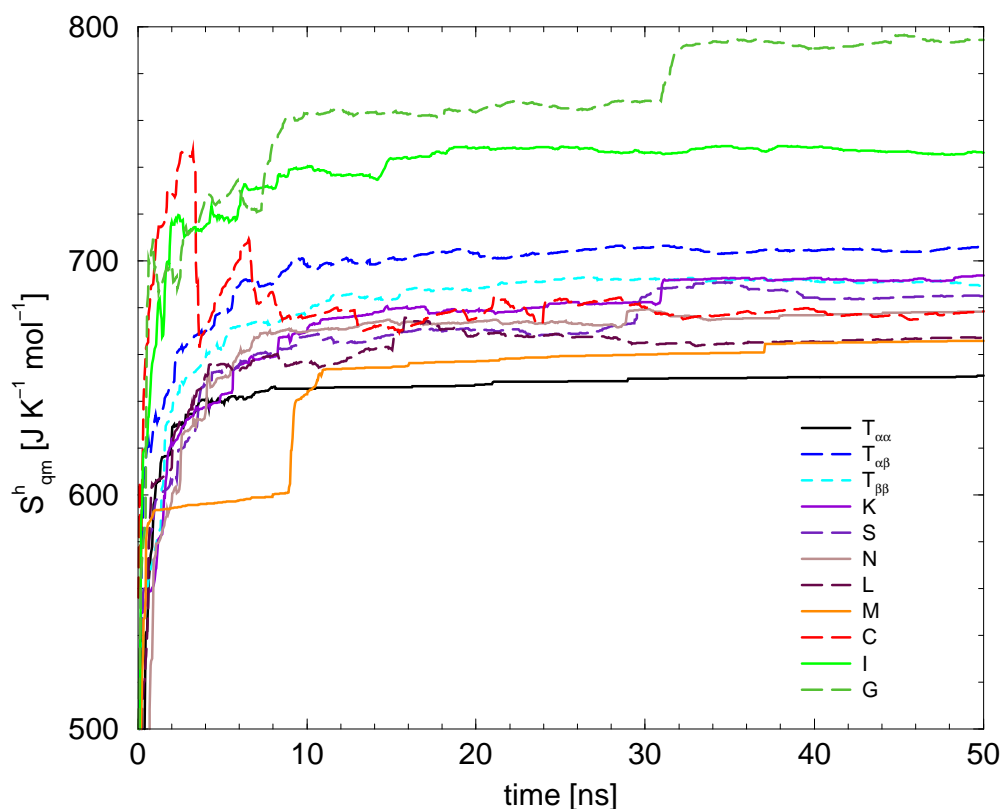


Figure 2.8: Build-up curves of the QH estimates S_{qm}^h for the solute configurational entropies as obtained from LEUS simulations of the 11 considered disaccharides in water at 300 K and 1 bar, displayed as a function of the sampling time. The codes employed for the different disaccharides are provided in the legend of Figure 2.1

$k_B = 8.314 \text{ J K}^{-1} \text{ mol}^{-1}$). Higher-order (beyond pairwise) correlations are expected to further decrease the entropy, but are increasingly difficult to estimate (requirement for more extensive sampling, large computational and memory requirement involved in their evaluation).

The corrected entropy values S_{sol} range between 182.7 and 430.2 $\text{J mol}^{-1}\text{K}^{-1}$. This represents a range of about 22 to 52 in units of k_B . This range is not unreasonable, considering the number of unconstrained internal coordinates (and excluding rigid-body translation and rotation) within the disaccharides (55 coordinates). For the (1 \rightarrow n)-linked disaccharides ($n = 2, 3, 4$) most of the corrected entropy values S_{sol} fall within a narrow range of about 250-255 $\text{JK}^{-1}\text{mol}^{-1}$, except for **K** (280.1) and **C** (208.8). For the (1 \rightarrow 1)-linked disaccharides (trehaloses) the entropy values S_{sol} for **T $_{\alpha\beta}$** (268.9) marginally higher, while the values for **T $_{\alpha\alpha}$** (182.7) and **T $_{\beta\beta}$** (323.8) are significantly lower and higher, respectively. Finally, the entropy values S_{sol} for the $\alpha(1 \rightarrow 6)$ -linked disaccharides (**I** and **G**) are significantly larger than the corresponding values for all other disaccharides. This observation is intuitively justified by the presence of an additional dihedral angle ($\tilde{\omega}$) within the glycosidic linkage. Note that a systematically smaller configurational entropy for the α -linked (axial-equatorial) compared to the β -linked (equatorial-equatorial) disaccharides (as reported previously^{53,54} and suggested by qualitative considerations based on QM or MM potential energy maps for the two types of linkages) is no longer observed. Based on the present

| System | S_{qm}^h [J K ⁻¹ mol ⁻¹] | ΔS_{cl}^{ah} [J K ⁻¹ mol ⁻¹] | ΔS_{cl}^{pc} [J K ⁻¹ mol ⁻¹] | S_{sol} [J K ⁻¹ mol ⁻¹] | $-T\Delta S_{sol}$ [kJ mol ⁻¹] |
|-----------------------|--|--|--|---|---|
| T_{αα} | 650.9 | -19.6 | -448.5 | 182.7 | 0.0 |
| T_{αβ} | 706.1 | -15.6 | -421.7 | 268.9 | 25.9 |
| T_{ββ} | 689.3 | -19.8 | -345.7 | 323.8 | 42.3 |
| K | 693.8 | -18.0 | -395.7 | 280.1 | 29.2 |
| S | 684.9 | -22.5 | -407.4 | 255.0 | 21.7 |
| N | 678.3 | -22.4 | -400.6 | 255.4 | 21.8 |
| L | 667.4 | -22.0 | -394.4 | 251.0 | 20.5 |
| M | 665.8 | -26.7 | -388.6 | 250.8 | 20.4 |
| C | 678.4 | -29.3 | -440.4 | 208.8 | 7.8 |
| I | 746.4 | -14.8 | -319.0 | 412.6 | 69.0 |
| G | 794.4 | -19.3 | -344.9 | 430.2 | 74.3 |

Table 2.6: Estimated solute configurational entropies and corresponding correction terms as obtained from LEUS simulations of the 11 considered disaccharides in water at 300 K and 1 bar. The reported quantities are the QH estimates for the solute configurational entropies S_{qm}^h (evaluated using the quantum-mechanical formula for the entropy of a harmonic oscillator; Figure 2.8) and the (classically derived) corrections for mode anharmonicities (ΔS_{cl}^{ah}) and pairwise supralinear correlations (ΔS_{cl}^{pc}), leading to the corrected values S_{sol} (Eqs. 2.7). The quantity $-T\Delta S_{sol}$, where T is the temperature (300 K) and ΔS_{sol} the entropy difference between **T_{αα}** and the specific disaccharide, is also reported as an indication of the corresponding free-energy contribution (relative to **T_{αα}**). The codes employed for the different disaccharides are provided in the legend of Figure 2.1.

results, such a trend only applies to the pair **T_{αβ}** vs. **T_{ββ}** and **I** vs. **G**, the opposite being observed for **K** vs. **S**, **N** vs. **L** and **M** vs. **C**. However, it should be stressed that the presence of additional metastable minima on the Ramachandran free-energy map for β -linkages compared to α -linkages (Figure 2.4 and 2.5) does not automatically imply a higher conformational entropy, in view of the high relative free-energies associated with these alternative states (Table 2.4). In the present situation, it is probably much more the breadth of the most stable state that is a determinant factor for the magnitude of this entropy.

The trends in configurational entropies suggested by the present simulation study are thus **G** > **I** >> **T_{ββ}** >> **K** > **T_{αβ}** > **N** ~ **S** > **L** ~ **M** >> **C** >> **T_{αα}**. These results are quantitatively in line with the results of a chromatographic study comparing estimated conformational entropies for **T_{ββ}**, **M**, **C**, **I** and **G** in aqueous solution¹⁹⁶, which suggested a ranking **G** > **T_{ββ}** > **I** > **C** > **M** in terms of decreasing entropy (the present results show, however, inversions between **T_{ββ}** and **I**, as well as between **C** and **M** in this series).

2.5 Conclusions

The goal of the present work was to revisit (and extend) the conclusions of a previous study^{53,54} (Pereira *et al.*, 2006) concerning the conformational and dynamical properties of glucose-based disaccharides in water, as probed by explicit-solvent MD simulations using the GROMOS 45A4 force-field⁵⁵⁻⁵⁹. To overcome the lack of sampling in the glycosidic dihedral angles ϕ and ψ on the 50 ns timescale^{53,54}, a sampling enhancement method⁶⁰ (LEUS) was applied in the conformational subspace defined by these degrees of freedom. The study was further extended to all possible glucose-based disaccharides (non-reducing as well as reducing; β -anomer in the latter case), *i.e.* it includes the compounds (Figure 2.1) trehalose ($\mathbf{T}_{\alpha\alpha}$), isotrehalose ($\mathbf{T}_{\alpha\beta}$), neotrehalose ($\mathbf{T}_{\beta\beta}$), kojibiose (\mathbf{K}), sophorose (\mathbf{S}), nigerose (\mathbf{N}), laminarabiose (\mathbf{L}), maltose (\mathbf{M}), cellobiose (\mathbf{C}), isomaltose (\mathbf{I}) and gentiobiose (\mathbf{G}).

The results demonstrate that the LEUS sampling enhancement promotes (within the same 50 ns sampling time) an extensive (*i.e.* nearly exhaustive in terms of the sterically allowed regions) coverage of the Ramachandran (ϕ, ψ) maps and a statistically meaningful number of transitions (hundred or more, compared to a few or none in the free MD simulations^{53,54}) between the different (meta)stable conformational states for all disaccharides considered.

In terms of average properties and agreement with experimental data, the results of the LEUS simulations do not differ dramatically from those obtained previously for the non-reducing disaccharides^{53,54}. The reason is that all disaccharides considered are characterized either by a deep (*i.e.* overwhelmingly populated) free-energy basin in the neighborhood of the experimental structure ((1 \rightarrow n)-linked disaccharides with $n = 1, 2, 3$ or 4) or by two dominant basins with a low interconversion barrier (*i.e.* corresponding to transitions amenable to appropriate sampling on the 50 ns timescale; (1 \rightarrow 6)-linked disaccharides). However, additional information can be obtained from the present LEUS simulations concerning the properties of the (meta)stable conformational states and their interconversion dynamics. Although these metastable states are not relevant in the specific case of disaccharides (due to their negligible populations), they may become important in the context more complex oligo- and polysaccharides (either at equilibrium or along conformational transitions).

Based on the results of these simulations, the 11 glucose-based disaccharides can be roughly classified into the four following families :

1. The only disaccharide with an axial-axial linkage ($\mathbf{T}_{\alpha\alpha}$) presents no (significant) metastable state, no intramolecular H-bonds and the lowest configurational entropy.
2. The four $\alpha(1 \rightarrow n)$ -linked disaccharides ($n = 1, 2, 3$ or 4) with an axial-equatorial linkage ($\mathbf{T}_{\alpha\beta}$, \mathbf{K} , \mathbf{N} and \mathbf{M}) present a stable conformational state involving two alternative intramolecular H-bonds (occurrences of about 10-40%; none for $\mathbf{T}_{\alpha\beta}$), a single (significant) metastable state (differing from the most stable state by the value of ψ and with a relative free energy of the order of 10-18 kJ mol⁻¹), and intermediate configurational entropies (noticeably higher for $\mathbf{T}_{\alpha\beta}$ and \mathbf{K}).
3. The four $\beta(1 \rightarrow n)$ -linked disaccharides ($n = 1, 2, 3$ or 4) with an equatorial-equatorial linkage ($\mathbf{T}_{\beta\beta}$, \mathbf{S} , \mathbf{L} and \mathbf{C}) present a stable conformational state involving a single intramolecular H-bond (occurrence 40-80%; none for $\mathbf{T}_{\beta\beta}$), two (significant) metastable states (differing from the most stable state by the value of either ϕ or ψ and with relative free energies of the

order of 11-22 kJ mol⁻¹), and intermediate configurational entropies (noticeably lower for **C**).

4. The two (1 → 6)-linked disaccharides (**I** and **G**) are characterized by the presence of an extra glycosidic dihedral angle $\tilde{\omega}$ within the linkage and present two dominant (*i.e.* significantly populated) conformational states (differing by the value of $\tilde{\omega}$ and with a free energy difference of the order of 1.5-2.5 kJ mol⁻¹) involving no intramolecular H-bonds, no (**I**) or a pair of (**G**) metastable conformational states (in the latter case, differing from the most stable state pair by the value of ϕ and with a relative free energy of the order of 9 kJ mol⁻¹), and the highest configurational entropies (noticeably higher for **G**).

This classification clearly correlates with approximate topological symmetries (Figure 2.6). Considering the 8 (1 → n)-linked disaccharides ($n = 1, 2, 3$ or 4; excluding **T_{αα}**), these symmetries disregard the detailed positioning of the exocyclic hydroxyl and hydroxymethyl groups, as well as of the ring oxygen atoms, relative to the glycosidic linkage (except for the *exo*-anomeric influence of a ring oxygen atom next to the linkage). The three other disaccharides (**T_{αα}**: only axial-axial linkage; **I** and **G**: only ones with a three-bond glycosidic linkage) are obviously special and not involved in these symmetries. The existence of these approximate symmetries suggests that the conformational properties of the 11 glucose-based disaccharides are determined by four main conformational driving forces : (i) conformational preferences of the residue rings¹⁹⁷⁻²⁰² (leading here to a unique and relatively rigid ⁴C₁ chair conformation); (ii) *exo*-anomeric effect^{1, 2, 158-165, 203} (stereoelectronic bias of the ϕ values, also acting on ψ for the non-reducing trehaloses); (iii) steric constraints (the latter considering the relative positioning of the rings and omitting the details of their local substitution patterns; leading to forbidden regions on the Ramachandran map and to a more restrictive influence of the *exo*-anomeric effect for α -linkages as opposed to β -linkages); (iv) presence of a third glycosidic dihedral angle in $\alpha(1 \rightarrow 6)$ -linked disaccharides (leading to an enhanced flexibility). The observed approximate symmetries also suggest that intramolecular H-bonding and specific solvation effects play secondary roles in determining the conformational preferences of these disaccharides (*i.e.* only slightly modulate the basic conformational preferences of the specific linkage). Note that the above considerations remain to be generalized (work in progress) to the case of residues presenting alternative chair conformations (*e.g.* D-idose^{41, 204-206}) and one or more axial (non-glycosidic) hydroxyl groups (*i.e.* to disaccharides involving other residues than D-glucopyranose).

The above observations, along with the result of the H-bond analysis, are in line with the hypothesis that the formation of a solvent-exposed intramolecular H-bond in aqueous environment is an “opportunistic” consequence of the close proximity of two H-bonding groups in a given molecular conformation, rather than a major conformational driving force promoting this proximity⁴¹. The reasoning underlying this hypothesis is that since water is a protic solvent of high dielectric permittivity, it dramatically weakens the strength of intramolecular electrostatic interactions (such as H-bonds) and at the same time, efficiently competes against other solute groups for the formation of H-bonds with a given solute group. Note that the above statement concerning the presumably weak conformational driving force associated with H-bonding⁴¹ pertains to small molecules in an aqueous environment. It may not be applicable to larger system (*e.g.* due to H-bonding cooperativity effects^{178, 207-217} in extended chains or reduced local solvation in folded chains) and to other environments (*e.g.* crystals, fibers, solutions with non-polar solvents, or vacuum). In addition, this hypothesis does not imply that intramolecular H-bonding has no

effect on the physico-chemical properties of a specific sugar, because many of these properties are actually defined by a change of environment (relative to the bulk aqueous environment at high dilution). For example, when comparing a closely related series of compounds (such as the present glucose-based disaccharides; see also previous work on monosaccharides⁴¹), the absence of high-occurrence intramolecular H-bonds (compensated by a more extensive interaction with the solvent) may lead to a higher hydrophilicity^{218,219} or, equivalently, a lower apparent hydrophobicity (relative affinity of the compound for less polar environments), a higher propensity of the compound to self-aggregate^{220–226} (*e.g.* tendency to cluster, higher glass transition temperature, more limited solubility) or interact with other polar solutes, and a slowing down of the dynamics in aqueous solution¹³⁷ (*e.g.* lower transport coefficients, higher viscosity).

The conformational peculiarities of the different linkage types reported above clearly correlate (at least to some extent) with the biological functions of these linkages (within di-, oligo- or polysaccharides involving them). For example, the disaccharide trehalose ($\mathbf{T}_{\alpha\alpha}$) is prominent for its role in the preservation of biostructures (proteins, membranes) under stressful environmental conditions^{28–31,50}. Different possible reasons have been put forward to explain the outstanding bioprotective ability of trehalose (over other disaccharides), including a high glass-transition temperature²²⁷, a large hydration radius²²⁸, a reduction of the water dynamics¹³⁷, or peculiarities in the phase diagram of trehalose-water mixtures. In particular, isotrehalose $\mathbf{T}_{\alpha\beta}$ and neotrehalose $\mathbf{T}_{\beta\beta}$ do not evidence the same bioprotective abilities²²⁹. In the present work, trehalose appears special in two respects: (i) it is the least flexible disaccharide in the series (due to the influence of the *exo*-anomeric effect, in its most restrictive form for α -linkages, on its two glycosidic dihedral angles); (ii) it presents a complete lack of intramolecular H-bonds (suggesting a high hydrophilicity). The low flexibility is expected to enhance the affinity of the molecule for biostructures²³⁰ (more limited entropy loss upon binding according to the sugar-like mechanism). The high hydrophilicity is expected to enhance the propensity to clustering and glass formation (self-association). Thus, the specific features of the diaxial linkage within this molecule might be at the origin of its bioprotective properties. On the other hand of the spectrum, the (1→6)-linkage is known to serve as a branching point in many polysaccharides (*e.g.* schizophyllan²³¹, amylopectin^{232,233}). In the present work, the (1→6)-linked disaccharides (\mathbf{I} and \mathbf{G}) appear special in two respects: (i) they are the most flexible disaccharide in the series (due to the presence of an extra glycosidic dihedral angle); (ii) they present a complete lack of intramolecular H-bonds (suggesting a high hydrophilicity). These features may explain why nature has selected this particular linkage for chain branching points where flexibility is required for function.

From an even more general perspective, the finding of a single (significantly populated) conformation at the disaccharide level for all (1 → *n*)-linkages (*n* = 1, 2, 3 or 4) in aqueous solution, with all metastable states presenting a high relative free energy (10–22 kJ mol⁻¹) may also be of relevance to the specific functions of carbohydrates in nature. For illustrative purposes, a comparison of the Ramachandran free-energy map for \mathbf{C} (Figure 2.4), with a corresponding free-energy map for a prototypical peptide linkage (blocked dialanine peptide) in water (calculated using the same force-field and a similar LEUS simulation approach; H. Hansen, unpublished results) is shown in Figure 2.9. The latter map is significantly “flatter” and presents three metastable states with relative free-energies of only 0.3, 1.4 and 6.1 kJ mol⁻¹. Even if the presence of longer sidechains on the amino-acid residues will affect the number, breadths and relative free energies of the different states, the peptide linkage remains generally speaking considerably more flexible than the glycosidic one. This difference suggests that oligo- and polysaccharides are characterized by a stronger

encoding (compared to oligo- and polypeptides) of their conformation into their sequence *via* the local conformational preferences of the successive glycosidic linkages. However, only oligo- and polypeptides are flexible enough to achieve folding, *i.e.* the formation of a unique and conformationally restricted structure (given a carefully engineered residue sequence). Polysaccharides, in contrast, typically achieve conformationally more restricted structures by lateral chain association. In the context of polypeptide folding, the encoding of the structure into the sequence occurs predominantly *via* the numerous non-bonded interactions between backbone and sidechain atoms (that are generally non-local along the sequence).

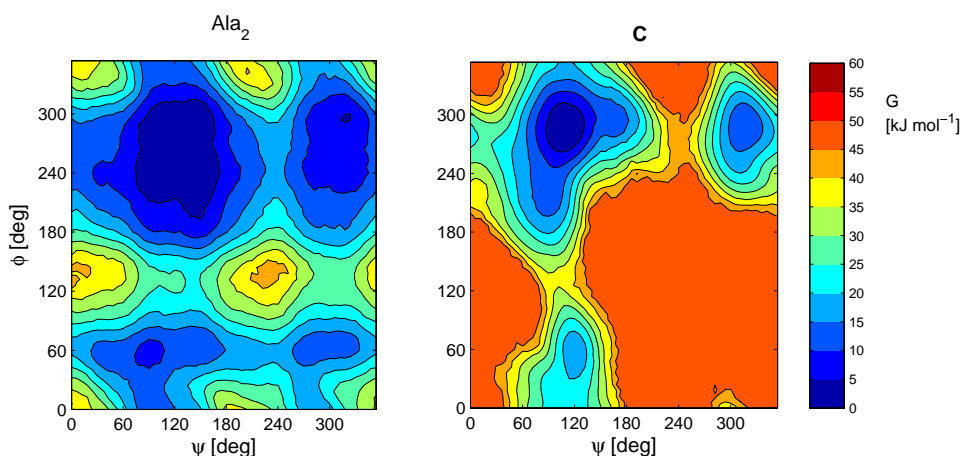


Figure 2.9: Comparison of the Ramachandran free-energy map of a blocked *D*-alanine peptide (Ala_2 ; left; $\text{C}_7\text{H}_{10}\text{N}_2\text{O}_2$) and of cellobiose (**C**; right; Figure 2.4) as obtained from LEUS simulations of the two compounds (using the same force-field; 53A6¹⁴¹, equivalent to 45A4^{55–59} for **C**). The LEUS parameters employed for Ala_2 are : $N_l = 2$, $N_g = 32$, $\sigma = 360^\circ / (2N_g) = 5.625^\circ$, $c = 2 \cdot 10^{-4} \text{ kJ mol}^{-1} \text{ deg}^2$, $t_{LE} = 15 \text{ ns}$, $t_{US} = 50 \text{ ns}$. The definition of the ϕ , ψ dihedral angles for Ala_2 are : $\phi_{\text{C}_O\text{-N-C}_\alpha\text{-C}_O}$ and $\psi_{\text{N-C}_\alpha\text{-C}_O\text{-N}}$

These differences may explain why polypeptides are advantageous for the design of globular structures (proteins) where the (evolutionary) fine-tuning of local residue arrangements is essential, *i.e.* for the design of the cell machinery (*e.g.* enzymes). In contrast, polysaccharides are more amenable to the design of regular structures (those compatible with the local linkage preferences), *i.e.* for the design of elements important for the cell mechanical support (*e.g.* cellulose, polyuronates, capsular polysaccharides), as well as for the conformational encoding of signaling information, *i.e.* for processes where conformational rigidity may enhance the precision of the sequence recognition (*e.g.* glycolipids and glycoproteins at the cell surface). This view is, of course, somewhat reductionistic and many other aspects certainly come into play. For example, oligo- and polypeptides are still associated with the bulk of cell signaling (*e.g.* hormones, signaling proteins), while the role of carbohydrates in the energetic metabolism is probably related to their specific chemical rather than conformational properties.

As a final note concerning this comparison between the peptidic and glycosidic linkages, it is often stated (Chapter 1) that the experimental characterization of oligo- and polysaccharides is

particularly difficult (compared *e.g.* to polypeptides) because of the high conformational heterogeneity (disorder) of these biopolymers. This statement is certainly correct when referring to a comparison between folded polypeptide chains and free polysaccharide chains in solution, and to a large extent also, associated polysaccharide chains in crystals or fibers. However, the present comparison suggests that for the same number of linkages (and in the absence of folding), an oligosaccharide is considerably more ordered compared to an oligopeptide (this statement is supported by the simple comparison of the relative persistence lengths of the two types of chains in solution). Another common statement is that the folding of polysaccharide chains is prevented by the high conformational entropy (disorder) in solution. This second statement is almost certainly incorrect. The present results suggest that it is the conformational rigidity of polysaccharide chains that prevents them from folding (for enthalpic reasons), rather than their presumed flexibility (for entropic reasons). In the absence of folding, both types of biomolecules are capable of adopting labile and short-ranged secondary structures in solution. These are likely to be less diverse and more stable in the case of saccharides, owing to their more limited linkage flexibility (Figure 2.9), although the dominant free-energy basin is still generally compatible with multiple types of regular helical structures. And in both cases, these local and labile secondary structure motives may become extended and permanent in the presence of tertiary interactions, *i.e.* upon folding (polypeptide), lateral association with other chains (polysaccharide; crystals, fibers, junction zones in gels) or binding to other biomolecules (*e.g.* protein-protein or protein-carbohydrate association).

Along with previous work^{41,53–55,60}, the present results also permit to suggest a complete spectrum of approximate transition timescales for carbohydrates up to the disaccharide level, namely : ~ 30 ps (hydroxyl groups^{41,53}), ~ 1 ns (free lactol group^{41,53}, free hydroxymethyl groups^{41,53}, glycosidic dihedral angle $\tilde{\omega}$ in (1 \rightarrow 6)-linked disaccharides⁵³), ~ 10 ns to $2 \mu\text{s}$ (ring conformation⁶⁰, glycosidic dihedral angles ϕ and ψ). For completeness, mutarotation (ring opening and randomization of the anomeric configuration for reducing residues) is associated with timescales^{82,234–240} of the order of 10^3 s.

These results complement those previously reported by Lins & Hünenberger⁵⁵, Pereira *et al.*^{53,54}, Kräutler *et al.*⁴¹ and Hansen & Hünenberger⁶⁰ in providing a detailed characterization of the (thermodynamical and dynamical) properties of mono- and disaccharides within the GRO-MOS 45A4 force-field^{55–59}. Good agreement with available experimental and theoretical data also represents a validation for the latter force-field. In turn, these simulations provide insight into a number of properties of mono- and disaccharides that are not directly (or not unambiguously) accessible by experimental measurements.

Acknowledgments

The authors would like to thank Cristina Pereira for her valuable help and support during the initial stages of this work.

2.6 References

- [1] Lehmann, J. Carbohydrates structure and biology. Georg Thieme Verlag, Stuttgart, Germany. (1998)
- [2] Rao, V.S.R., Qasba, P.K., Balaji, P.V. & Chandrasekaran, R. Conformation of carbohydrates. Harwood Academic Publishers, Amsterdam, The Netherlands. (1998)
- [3] Dwek, R.A. *Chem. Rev.* **96** 683-720 (1996)
- [4] Gabius, H.-J., André, S., Kaltner, H. & Siebert, H.-C. *Biochim. Biophys. Acta* **1572** 165-177 (2002)
- [5] Lowe, J.B. & Marth, J.D. *Annu. Rev. Biochem.* **72** 643-691 (2003)
- [6] Gabius, H.-J., Siebert, H.-C., André, S., Jiménez-Barbero, J. & Rüdiger, H. *ChemBioChem* **5** 740-764 (2004)
- [7] Gabius, H.-J. *Crit. Rev. Immunol.* **26** 43-79 (2006)
- [8] Gardner, K.H. & Blackwell, J. *Biopolymers* **13** 1975-2001 (1974)
- [9] Ovodov, Y.S. *Pure. Appl. Chem.* **42** 351-369 (1974)
- [10] Rees, D.A. & Welsh, E.J. *Angew. Chem. Int. Ed.* **16** 214-224 (1977)
- [11] Berg, J.M., Tymoczko, J.L. & Stryer, L. Biochemistry Volume . Edition 5th. W.H. Freeman, Co. Ltd. (2007)
- [12] Nelson, D.L. & Cox, M.M. Lehninger Principles of Biochemistry Volume . Edition 5th. Palgrave Macmillan (2008)
- [13] Peric, L., Pereira, C.S., Perez, S. & Hünenberger, P.H. *Molecular Simulation* **34** 421-446 (2008)
- [14] Di Pippo, Bohn, A., Congestri, R., De Philippis, R., Albertano, *Biofouling.* **6** 495-504. (2009)
- [15] Ovodov, Y.S. *Russian J. Bioorg. Chem.* **35** 269-284 (2009)
- [16] Ridley, B.L., O'Neill, M.A. Mohnen, *Phytochemistry* **57** 929-967 (2001)
- [17] Peat, S., Bourne, E.J. & Barker, S.A. *Nature* **161** 127- (1948)
- [18] Flatt, J.P. *Journal of lipid research* **11** 131- (1970)
- [19] Dwek, R.A., Edge, C.J., Harvey, D.J. & Wormald, M.R. *Annu. Rev. Biochem.* **62** 65-100 (1993)
- [20] Sears, P. & Wong, C.H. *Proc. Natl. Acad. Sci.* **93** 12086-12093 (1996)
- [21] Davis, A.P. & Wareham, R.S. *Angew. Chem. Int. Ed.* **38** 2978-2996 (1999)

- [22] Bush, C.A., Martin-Pastor, M. & Imberty, A. *Annu. Rev. Biophys. Biomol. Struct.* **28** 269-293 (1999)
- [23] Imberty, A. & Pérez, S. *Chem. Rev.* **100** 4567-4588 (2000)
- [24] Rüdiger, H., Siebert, H.-C., Solís, D., Jiménez-Barbero, J., Romero, A., von der Lieth, C.-W., Diaz-Mauriño, T. & Gabius, H.-J. *Curr. Med. Chem.* **7** 389-416 (2000)
- [25] Hricovini, M. *Curr. Med. Chem.* **11** 2565-2583 (2004)
- [26] Dyekjaer, J.D. & Woods, R.J. *ACS Symposium Series* **930** 203-219 (2006)
- [27] Garrett, B.C., Schenter G.K., Morita, *Chem. Rev.* **106** 1355-1374 (2006)
- [28] Crowe, J.H., Crowe, L.M. & Chapman, D. *Science* **223** 701-703 (1984)
- [29] Crowe, J.H., Carpenter, J.F. & Crowe, L.M. *Annu. Rev. Physiol.* **60** 73-103 (1998)
- [30] Tunnacliffe, A. & Lapinski, J. *Phil. Trans. R. Soc. London B* **358** 1755-1771 (2003)
- [31] Crowe, J.H., Crowe, L.M., Wolkers, W.F., Oliver, A.E., Ma, X., Auh, J.-H., Tang, M., Zhu, S., Norris, J. & Tablin, F. *Integr. Comp. Biol.* **45** 810-820 (2005)
- [32] Draget, K.I., Skjåk-Braek, G. & Smidsrød, O. *Int. J. Biol. Macromol.* **21** 47-55 (1997)
- [33] Ramesh, H.P. & Tharanathan, R.N. *Crit. Rev. Biotech.* **23** 149-173 (2003)
- [34] Ladmiral, V., Melia, E. & Haddleton, D.M. *Eur. Polym. J.* **40** 431-449 (2004)
- [35] Zimmermann, H., Ehrhart, F., Zimmermann, D., Müller, K., Katsen-Globa, A., Behringer, M., Feilen, P.J., Gessner, P., Zimmermann, G., Shirley, S.G., Weber, M.M., Metze, J. & Zimmermann, U. *Appl. Phys. A* **89** 909-922 (2007)
- [36] Liu, L., Fishman, M.L. & Hicks, K.B. *Cellulose* **14** 15-24 (2007)
- [37] Rinaudo, M. *Polym. Int.* **57** 397-430 (2008)
- [38] Quignard, F., Valentin, R. & Di Renzo, F. *New J. Chem.* **32** 1300-1310 (2008)
- [39] Tokstoguzov, V. *Origins of Life and Evolution of the Biosphere* **34** 571-597 (2004)
- [40] Stern, R. & Jedrzejak, M.J. *Chem. Rev.* **108** 5061-5085 (2008)
- [41] Kräutler, V., Müller, M. & Hünenberger, P.H. *Carbohydr. Res.* **342** 2097-2124 (2007)
- [42] Laine, R.A. *Glycobiology* **4** 759-767 (1994)
- [43] Meyer, B. Conformational aspects of oligosaccharides. In: *Topics in current chemistry. Volume 154*. Thiem, J., Ed. Springer-Verlag, Berlin, Germany. pp 141-208 (1990)
- [44] Engelsen, S.B., Cros, S., Mackie, W. & Pérez, S. *Biopolymers* **39** 417-433 (1996)

- [45] Pérez, S., Kouwijzer, M., Mazeau, K. & Engelsens, S.B. *J. Mol. Graphics* **14** 307-321 (1996)
- [46] Woods, R.J. *Glycoconjugate J.* **15** 209-216 (1998)
- [47] Wormald, M.R., Petrescu, A.J., Pao, Y.-L., Glithero, A., Elliott, T., *Chem. Rev.* **102** 371-386 (2002)
- [48] Gerbst, A.G., Grachev, A.A., Shashkov, A.S. & Nifantiev, N.E. *Russian J. Bioorg. Chem.* **33** 24-37 (2007)
- [49] Sanz, M.L., Sanz, J. & Martinez-Castro, I. *J. Chromatogr.A.* **1059** 143-148 (2004)
- [50] Crowe, J.H., Hoekstra, F.A. & Crowe, L.M. *Annu. Rev. Physiol.* **54** 579-599 (1992)
- [51] Walter, R.H. Polysaccharide association structures in food. Marcel Dekker Inc., New York, U.S.A. (1998)
- [52] Limbach, H.J. & Kremer, K. *Trends Food Sci. Technol.* **17** 215-219 (2006)
- [53] Pereira, C.S., Kony, D., Baron, R., Müller, M., van Gunsteren, W.F. & Hünenberger, P.H. *Biophys. J.* **90** 4337-4344 (2006)
- [54] Pereira, C.S., Kony, D., Baron, R., Müller, M., van Gunsteren, W.F. & Hünenberger, P.H. *Biophys. J.* **93** 706- (2007)
- [55] Lins, R.D. & Hünenberger, P.H. *J. Comput. Chem.* **26** 1400-1412 (2005)
- [56] Schuler, L.D. & van Gunsteren, W.F. *Mol. Simul.* **25** 301-319 (2000)
- [57] Schuler, L.D., Daura, X. & van Gunsteren, W.F. *J. Comput. Chem.* **22** 1205-1218 (2001)
- [58] Chandrasekhar, I., Kastenholz, M.A., Lins, R.D., Oostenbrink, C., Schüler, L.D., Tieleman, D.P. & van Gunsteren, W.F. *Eur. Biophys. J.* **32** 67-77 (2003)
- [59] Soares, T.A., Hünenberger, P.H., Kastenholz, M.A., Kräutler, V., Lenz, T., Lins, R.D., Oostenbrink, C. & van Gunsteren, W.F. *J. Comput. Chem.* **26** 725-737 (2005)
- [60] Hansen, H.S. & Hünenberger, P.H. *J. Comput. Chem.*, in press (2009)
- [61] Karplus, M. & Kushick, J.N. *Macromolecules* **14** 325-332 (1981)
- [62] Schlitter, J. *Chem. Phys. Lett.* **215** 617-621 (1993)
- [63] Schäfer, H., Mark, A.E. & van Gunsteren, W.F. *J. Chem. Phys.* **113** 7809-7817 (2000)
- [64] Andricioaei, I. & Karplus, M. *J. Chem. Phys.* **115** 6289-6292 (2001)
- [65] Baron, R., van Gunsteren, W.F. & Hünenberger, P.H. *Trends Phys. Chem.* **11** 87-122 (2006)

- [66] Huber, T., Torda, A.E. & van Gunsteren, W.F. *J. Comput. Aided Mol. Design* **8** 695-708 (1994)
- [67] Torrie, G.M. & Valleau, J.P. *J. Comput. Phys.* **23** 187-199 (1977)
- [68] Valleau, J.P. & Torrie, G.M. A guide to Monte Carlo for statistical mechanics: 1. Highways. In: *Modern theoretical chemistry. Volume 5*. Berne, B.J., Ed. Plenum Press, New York, U.S.A. pp 169-194 (1977)
- [69] Eyring, H. *Trans. Faraday Soc.* **34** 41-48 (1937)
- [70] Mezei, M. *J. Comput. Phys.* **68** 237-248 (1987)
- [71] Hooft, R.W.W., van Eijck, B.P. & Kroon, J. *J. Chem. Phys.* **97** 6690-6694 (1992)
- [72] Leitgeb, M., Schröder, C. & Boresch, S. *J. Chem. Phys.* **122** 084109/1-084109/15 (2005)
- [73] Darve, E. & Pohorille, A. *J. Chem. Phys.* **115** 9169-9183 (2001)
- [74] Darve, E., Wilson, M.A. & Pohorille, A. *Mol. Simul.* **28** 113-144 (2002)
- [75] Darve, E., Rodriguez-Gomez, D. & Pohorille, A. *J. Chem. Phys.* **128** 144120/1-144120/13 (2008)
- [76] Laio, A. & Parrinello, M. *Proc. Natl. Acad. Sci. USA* **99** 12562-12566 (2002)
- [77] Laio, A., Rodriguez-Forteza, A., Gervasio, F.L., Ceccarelli, M. & Parrinello, M. *J. Phys. Chem. B* **109** 6714-6721 (2005)
- [78] Pickett, H.M. & Strauss, H.L. *J. Am. Chem. Soc.* **92** 7281-7290 (1970)
- [79] Polacek, R., Stenger, J. & Kaatze, U. *J. Chem. Phys.* **116** 2973-2982 (2002)
- [80] Stenger, J., Cowman, M.K., Eggers, F., Eyring, E.M., Kaatze, U. & Petrucci, S. *J. Phys. Chem. B* **104** 4782-4790 (2000)
- [81] Hagen, R. & Kaatze, U. *J. Chem. Phys.* **120** 9656-9664 (2004)
- [82] Behrends, R. & Kaatze, U. *ChemPhysChem* **6** 1133-1145 (2005)
- [83] Hajduk, P.J., Horita, D.A. & Lerner, L.E. *J. Am. Chem. Soc.* **115** 9196-9201 (1993)
- [84] Horita, D.A., Hajduk, P.J. & Lerner, L.E. *Glycoconjugate J.* **14** 691-696 (1997)
- [85] Zhu, Y., Zajicek, J. & Serianni, A.S. *J. Org. Chem.* **66** 6244-6251 (2001)
- [86] Chu, S.S.C. & Jeffrey, G.A. *Acta Cryst. B* **24** 830- (1968)
- [87] Takeda, H., Yasuoka, N. & Kasai, N.. *Carbohydr. Res.* **53** 137-152 (1977)
- [88] Linden, A. *Acta Cryst.* **C51** 1007-1012 (1995)

- [89] Lee, C.K. & Koh, L.L. *Acta Cryst.* **49C** 621-624 (1993)
- [90] Taga, T. & Miwa, Y.M. *Acta Cryst.* **C53** 234-236 (1997)
- [91] Jeffrey, G.A. & Nanni, R. *Carbohydrate Res.* **137** 21-30 (1985)
- [92] Brown, G.M. et al. *Acta Cryst.* **B28** 3145- (1972)
- [93] Ohruji, H., Nishida, Y., Watanabe, M., Hori, H. & Meguro, H. *Tetrahedron Lett.* **26** 3251-3254 (1985)
- [94] Lipkind, G.M., Shashkov, A.S. & Kochetkov, N.K. *Carbohydr. Res.* **141** 191-197 (1985)
- [95] Shashkov, A.S., Lipkind, G.M. & Kochetkov, N.K. *Carbohydr. Res.* **147** 175-182 (1986)
- [96] Poppe, L. *J. Am. Chem. Soc.* **115** 8241-8426 (1993)
- [97] Leeftang, B.R., Vliegthart, J.F.G., Kroon-Batenburg, L.M.J., van Eijck, B.P. & Kroon, J. *Carbohydr. Res.* **230** 41-61 (1992)
- [98] Sugiyama, H., Hisamichi, K., Usui, T., Sakai, K. & Ishiyama, J. *J. Mol. Struct.* **556** 173-177 (2000)
- [99] Cheetham, N.W.H., Dasgupta, P. & Ball, G.E. *Carbohydr. Res.* **338** 955-962 (2003)
- [100] Zhang, P., Klymachyov A.N., Brown S., Ellington, J.G. & Grandinetti, P.J. *Solid State Nuclear Magnetic Resonance* **12** 221-225 (1998)
- [101] Andre, I., Mazeau, K., Taravel F.R., Tvaroska, *New J. Chem.* **19** 331-339 (1995)
- [102] Engelsen, S.B., Monteiro, C., Hervé de Penhoat, C. & Pérez, S. *Biophys. Chem.* **93** 103-127 (2001)
- [103] Roslund, M.U., Tähtinen, P., Niemitz, M., *Carbohydr. Res.* **343** 101-112 (2007)
- [104] Best, R.B., Jackson, G.E. & Naidoo, K.J. *J. Phys. Chem. B* **105** 4742-4751 (2001)
- [105] Matsuo, K. & Gekko, K. *Carbohydr. Res.* **339** 591-597 (2004)
- [106] Tvaroška, Imberty, A., Perez, *Biopolymers* **30** 369-379 (1990)
- [107] French, A.D., Kelterer, A-M., Johnson, G.P., Dowd, M.K. & Cramer, C.J. *J.Comp.Chem.* **22** 65-78 (2001)
- [108] French, A.D., Kelterer, A-M., Johnson, G.P., Dowd, M.K. & Cramer, C.J. *J. Mol. Struct.* **556** 303-313 (2000)
- [109] Strati, G.L., Willett, J.L. & Momany, F.A. *Carbohydr. Res.* **337** 1851-1859 (2002)
- [110] Strati, G.L., Willett, J.L. & Momany, F.A. *Carbohydr. Res.* **337** 1833-1849 (2002)
- [111] Javaroni, F., Ferreira, A.B.B. & da Silva, C.O. *Carbohydr. Res.* **344** 1235-1247 (2009)

- [112] Dowd, M.K., French, A.D., *Carbohydr. Res.* **233** 15-34 (1992)
- [113] Dowd, M.K., Zeng, J., French, A.D., *Carbohydr. Res.* **230** 223-244 (1992)
- [114] Dowd, M.K., French, A.D. & Reilly, P.J. *Biopolymers* **34** 625-638 (1994)
- [115] Stortz, C.A. & Cerezo, A.S. *Carbohydr. Res.* **338** 95-107 (2003)
- [116] Stortz, C.A. & Cerezo, A.S. *Carbohydr. Res.* **338** 1679-1689 (2003)
- [117] Starzak, M. & Mathlouthi, *Food Chemistry* **96** 346-370 (2006)
- [118] French, A.D. & Johnson, G.P. *Cellulose* **11** 449-462 (2004)
- [119] Seo, S., McClintock, M. L. & Wei, A. *Chem. Bio. Chem.* **7** 1959-1964 (2006)
- [120] Stortz, C.A. & French, A.D. *Molecular Simulation* **34** 373-389 (2008)
- [121] Ha, S.N., Madsen, L.J. & Brady, J.W. *Biopolymers* **27** 1927-1952 (1988)
- [122] Hardy, B.J. & Sarko, A. *J. Comput. Chem.* **14** 831-847 (1993)
- [123] Hardy, B.J. & Sarko, A. *J. Comput. Chem.* **14** 848-857 (1993)
- [124] Ott, K.-H. & Meyer, B. *J. Comput. Chem.* **17** 1068-1084 (1996)
- [125] Hardy, B.J., Gutierrez, A., Lesiak, K., Seidl, E. & Wildmalm, G. *J. Phys. Chem.* **100** 9187-9192 (1996)
- [126] Liu, Q., Schmidt, R.K., Teo, B., Karplus, P.A. & Brady, J.W. *J. Am. Chem. Soc.* **119** 7851-7862 (1997)
- [127] Bonanno, G., Noto, R. & Fornili, S.L. *J. Chem.Soc. Faraday Trans.* **94** 2755-2762 (1998)
- [128] Naidoo, K.J. & Kuttel, M. *J. Comput. Chem.* **4** 445-456 (2001)
- [129] Corzana, F., Motawia, M.S., Hervé du Penhoat, C., Perez, S., Tschampel, S.M., Woods, R.J. & Engelsen, S.B. *J. Comput. Chem.* **25** 573-586 (2004)
- [130] Yu, H., Amann, M., Hansson, T., Köhler, J., Wich, G. & van Gunsteren, W.F. *Carbohydr. Res.* **339** 1697-1709 (2004)
- [131] Kony, D., Damm, W., Stoll, S. & Hünenberger, P.H. *J. Phys. Chem. B* **108** 5815-5826 (2004)
- [132] Engelsen, S.B. & Pérez, S. *J. Phys. Chem. B* **104** 9301-9311 (2000)
- [133] Momany, F.A. & Willett, J.L. *Carbohydr. Res.* **326** 210-226 (2000)
- [134] Ekdawi-Sever, N.C., Conrad, P.B. & de Pablo, J.J. *J. Phys. Chem.* **105** 734-742 (2001)
- [135] Cheetham, N.W.H. & Dasgupta, P. *Aust. J. Chem.* **58** 803-809 (2005)

- [136] Lerbret, A., Bordat, P., Affouard, F., Descamps, M. & Migliardo, F. *J. Phys. Chem. B* **109** 11046-11057 (2005)
- [137] Choi, Y., Cho, K.W., Jeong, K. & Jung, S. *Carbohydr. Res.* **341** 1020-1028 (2006)
- [138] van Gunsteren, W.F., Billeter, S.R., Eising, A.A., Hünenberger, P.H., Krüger, P., Mark, A.E., Scott, W.R.P. & Tironi, I.G. Biomolecular simulation: The GROMOS96 manual and user guide. Verlag der Fachvereine, Zürich, Switzerland. (1996)
- [139] Scott, W.R.P., Hünenberger, P.H., Tironi, I.G., Mark, A.E., Billeter, S.R., Fennen, J., Torda, A.E., Huber, T., Krüger, P. & van Gunsteren, W.F. *J. Phys. Chem. A* **103** 3596-3607 (1999)
- [140] Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F. & Hermans, J. Interaction models for water in relation to protein hydration. In: *Intermolecular Forces*. Pullman, B., Ed. Reidel, Dordrecht, The Netherlands. pp 331-342 (1981)
- [141] Oostenbrink, C., Villa, A., Mark, A.E. & van Gunsteren, W.F. *J. Comp. Chem.* **25** 1656-1676 (2004)
- [142] Oostenbrink, C., Soares, T.A., van der Vegt, N.F.A. & van Gunsteren, W.F. *Eur. Biophys. J.* **34** 273-284 (2005)
- [143] Hockney, R.W. *Methods Comput. Phys.* **9** 136-211 (1970)
- [144] Ryckaert, J.-P., Ciccotti, G. & Berendsen, H.J.C. *J. Comput. Phys.* **23** 327-341 (1977)
- [145] Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., Di Nola, A. & Haak, J.R. *J. Chem. Phys.* **81** 3684-3690 (1984)
- [146] van Gunsteren, W.F. & Berendsen, H.J.C. *Angew. Chem. Int. Ed.* **29** 992-1023 (1990)
- [147] Barker, J.A. & Watts, R.O. *Mol. Phys.* **26** 789-792 (1973)
- [148] Tironi, I.G., Sperb, R., Smith, P.E. & van Gunsteren, W.F. *J. Chem. Phys.* **102** 5451-5459 (1995)
- [149] Heinz, T.N., van Gunsteren, W.F. & Hünenberger, P.H. *J. Chem. Phys.* **115** 1125-1136 (2001)
- [150] Insight II Molecular Modeling Package. Molecular Simulations Inc. (2000)
- [151] IUPAC, - *Biochemistry* **9** 3471-3479 (1970)
- [152] McNaught, A.D. *Pure Appl. Chem.* **68** 1919-2008 (1996)
- [153] Hünenberger, P.H., Granwehr, J.K., Aebischer, J.-N., Ghoneim, N., Haselbach, E. & van Gunsteren, W.F. *J. Am. Chem. Soc.* **119** 7533-7544 (1997)
- [154] Wu, D. & Kofke, D.A. *J. Chem. Phys.* **123** 054103/1-054103/10 (2005)

- [155] Wu, D. & Kofke, D.A. *J. Chem. Phys.* **123** 084109/1-084109/10 (2005)
- [156] Shell, M.S. *J. Chem. Phys.* **129** 144108/1-144108/7 (2008)
- [157] Shen, T. & Hamelberg, D. *J. Chem. Phys.* **129** 034103/1-034103/9 (2008)
- [158] Lemieux, R.U. & Koto, S. *Tetrahedron* **30** 1933-1944 (1974)
- [159] Pérez, S. & Marchessault, R.H. *Carbohydr. Res.* **65** 114-120 (1978)
- [160] Thogersen, H., Lemieux, R.U., Bock, K. & Meyer, B. *Can. J. Chem.* **60** 44-57 (1982)
- [161] Inagaki, S., Iwase, K. & Mori, Y. *Chem. Lett.* **3** 417-420 (1986)
- [162] Praly, J.-P. & Lemieux, R.U. *Can. J. Chem.* **65** 213-223 (1987)
- [163] Juaristi, E. & Cuevas, G. *Tetrahedron* **48** 5019-5087 (1992)
- [164] Graczyk, P.P. & Mikolajczyk, M. *Topics Stereochem.* **21** 159-349 (1994)
- [165] Tvaroška, I. & Carver, J.P. *Carbohydr. Res.* **309** 1-9 (1998)
- [166] Nishida, Y., Ohruai, H. & Meguro, H. *Tetrahedron Lett.* **25** 1575-1578 (1984)
- [167] Nishida, Y., Hori, H., Ohruai, H. & Meguro, H. *J. Carbohydr. Chem.* **7** 239-250 (1988)
- [168] Abraham, R.J., Chambers, E.J. & Thomas, W.A. *Magn. Reson. Chem.* **30** S60-S65 (1992)
- [169] Abraham, R.J., Chambers, E.J. & Thomas, W.A. *Magn. Reson. Chem.* **32** 248-254 (1994)
- [170] Tvaroška, I. & Gajdoš, J. *Carbohydr. Res.* **271** 151-162 (1995)
- [171] Brochier-Salon, M.-C. & Morin, C. *Magn. Reson. Chem.* **38** 1041-1042 (2000)
- [172] Tvaroška, I., Tavel, F.R., Utille, J.P. & Carver, J.P. *Carbohydr. Res.* **337** 353-367 (2002)
- [173] Stenutz, R., Carmichael, I., Widmalm, G. & Serianni, A.S. *J. Org. Chem.* **67** 949-958 (2002)
- [174] Thibaudeau, C., Stenutz, R., Hertz, B., Klepach, T., Zhao, S., Wu, Q., Carmichael, I. & Serianni, A.S. *J. Am. Chem. Soc.* **126** 15668-15685 (2004)
- [175] Suzuki, T., Kawashima, H. & Sota, T. *J. Phys. Chem. B* **110** 2405-2418 (2006)
- [176] Mayato, C., Dorta, R.L. & Vázquez, J.T. *Tetrahedron: Asymmetry* **15** 2385-2397 (2004)
- [177] Marchessault, R.H. & Pérez, S. *Biopolymers* **18** 2369-2374 (1979)
- [178] Jeffrey, G.A. *Acta Crystallogr. B* **46** 89-103 (1990)
- [179] Abraham, R.J., Chambers, E.J. & Thomas, W.A. *Carbohydr. Res. C* **226** 1-5 (1992)
- [180] de Vries, N.K. & Buck, H.M. *Carbohydr. Res.* **165** 1-16 (1987)

- [181] de Vries, N.K. & Buck, H.M. *Rec. Trav. Chim. Pays-Bas* **106** 453-460 (1987)
- [182] Rockwell, G.D., *J. Am. Chem. Soc.* **120** 100953-100963 (1998)
- [183] Tvaroška, I. & Kožar, T. *Theor. Chim. Acta* **70** 99-114 (1986)
- [184] Cramer, C.J. & Truhlar, D.G. *J. Am. Chem. Soc.* **115** 5745-5753 (1993)
- [185] Kuttel, M., Brady, J.W. & Naidoo, K.J. *J. Comput. Chem.* **23** 1236-1243 (2002)
- [186] Brady, J.W. *J. Am. Chem. Soc.* **111** 5155-5165 (1989)
- [187] Kroon-Batenburg, L.M.J. & Kroon, J. *Biopolymers* **29** 1243-1248 (1990)
- [188] Ha, S., Gao, J., Tidor, B., Brady, J.W. & Karplus, M. *J. Am. Chem. Soc.* **113** 1553-1557 (1991)
- [189] Kirschner, K.N. & Woods, R.J. *Proc. Natl. Acad. Sci. USA* **98** 10541-10545 (2001)
- [190] Haller, J. & Kaatze, U. *Chem. Phys. Lett.* **463** 413-417 (2008)
- [191] Benbrahim, N., Sekkal-Rahal, M. & Vergoten, G. *Spectrochimica Acta A* **58** 3021-3031 (2002)
- [192] Kolpak, F.J. & Blackwell, J. *Macromolecules.* **9** 273-278 (1976)
- [193] Kadla, J.F. & Gilbert, R.D. *Cell. Chem. Technol.* **34** 197-216 (2000)
- [194] Ichiye, T. & Karplus, M. *Proteins.* **2** 236-259 (1987)
- [195] Perahia, D., Levy, R.M. & Karplus, M. *Biopolymers.* **29** 645-677 (1990)
- [196] Striegel, A.M. *J. Am. Chem. Soc.* **125** 4146-4148 (2003)
- [197] Reeves, R.E. *J. Am. Chem. Soc.* **72** 1499-1506 (1950)
- [198] Angyal, S.J. *Angew. Chem. Int. Ed.* **8** 157-166 (1969)
- [199] Jensen, F.R. & Bushweller, C.H. *Adv. Alicyclic. Chem.* **3** 139-194 (1971)
- [200] Bentley, R. *Annu. Rev. Biochem.* **41** 953-996 (1972)
- [201] Mäler, L., Widmalm, G. & Kowalewski, J. *J. Phys. Chem.* **100** 17103-17110 (1996)
- [202] Kleinpeter, E. *Adv. Heterocyclic Chem.* **86** 41-127 (2004)
- [203] Yamaguchi, T. & Mizuno, S. *Bull. Chem. Soc. Jpn.* **10** 453-464 (1935)
- [204] Reuben, J. *J. Am. Chem. Soc.* **107** 5867-5870 (1985)
- [205] Snyder, J.R. & Serianni, A.S. *J. Org. Chem.* **51** 2694-2702 (1986)
- [206] Kurihara, Y. & Ueda, K. *Carbohydr. Res.* **341** 2565-2574 (2006)

- [207] Tse, Y.-C. & Newton, M.D. *J. Am. Chem. Soc.* **99** 611-613 (1977)
- [208] Jeffrey, G.A. & Lewis, L. *Carbohydr. Res.* **60** 179-182 (1978)
- [209] Jeffrey, G.A. & Mitra, J. *Acta Crystallogr.* **B39** 469-480 (1983)
- [210] Engelsen, S.B., Hervé du Penhoat, C. & Pérez, S. *J. Phys. Chem.* **99** 13334-13350 (1995)
- [211] Luque, F.J., López, J.M., López de la Paz, M., Vicent, C. & Orozco, M. *J. Phys. Chem. A* **102** 6690-6696 (1998)
- [212] López de la Paz, M., Ellis, G., Pérez, M., Perkins, J., Jiménez-Barbero, J. & Vicent, C. *Eur. J. Org. Chem.* **5** 840-855 (2002)
- [213] Hawley, J., Bampos, N., Aboitiz, N., Jiménez-Barbero, J., López de la Paz, M., Sanders, J.K.M., Carmona, P. & Vicent, C. *Eur. J. Org. Chem.* **12** 1925-1936 (2002)
- [214] Carcabal, P., Jockusch, R.A., Hünig, I., Snoek, L.C., Kroemer, R.T., Davis, B.G., Gambelin, D.P., Compagnin, I., Oomens, J. & Simons, J.P. *J. Am. Chem. Soc.* **127** 11414-11425 (2005)
- [215] Simons, J.P., Jockusch, R.A., Carcabal, P., Hünig, I., Kroemer, R.T., McLeod, N.A. & Snoek, L.C. *Int. Rev. Phys. Chem.* **24** 489-531 (2005)
- [216] Simperler, A., Watt, S.W., Bonnet, P.A., Jones, W. & Motherwell, W.D.S. *Cryst. Eng. Commun.* **8** 589-600 (2006)
- [217] Dashnau, J.L., Sharp, K.A. & Vanderkooi, J.M. *J. Phys. Chem. B* **109** 24152-24159 (2005)
- [218] Galema, S.A., Howard, E., Engberts, J.B.F.N. & Grigera, J.R. *Carbohydr. Res.* **265** 215-225 (1994)
- [219] Cheetham, N.W.H. & Lam, K. *Carbohydr. Res.* **282** 13-23 (1996)
- [220] Williams, R.J. & Leopold, A.C. *Plant. Physiol.* **89** 977-981 (1989)
- [221] Green, J.L. & Angell, C.A. *J. Phys. Chem.* **93** 2880-2882 (1989)
- [222] Sun, W.Q. & Leopold, A.C. *Ann. Bot.* **74** 601-604 (1994)
- [223] Koster, K.L., Webb, M.S., Bryant, G. & Lynch, D.V. *Biochim. Biophys. Acta* **1193** 143-150 (1994)
- [224] Sun, W.Q., Leopold, A.C., Crowe, L.M. & Crowe, J.H. *Biophys. J.* **70** 1769-1776 (1996)
- [225] Sun, W.Q. & Leopold, A.C. *Comp. Biochem. Physiol. A* **117** 327-333 (1997)
- [226] Koster, K.L., Lei, Y.P., Anderson, M., Martin, S. & Bryant, G. *Biophys. J.* **78** 1932-1946 (2000)
- [227] Chen, T., Fowler, A. & Toner, M. *Cryobiology* **40** 277-282 (2000)

- [228] Lerbret, A. et al. *Carbohydr. Res.* **340** 881-887 (2005)
- [229] Albertorio, F., Chapa, V.A., Chen, X., Diaz, A.J. & Cremer, P.S. *J. Am. Chem. Soc.* **129** 10567-10574 (2007)
- [230] Pereira, C.S. & Hünenberger, P.H. *Biophys. J.* **95** 3525-3534 (2008)
- [231] Kony, D., Damm, W., Stoll, S., van Gunsteren, W.F. & Hünenberger, P.H. *Biophys. J.* **93** 442-455 (2007)
- [232] Imberty, A. & Pérez, S. *Int. J. Biol. Macromol.* **11** 177- (1989)
- [233] Buléon, A. & Tran, V. *Int. J. Biol. Macromol.* **12** 345- (1990)
- [234] Pigman, W. & Isbell, H.S. *Adv. Carbohydr. Chem. Biochem.* **23** 11-57 (1968)
- [235] Lee, C.Y., Acree, T.E. & Schallenberger, R.S. *Carbohydr. Res.* **9** 356-360 (1969)
- [236] Acree, T.E., Schallenberger, R.S., Lee, Y. & Einset, J.W. *Carbohydr. Res.* **10** 355-360 (1969)
- [237] Wertz, P.W., Garver, J.C. & Anderson, L. *J. Am. Chem. Soc.* **103** 3916-3922 (1981)
- [238] Serianni, A.S., Pierce, J., Huang, S.-G. & Barker, R. *J. Am. Chem. Soc.* **104** 4037-4044 (1982)
- [239] Snyder, J.R., Johnston, E.R. & Serianni, A.S. *J. Am. Chem. Soc.* **111** 2681-2687 (1989)
- [240] Lewis, B.E., Choytun, N., Schramm, V.L. & Bennet, A.J. *J. Am. Chem. Soc.* **128** 5049-5058 (2006)

Chapter 3

Interaction of the disaccharides trehalose and gentiobiose with lipid bilayers: a comparative molecular dynamics study

3.1 Summary

The non-reducing glucose disaccharide α,α -trehalose (TRH) is known for its bioprotective action in organisms subject to stressful environmental conditions. However, the mechanisms whereby TRH stabilizes biomolecules under harsh conditions remains matter of debate, the five main hypotheses (not all mutually exclusive) being : (i) water replacement (WRH); (ii) headgroup bridging (HBH); (iii) vitrification (VIH); (iv) water entrapment (WEH); (v) hydration forces (HFH). Four of these hypotheses (all except HFH) are in principle compatible with a preferential affinity of the sugar molecules (compared to water) for the biomolecular surface. According to the recently proposed sugar-like mechanism, preferential affinity would result from the large entropy gain associated with the release of water molecules from the surface region at the cost of immobilizing and rigidifying fewer sugar molecules. Assuming the validity of this mechanism, a more flexible disaccharide such as gentiobiose (GNT) should evidence a weaker preferential affinity and decreased bioprotective ability. In the present work, explicit solvent molecular dynamics (MD) simulation of a dipalmitoyl-phosphatidylcholine (DPPC) bilayer patch (at constant lateral area and normal pressure) in the presence of either pure water or of GNT or TRH solutions are performed (at 0.8 or 1.6 m sugar concentration and considering temperatures of 323 or 475 K) in order to assess the validity of this suggestion. TRH (at 1.6 m concentration) is indeed found to essentially preserve the bilayer structure at 475 K, while GNT does not. However, the comparison of the system properties at 323 K does not permit to unambiguously ascribe this difference to one of the four above mechanisms (all except HBH) compatible with the (observed) preferential affinity, because the simulations reveal some characteristic features of all of them. This suggests either that multiple mechanisms may be simultaneously active or that the definitions of these proposed mechanisms is not precise enough for an unambiguous assessment based on experimental or theoretical data.

3.2 Introduction

Nature has evolved numerous strategies for the long-term survival of organisms exposed to potentially damaging conditions like extreme dryness, cold, heat, pressure, salt concentration, acidity or oxygen deprivation^{1,2}. One of these strategies, cryptobiosis, involves a reversible suspension of the metabolism and a temporary effective isolation from the environmental stress^{3,4}. Probably the best studied case of cryptobiosis concerns anhydrobiosis, namely the resistance to nearly complete dehydration, as observed in the life cycle of some yeasts, rotifers, nematodes, tardigrades, shrimps and plants^{1,4-8}. A characteristic feature of the anhydrobiotic process (as well as of similar processes triggered by stressful temperature, pressure, osmotic or chemical conditions) is the accumulation of large amounts of saccharides in the cell^{9,10}, and in particular of the disaccharide trehalose^{5,11} (α,α -trehalose; TRH; Figure 3.1). As a particularly spectacular evidence of the role of sugars in bioprotection, it has been shown that the expression in human fibroblasts of the genes responsible for encoding the TRH biosynthetic enzymes of *Escherichia coli* permits the survival of these cells in the dry state for up to a few days¹². Note, however, that other types of cosolutes including *e.g.* fructans^{13,14}, raffinose family oligosaccharides¹⁵ or heat-shock proteins¹⁶⁻¹⁸ (HSPs) are often involved in interplay with sugars for biopreservation.

It is generally accepted that bioprotection involves the stabilization of biomolecules such as proteins¹⁹⁻²⁵ (against irreversible denaturation) and lipid membranes^{7,9,26-33} (against mechanical disruption). A possible stabilization of nucleic acids³⁴ (against denaturation) has also been recently proposed. Five possible mechanisms have been put forward to tentatively explain the bioprotective action of sugars, principally concerning the best documented example of protection against dehydration^{35,36}:

1. The *water replacement hypothesis* (WRH) suggests that during dehydration, sugars molecules can substitute water molecules in forming hydrogen bonds (H-bonds) with the polar and charged groups present at the biomolecular surface, thereby inhibiting the denaturation (proteins) or a transition to the gel phase (membranes)^{5,37-47}.
2. The *headgroup-bridging hypothesis* (HBH) is an extension of the WRH formulated based on simulations in the context of membrane bioprotection, which proposes that sugar form a scaffold of H-bonds bridging multiple headgroups, thereby inhibiting a transition to the gel phase upon dehydration³⁶ (this scaffold is labile in the dilute regime^{36,48-50} but expected to strengthen upon dehydration, as a result of the removal of the water molecules competing for the headgroup H-bonding sites and of the reduction of dielectric screening effects).
3. The *vitrification hypothesis* (VIH) suggests that sugars found in anhydrobiotic systems, known to be good vitrifying agents, protect biomolecules through the formation of amorphous glasses, thereby reducing structural fluctuations and preventing the denaturation (proteins) or mechanical disruption (membranes)⁵¹⁻⁵⁷.
4. The *water-entrapment hypothesis* (WEH) proposes that sugars concentrate residual water molecules close to the biomolecular surface, thereby preserving to a large extent its solvation and native properties in the quasi-dry state⁵⁸⁻⁶⁴.
5. The *hydration forces hypothesis* (HFH), introduced more recently in the context of membrane bioprotection, proposes that sugar molecules are preferentially excluded from the

biomolecular surface,^{65,66} and indirectly reduce the compressive stress in the membrane upon dehydration (“hydration repulsion” between bilayers brought in contact), thereby preventing a transition to the gel phase^{65,67}.

Note that these mechanisms are not necessarily all mutually exclusive. For example, the WRH and VIH hypotheses are often invoked simultaneously to explain the bioprotective effect of sugars against dehydration^{4,9,10,41,56,68}. They are also compatible with the HBH (an extension of WRH) and WEH (concerning the preferential location of the residual water molecules in the quasi-dry state) hypotheses. Similarly, the HFH hypothesis is compatible with the VIH hypothesis, and could also be combined with the WEH hypothesis (now viewed as the limiting case of preferential exclusion in the quasi-dry state). However, the two alternative types of possible combinations (WRH+HBH+VIH+WEH or HFH+WEH+VIH) cannot be reconciled. Since there is experimental or/and theoretical evidence supporting each of the four above hypotheses in specific situations, it is not impossible that the bioprotection mechanism varies depending on the nature of the environmental stress (*e.g.* dehydration, temperature, pressure, osmotic or chemical conditions) and its intensity (*e.g.* for anhydrobiosis, extent of dehydration), as well as on the type of protected biomolecule (*e.g.* protein or membrane).

Recent molecular dynamics (MD) simulation studies indeed suggested that there is a qualitative difference in the way sugars interact with membranes and with proteins. In the case of membranes, the simulations^{36,48–50,69–78} systematically revealed a preferential affinity of the sugar molecules for the bilayer surface and a direct interaction (through H-bonds) with the lipid headgroups, even in dilute solutions, providing support for the WRH hypothesis (possibly along with the HBH hypotheses^{36,48–50}). In contrast, in the case of proteins, the simulations^{58,60,61,63,64,64,79–83} rather suggested a preferential exclusion of the sugar molecules from the biomolecular surface (trapped water layer), even in concentrated solutions, providing support for the WEH hypothesis (along with the VIH hypothesis for concentrated sugar solutions⁸⁰).

Closely related to the question of the bioprotection mechanism by sugars is that of the particular efficiency of TRH (compared to *e.g.* other disaccharides or oligosaccharides) in terms of bioprotective action. TRH is the symmetrical non-reducing disaccharide composed of two D-glucopyranose units in an (axial-axial) $\alpha(1\rightarrow1)\alpha$ linkage. Recent MD simulation studies of the 11 glucose-based disaccharides in water⁸⁴ (see Chapter 2 of this thesis) have revealed two peculiar features of TRH which, in combination, confer to this compound rather unique properties in this series :

1. TRH presents the highest conformational rigidity (lowest conformational entropy) at the level of its glycosidic linkage, resulting from the action of the *exo*-anomeric effect (in its most restrictive form for α -anomers) on both of its glycosidic dihedral angles.
2. TRH presents no intramolecular H-bonds in its single dominant “clam-shell” glycosidic linkage conformation (the only conformation significantly populated at room temperature, as well as the one found in the crystallographic structure⁸⁵), resulting in the highest hydrophilicity and, presumably, the highest propensity to intermolecular H-bonded interactions (*e.g.* self-association or interaction with other H-bonding solutes).

The fact that it is a highly rigid disaccharide, rather than a (smaller) monosaccharide or a (more flexible) oligosaccharide suggests that TRH might represent an optimal balance in terms of providing a high density of molecular hydroxyl groups in essentially fixed relative positions

and orientations, and exempt of mutual (intramolecular H-bonding) interactions (*i.e.* available for intermolecular interactions with the solvent, with other sugar molecules or with biomolecules). In pictorial terms, TRH could thus be viewed as a kind of “hydroxyl porcupine”. The specific disposition of the hydroxyl groups in the “clam-shell” conformation (*i.e.* mostly on one side of the molecule) may also itself facilitate H-bonded interactions between the sugar and (polar) planar or quasi-planar surfaces (*e.g.* proteins or membranes)^{86,87}.

The two above conformational factors may also in turn be largely responsible for a number of peculiar physico-chemical properties of this disaccharide^{26,27,88}. The phase diagram of TRH-water mixtures presents unusual features^{88,89}. Trehalose dihydrate melts at 97°C. However, upon additional heating, the disaccharide recrystallizes at 130°C in an anhydrous form, that melts again at 203°C. TRH-water mixtures also present particularly high glass transition temperatures (*e.g.* $T_g = 243.7^\circ\text{C}$ at 20 wt% H₂O)⁸⁹ compared to other disaccharides. Experimental⁹⁰ and theoretical^{91–98} studies evidenced strong kosmotropic properties for this disaccharide (*i.e.* a propensity of the TRH molecules to promote water ordering in their vicinity, thereby significantly perturbing the water H-bonding network). Along with the two main conformational features described above, these peculiar physico-chemical properties may also play a direct role in the bioprotective action of TRH. In addition, the difficulty to perform the hydrolysis of TRH in acidic solutions and the inability of α -glucosidases to cleave its glycosidic linkage make TRH a highly stable disaccharide *in vivo*⁸⁸. Finally, it has been suggested that this molecule may also act as an antioxidant^{99,100} (scavenger of reactive oxygen species). The particular (physico-chemical, chemical and bioprotective) properties of the TRH molecule have given rise to a wide spectrum of technological applications^{11,101,102} *e.g.* in the food (*e.g.* microencapsulation^{103,104}, crystallization inhibition^{105,106}, dried starter cultures stabilization³⁵), pharmaceutical (*e.g.* dry cell stabilization^{7,107}, amorphization of pharmaceuticals¹⁰⁸, inhibition of cataractogenesis²¹, inhibition of peptide aggregation in neurodegenerative diseases¹⁰⁰) and cosmetic industries.

In a recent study, Pereira and Hünenberger³⁶ proposed (based on MD simulations and available experimental data) two mechanisms for the interaction of polyhydroxylated cosolutes with lipid bilayers. The *alcohol-like mechanism* (active for small alcohols and polyols) involves preferential affinity of the cosolute (compared to water) for the superficial region of the bilayer interior, and is driven by the hydrophobic effect. It results in a lateral expansion of the membrane, a disorder increase within bilayer, and a partial substitution of water by cosolute molecules at the H-bonding sites provided by the membrane (predominantly at the level of the ester groups). The *sugar-like mechanism* (active for larger polyols including saccharides) involves preferential affinity of the cosolute (compared to water) for the bilayer surface and is driven by entropic effects. It results in the absence of lateral expansion and change in disorder within the bilayer, and in a partial substitution of water by cosolute molecules (predominantly at the level of the phosphate groups). It also involves the clustering of the cosolute molecules at the membrane surface (formation of a coating layer) and the bridging of lipid molecules *via* H-bonded cosolute molecules. H-bonding itself is not viewed as a driving force for these two mechanisms, which only involve the (partial) substitution of water-lipid by cosolute-lipid H-bonds (the sum of the two remaining essentially constant, irrespective of the nature and concentration of the cosolute).

If one accepts the sugar-like mechanism, which is compatible with the WRH, HBH, VIH and possibly WEH (but not the HEH) hypotheses, as providing an essentially correct description of sugar-membrane interactions, the main driving force for the clustering of sugar molecules at the bilayer surface is the release of water molecules from the surface region to the bulk (large entropy

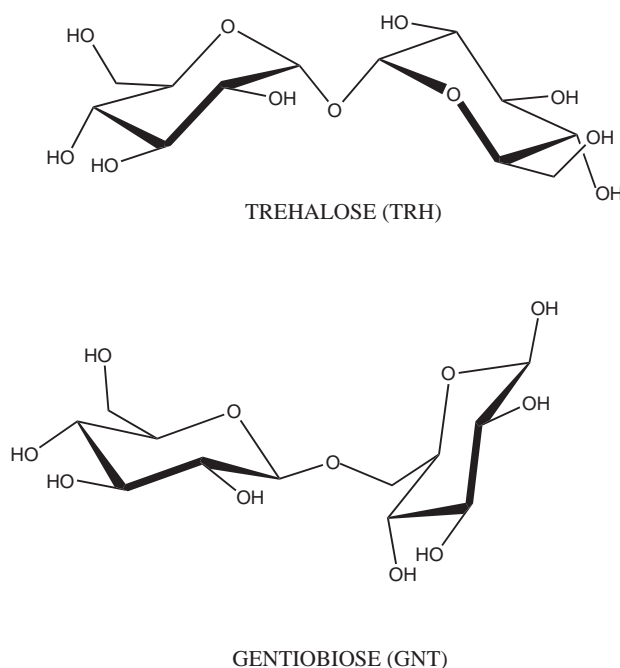


Figure 3.1: Structures of the two disaccharides considered in the present study. Trehalose (more precisely α,α -trehalose; TRH; $\text{Glc } \alpha(1 \rightarrow 1)\alpha\text{Glc}$) is the non-reducing disaccharide consisting of two D-glucopyranose (Glc) residues in a $\alpha(1 \rightarrow 1)\alpha$ (two-bond, axial-axial) glycosidic linkage. Gentiobiose (GNT; represented here in its β -anomeric form at the reducing residue; $\text{Glc } \beta(1 \rightarrow 6)\text{Glc}\beta$) is the reducing disaccharide consisting of two Glc residues in a $\beta(1 \rightarrow 6)$ (three-bond, equatorial-equatorial) glycosidic linkage.

increase) at the cost of immobilizing and rigidifying fewer (already inherently rather rigid) sugar molecules in this same region (limited entropy cost), the process being enthalpically essentially neutral (exchange of H-bonding partners for the lipid headgroups). In this case, the particularly strong bioprotective capacity of TRH could be in large part a consequence of its inherently limited conformational flexibility (low entropy cost of rigidification), other contributing factors being a favorable hydroxyl group disposition^{86,87} (to interact with a planar surface), a strong propensity to self-aggregate (surface clustering), an elevated glass transition temperature (contribution of the VIH mechanism) and a high chemical stability (against acidic or enzymatic hydrolysis).

If the above reasoning is correct, a more flexible disaccharide should evidence a reduced preferential affinity and bioprotective ability for lipid bilayers compared to TRH. Recent MD simulations of the 11 glucose-based disaccharides in water⁸⁴ (see Chapter 2 of this thesis) suggested that the $\beta(1 \rightarrow 6)$ -linked disaccharide gentiobiose (GNT; Figure 3.1), which has one more torsional degree of freedom in its glycosidic linkage, is, at the opposite of the TRH, the most flexible compound in this series. This observation is also corroborated by other experimental and theoretical studies^{109–113}.

In the contrast (to the two-bond, axial-axial) glycosidic linkage of TRH, the (three-bond, equatorial-equatorial) linkage of GNT is only affected by the *exo*-anomeric effect (in its less restrictive form for β -anomers) on its ϕ dihedral angle, and includes an additional dihedral angle ω . As a result, this compound is characterized by two alternative significantly populated conformers (differing by the value of ω), a broad probability distribution around ψ , and the highest

conformational entropy (flexibility) among all possible glucose-based disaccharides. Note, however, that neither TRH nor GNT present persistent intramolecular H-bonds in aqueous solution, which probably results in the former case from the low flexibility (in a conformation incompatible with the formation of intramolecular H-bonds) and in the latter case from the high flexibility (solvent exposed H-bonding in water presumably representing a weak conformational driving force compared to entropy). As a consequence, the hydroxyl groups of the two molecules are engaged to a similar extent in interactions with water in the dilute regime, and *a priori* similarly prone to intermolecular interaction in the more concentrated regime or in the presence of a lipid bilayer. Therefore, a comparison of TRH and GNT in terms of their interaction with membranes appears to represent an interesting test for the above ideas.

In the present work, explicit-solvent MD simulations of a dipalmitoyl-phosphatidylcholine (DPPC) bilayer patch in the presence of pure water or of aqueous solutions of the two disaccharides are reported and compared, in order to investigate the effect of the difference in flexibility between the two sugars on their interaction with the bilayer (and the resulting alteration of the bilayer properties), and on their potential for bioprotection. To this purpose, a total of ten simulations (10 ns each) have been carried out (at constant lateral area and normal pressure), differing in the disaccharide identity and concentration (pure water, GNT or TRH at either 0.8 or 1.6 m concentration, “m” indicating the solution molality in mol·kg⁻¹) and temperature (323 or 475 K).

3.3 Computational details

3.3.1 Molecular dynamics simulations

All MD simulations were performed using the GROMOS96 program^{114,115} together with the GROMOS 45A4 force-field^{116–120} (including recently reoptimized parameter sets for lipids¹¹⁹ and carbohydrates^{111,112,112,116,121,122}) and the SPC water model.¹²³ The simulations were carried out under periodic boundary conditions based on rectangular boxes containing a DPPC lipid bilayer patch of 2×6×6 lipid molecules (in the *xy*-plane, *i.e.* normal to the *z*-axis) in the presence of either pure water or a disaccharide (GNT or TRH) solution (Section 3.3.2). Newton’s equations of motion were integrated using the leap-frog scheme¹²⁴ with a timestep of 2 fs. All bond lengths were constrained using the SHAKE procedure¹²⁵ with a relative geometric tolerance of 10⁻⁴. The center of mass motion was removed every 100 ps. The simulations were performed in the $NA_{xy}P_zT$ ensemble (constant number of particles N , area in the *xy*-plane A_{xy} , pressure along the *z*-axis P_z and temperature T), with $P_z=1$ bar, $T = 323$ or 475 K and $A_{xy}=23.04$ nm², the latter using identical box dimensions along the *x*- and *y*-axes and based on an area per lipid of 0.64 nm² (consensus experimental estimate¹²⁶ for DPPC in the liquid-crystalline phase at full hydration and at 323 K). The temperature was maintained by weakly coupling the solute and solvent degrees of freedom separately to a temperature bath¹²⁷ at temperature T , with a relaxation time of 0.1 ps. The pressure along the *z*-axis was maintained by weakly coupling the particle coordinates and box dimension in the *z*-direction to a pressure bath¹²⁷ at pressure P_z , with a relaxation time of 0.5 ps and an isothermal compressibility of $4.575 \cdot 10^{-4}$ (kJ mol⁻¹ nm⁻³)⁻¹ as appropriate for water¹¹⁴. The non-bonded interactions were computed using a twin-range scheme,^{114,128} with short- and long-range cutoff distances set to 0.8 and 1.4 nm, respectively, and a frequency of 5 timesteps for the update of the short-range pairlist and intermediate-range interactions. A reaction-field cor-

rection^{129,130} was applied to account for the mean effect of electrostatic interactions beyond the long-range cutoff distance, using a relative dielectric permittivity of 61 as appropriate for the SPC water model¹³¹. Configurations were saved every 5 ps for analysis.

3.3.2 Molecular systems

The (periodic) DPPC bilayer patch ($2 \times 6 \times 6$ lipid molecules) was simulated in the presence of either pure water or of aqueous solutions of the disaccharides GNT or TRH, at different temperatures (323 or 475 K) and concentrations (0.8 or 1.6 m, “m” indicating the solution molality in mol kg^{-1}). The disaccharide GNT was simulated in its β -anomeric form at the reducing residue, expected to account for about 62% of the total anomeric population.¹³² The ten resulting simulations are labelled using a three-letter code : the successive letters refer to the boundary conditions (A: constant A_{xy} and P_z), nature of the solution environment (W: pure water; G_l and G_h : GNT 0.8 or 1.6 m; T_l and T_h : TRH 0.8 or 1.6 m) and temperature (L: lower temperature of 323 K; H: higher temperature of 475 K). The lower temperature of 323 K corresponds to a common choice in experimental studies concerning DPPC bilayers in the liquid crystalline phase¹²⁶. The higher temperature of 475 K is not meant to represent a physical situation. It is used here as a mere device to place the bilayer under stress conditions and investigate how the disaccharides affect the response of the membrane.

The two considered concentrations were modeled by inclusion of disaccharide molecules in ratios one-to-two (0.8 m; 36 disaccharide molecules) or one-to-one (1.6 m; total 72 disaccharide molecules) relative to the lipid molecules and appropriately selecting the corresponding numbers of water molecules (see below).

The initial structure of the DPPC bilayer was constructed by applying 36 lattice translations

| System | Ensemble | N_{GNT} | N_{TRH} | N_{WAT} | c [m] | T [K] | L_z [nm] |
|---------------------|---------------|-----------|-----------|-----------|---------|-----------|-------------|
| AWL (AWH) | $NA_{xy}P_zT$ | - | - | 2166 | 0 | 323 (475) | 6.71 (8.08) |
| AG_lL (AG_lH) | $NA_{xy}P_zT$ | 36 | - | 2518 | 0.79 | 323 (475) | 7.88 (9.23) |
| AT_lL (AT_lH) | $NA_{xy}P_zT$ | - | 36 | 2577 | 0.78 | 323 (475) | 7.90 (9.29) |
| AG_hL (AG_hH) | $NA_{xy}P_zT$ | 72 | - | 2581 | 1.55 | 323 (475) | 8.49 (9.92) |
| AT_hL (AT_hH) | $NA_{xy}P_zT$ | - | 72 | 2577 | 1.55 | 323 (475) | 8.53 (9.82) |

Table 3.1: Simulated systems and simulation conditions. The different columns report the simulation code (Section 3.3.2), the thermodynamical ensemble (Section 3.2.1), the composition of the computational box (number N of molecules of the different species; GNT: gentiobiose; TRH: trehalose; WAT: water; the bilayer patch consists of $2 \times 6 \times 6$ DPPC molecules in all cases), the disaccharide concentration (c ; “m” indicating the solutions molality in $\text{mol}\cdot\text{kg}^{-1}$), the simulation (thermostat) temperature, and the dimension of the (rectangular) computational box along the z-axis (bilayer normal; the corresponding dimensions in the xy-plane are $L_x = L_y = 4.8$ nm in all cases). All simulations were carried out for a total duration of 10 ns (after equilibration; Section 3.2.2). For compactness, entries corresponding to the simulations performed at high temperature are reported between parentheses.

and random axial rotations to a single pair of tail-to-tail lipid molecules, so as to generate two leaflets of areas A_{xy} . The resulting bilayer was solvated using 2166 SPC water molecules to reproduce full hydration¹²⁶ (about 30 water molecules per lipid). The system was slowly heated up to 323 K in the presence of progressively decreasing position restraints on the lipid atoms, over a total period of 0.27 ns. Equilibration was then carried out for 30 ns at 323 K. The need for such an extensive equilibration period is justified by the long timescales associated with the rotation of the lipids around their long axis and with the conformational relaxation of the lipid alkyl chains¹³³. The final configuration of this trajectory was used directly as initial configuration for the simulation of the bilayer in pure water at 323 K (for the corresponding simulation at 475 K, the temperature was first progressively increased over a time period of 0.24 ns). For the simulations in the presence of disaccharide solutions, the final configuration of the 323 K equilibration trajectory was desolvated (all water molecules deleted). The different systems (low and high concentration of GNT or TRH) were prepared by randomly placing the appropriate number of sugar molecules in the computational box. The systems were then resolvated to full hydration, imposing a minimal solute-solvent distance of 0.23 nm. A second equilibration procedure was then applied in order to allow for a proper distribution of the sugar molecules around the bilayer. The procedure is similar to the one described by Pereira *et al.*⁴⁸. The systems were heated up to 600 K and equilibrated for 0.56 ns maintaining the coordinates of all lipid atoms constrained. The energy of the systems were then minimized with 20000 steps of the steepest descent algorithm (constraints for the bilayer atoms were still used in this step). Finally, the systems were slowly heated up to the desired value (323 K or 475 K) in the presence of progressively decreasing position restraints on the lipid atoms, over a period of 0.24 ns. The final configurations of these equilibration trajectories were then used directly for the simulations of the bilayer in disaccharide solutions at 323 and 475 K.

The ten simulations were carried out for a total duration of 10 ns each (after equilibration). A summary of the simulated systems and simulation conditions is provided in Table 3.1.

3.3.3 Trajectory analysis

The simulations were analyzed in terms of normalized probability distributions $P_X(z)$ of specific atoms X along the z -axis (bilayer normal), overlaps C_{XY} between the corresponding sugar or water (WAT) distribution peaks (X) and lipid headgroup peaks (Y), order parameters $S_{CH}(C_n)$ of the methylene ($C_n; n = 2 \dots 15$) groups in the two lipid acyl chains, and intermolecular H-bonding.

The normalized probability distributions $P_X(z)$ were calculated for selected atoms X of DPPC, GNT, TRH and WAT. The selected atoms were the headgroup phosphorus atom ($X = P$) and the CH_2 glycerol united atom at the sn_1 position ($X = G$) of the lipid molecules (DPPC), the glycosidic oxygen atom ($X = S$) of the sugar molecules (GNT or TRH), and the oxygen atom ($X = W$) of the water molecules (WAT). These distribution profiles were calculated considering last 1 ns of each simulation.

The overlaps between the sugar ($X = S$) or WAT ($X = W$) distributions on the one hand, and the DPPC phosphate ($Y = P$) or glycerol ($Y = G$) distributions on the other hand, were further quantified in terms of an effective concentration C_{XY} of X in the neighborhood of Y . This quantity is defined as

$$C_{XY} = \frac{N_X}{A_{xy}} \int dz P_X(z) P_Y(z) \quad , \quad (3.1)$$

where N_X is the total number of molecules of species X in the simulation box. The meaning of C_{XY} is most easily understood in the idealized situation where P_Y is homogeneous over some interval Δz and zero outside. In this case, C_{XY} is equal to the total number of molecules of species X present within the region Δz divided by the total volume $A_{xy}\Delta z$ of this region, *i.e.* to the effective concentration of X in this region. In the general case, C_{XY} can be viewed as a measure of the local concentration of species X at the depth corresponding to species Y within the bilayer. In addition, the ratios C_{SP}/C_{WP} and C_{SG}/C_{WG} can be expressed in terms of local molalities m_P and m_G of the sugar in the regions of the phosphate and glycerol groups, respectively, which can be compared with the corresponding bulk molalities of 0.8 and 1.6 m, and provide an indication concerning the preferential affinity or exclusion of the sugar from the bilayer surface.

Deuterium order parameters (S_{CD}) can be derived from deuterium quadrupole splitting experiments and have been widely used to study biological membranes^{134–143} and to validate biomolecular force-fields^{119,144}. The corresponding carbon-hydrogen order parameters (S_{CH}) were calculated for the 14 methylene groups in each of the acyl tails of DPPC (sn_1 and sn_2 chains) by computing the correlation functions describing the reorientation of the carbon-hydrogen vectors. More precisely, for each methylene group along the chain, an order parameter tensor \underline{S} can be defined as $S_{ij} = \frac{1}{2}\langle 3\cos\theta_i \cos\theta_j - \delta_{ij} \rangle$, where θ_i is the angle between the i^{th} local molecular axis (x' , y' or z') and the bilayer normal (z -axis), δ_{ij} is the Kronecker delta symbol and $\langle \dots \rangle$ stands for trajectory averaging. Usually, for the n^{th} methylene group C_n , the direction of the vector $C_{n-1} - C_{n+1}$ is taken as z' , the direction of the vector normal to z' in the plane C_{n-1} , C_n , and C_{n+1} defines y' , while x' is defined by the direction of the vector perpendicular both to z' and y' ¹⁴⁵. The quantity $S_{CH} = -(2/3S_{xx} + 1/3S_{yy})$ is the value to be compared with the experimental S_{CD} value¹³⁴. These values were calculated considering the last 4 ns of each simulation, and averaged over the 72 DPPC molecules.

The presence of a H-bond was defined by a maximal hydrogen-oxygen distance of 0.25 nm and a minimal oxygen-hydrogen-oxygen angle of 135° . The occurring intermolecular H-bonds were classified according to the different pairs of species present in the simulations (sugar-sugar, sugar-DPPC, sugar-WAT and DPPC-WAT). All the oxygen atoms of the sugars, WAT and DPPC molecules were considered as H-bond acceptors and all the hydroxyl groups of the sugar and WAT molecules as H-bond donors. In addition, the H-bonded interaction between the sugar and DPPC molecules was further analyzed in terms of H-bonding patterns. A pattern is encoded by a series of integers (listed in decreasing order) indicating the number of H-bonds formed between a sugar molecule and a number of distinct DPPC molecules in a given system configuration. For example, the pattern 311 for a given sugar molecule represents a configuration in which this molecule is simultaneously triply H-bonded to a given DPPC molecule and singly H-bonded to two other DPPC molecules. One characteristic feature of a given pattern is the involved degree of bridging (number of integers in the pattern code), *i.e.* the number of distinct DPPC molecules forming at least one H-bond with the sugar molecule. The H-bonding analyzes were performed considering the last 4 ns of each simulation.

3.4 Results

3.4.1 Distribution profiles

The normalized probability distributions $P_X(z)$ corresponding to specific DPPC, GNT or TRH, and WAT atoms (X) along the bilayer normal (calculated considering the final 1 ns of the ten simulations) are displayed in Figure 3.2. The corresponding peak overlaps C_{XY} (effective concentration of X in the region Y) are reported in Table 3.2.

At 323K, all profiles present features characteristic of an intact bilayer structure. The peaks corresponding to the headgroup glycerol and phosphate atoms of the bilayer leaflets are sharp and well separated (at distances of about 2.9 and 3.5 nm, respectively), and water penetrates the bilayer no further than the region of the glycerol atoms. When comparing the simulation in pure water (AWL) to the other simulations involving disaccharide solutions (AG_lL , AG_hL , AT_lL , and AT_hL), the sugars clearly evidence preferential affinity (compared to water) for the bilayer surface, *i.e.* the sugar molecules cluster at the membrane surface, thereby reducing the water density in this region. However, it is important to stress that : (i) the water replacement is only partial and the membrane headgroups remain significantly hydrated; (ii) the reaching depth of the sugar molecules into the bilayer remains slightly smaller compared to that of water. In the simulations involving 0.8 m sugar solutions (AG_lL and AT_lL), the surface clustering entirely depletes the bulk from disaccharide molecules. This is no longer the case in the simulations involving 1.6 m sugar solutions (AG_hL and AT_hL), suggesting that the saturation of the bilayer surface by sugar molecules has been reached at this bulk concentration.

Experimentally¹⁴⁶, the maximal effect of disaccharides on lipid bilayers in terms of the inhibition of the liquid-crystalline to gel phase transition upon dehydration (which is expected to correlate with the threshold for surface saturation) corresponds to a sugar-to-lipid ratio of about 0.75. This is qualitatively consistent with the saturation observed in the present simulations (sugar-to-lipid ratios of 0.5 and 1.0 for the 0.8 and 1.6 m systems, respectively). It is important to mention that, according to the present simulations, after the surface of the bilayer reaches a state where it is completely saturated, the excess sugar will contribute only to the increase of bulk sugar concentration. This provides an explanation for the experimental results without the need of invoking the concept of sugar exclusion. At 325 K, the local concentrations of the disaccharides in the regions of the phosphate (C_{SP}) and glycerol (C_{SG}) groups are of the order of 0.3-0.4 and 0.06-0.13 nm⁻³, respectively. Upon increasing the bulk sugar concentration from 0.8 to 1.6 m, the former local concentration slightly decreases while the latter slightly increases (for both sugars), suggesting a somewhat deeper penetration of the sugar molecules into the bilayer.

However, the small relative magnitude of the change in C_{SP} (decrease by about 5-10% rather than twofold increase) confirms that the membrane surface is already essentially saturated in disaccharide molecules in the presence of a 0.8 m sugar solution. Interestingly, both C_{SP} and C_{SG} are systematically higher for TRH compared to GNT (at both bulk concentrations), indicating a denser coating by the former sugar. The local concentrations of WAT in the regions of the phosphate (C_{WP}) and glycerol (C_{WG}) groups are of the order of 6.5-8.1 and 3.4-4.2 nm⁻³, respectively. The corresponding concentrations in the absence of sugars are significantly higher (11.7 and 4.9 nm⁻³, respectively), indicating a partial replacement of the water molecules by sugar molecules at the bilayer surface (most pronounced in the phosphate region). Upon increasing the bulk sugar concentration from 0.8 to 1.6 m, the two local concentrations slightly increase for GNT or de-

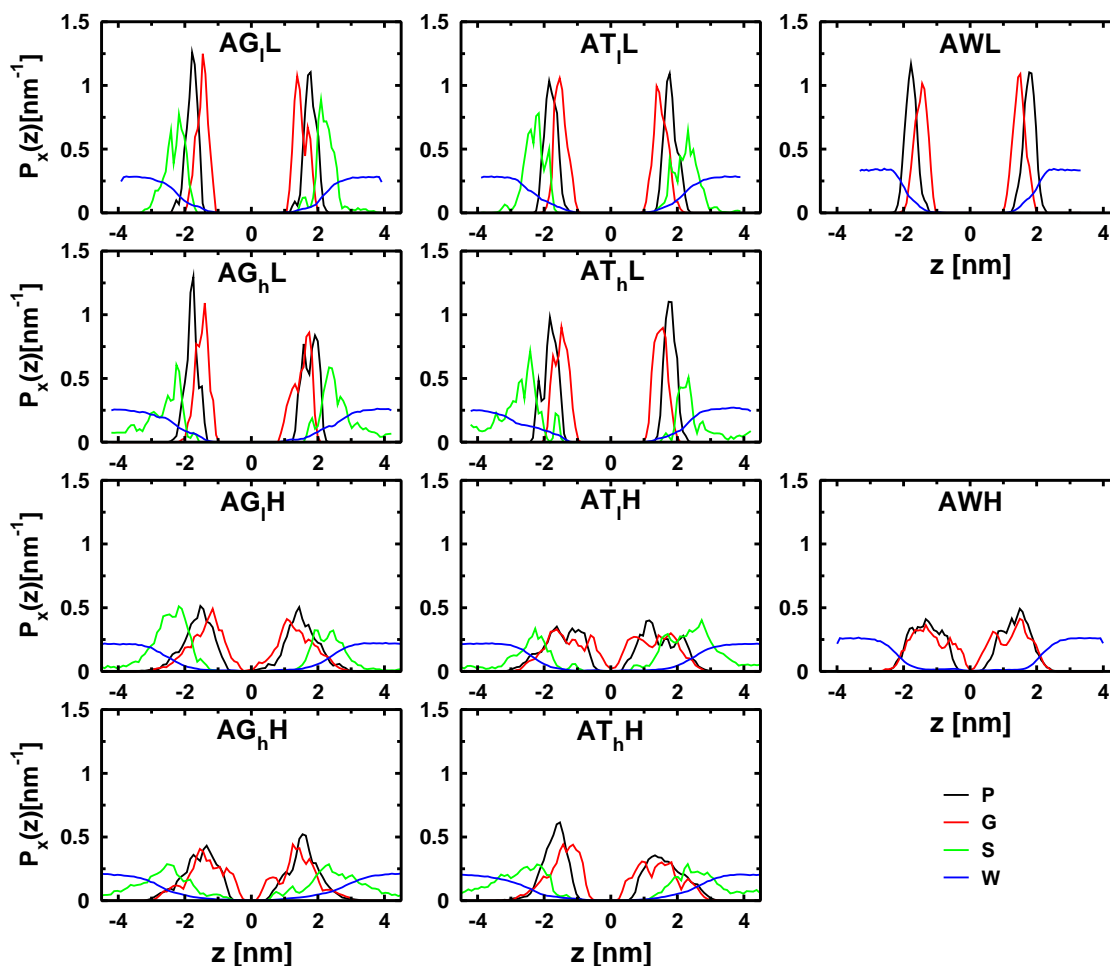


Figure 3.2: Normalized probability distributions $P_X(z)$ corresponding to specific DPPC, GNT or TRH, and WAT atoms (X) along the bilayer normal, as calculated from the ten simulations performed at constant bilayer area ($NA_{xy}P_zT$ ensemble). The different profiles correspond to the phosphorous atom ($X = P$) and the CH_2 glycerol united atom at the sn_1 position ($X = G$) of the lipid molecules (DPPC), the glycosidic oxygen atom ($X = S$) of the sugar molecules (GNT or TRH), and the oxygen atom ($X = W$) of the water molecules (WAT). The histogram bin width was set to $L_z/100$ and the profiles are centered at the bilayer midplane (as defined by the center of mass of all lipid atoms along the z -axis). The data was calculated considering the final 1 ns of each simulation. See Table 3.1 for the definition of the simulation codes.

crease for TRH (by about 7% in both cases). Here also, both C_{WP} and C_{WG} are systematically higher for TRH compared to GNT (at both bulk concentrations), suggesting a better extent of hydration of the bilayer in the presence of the former sugar, even though this sugar is also the one with the densest surface coating (see above). Finally, the effective sugar molalities in the regions of the phosphate (m_P) and glycerol (m_G) groups are of the order of 2.4-2.7 and 0.9-1.3 (0.8 m bulk concentration) or 1.6-1.7 (1.6 m bulk concentration) $\text{mol}\cdot\text{kg}^{-1}$ respectively. The effective sugar

| System | C_{SP} [nm ⁻³] | C_{SG} [nm ⁻³] | C_{WP} [nm ⁻³] | C_{WG} [nm ⁻³] | m_P [mol kg ⁻¹] | m_G [mol kg ⁻¹] |
|-------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|----------------------------------|----------------------------------|
| AWL | - | - | 11.65 | 4.90 | - | - |
| AWH | - | - | 2.71 | 2.66 | - | - |
| AG _l L | 0.315 | 0.057 | 6.53 | 3.40 | 2.68 | 0.93 |
| AT _l L | 0.400 | 0.096 | 8.11 | 4.21 | 2.74 | 1.27 |
| AG _h L | 0.299 | 0.113 | 6.95 | 3.87 | 2.39 | 1.62 |
| AT _h L | 0.354 | 0.127 | 7.52 | 4.09 | 2.61 | 1.72 |
| AWH | - | - | 2.71 | 2.66 | - | - |
| AG _l H | 0.211 | 0.152 | 3.57 | 2.86 | 3.28 | 2.96 |
| AT _l H | 0.195 | 0.171 | 3.68 | 2.99 | 2.94 | 3.18 |
| AG _h H | 0.343 | 0.290 | 3.58 | 3.02 | 5.32 | 5.33 |
| AT _h H | 0.347 | 0.251 | 4.31 | 3.45 | 4.47 | 4.04 |

Table 3.2: Peak overlaps C_{XY} (effective concentration of X in the region Y) derived from the normalized probability distributions of specific atoms (X, Y) along the z -axis (bilayer normal). The different values were calculated according to Eqs. 3.1 based on the distributions $P_X(z)$ and $P_Y(z)$ reported in Figure 3.2, considering the phosphorous atom ($Y = P$) or the CH_2 glycerol united atom ($Y = G$) at the sn_1 position of the lipid molecules (DPPC), along with the glycosidic oxygen atom ($X = S$) of the sugar molecules (GNT or TRH) or the oxygen atom ($X = W$) of the water molecules (WAT). The quantities m_P and m_G correspond to the ratios C_{SP}/C_{WP} and C_{SG}/C_{WG} expressed in terms of molalities.

molality in the phosphate region is significantly higher than in the bulk (preferential affinity) and essentially unaffected by the change in bulk sugar concentration (saturation). The effective sugar molality in the glycerol region is only slightly higher than in the bulk and its dependence on the bulk concentration suggests that saturation has not yet been reached in this region at the 0.8 m bulk concentration. Here also, the values of m_P and m_G are systematically higher for TRH compared to GNT.

At 475K all profiles present features characteristic of a largely perturbed bilayer structure (with the possible exception of simulation AT_hH ; see below). The peaks corresponding to the headgroup glycerol and phosphate atoms of the two bilayer leaflets are broad and no longer well separated, indicating that the headgroup and tail regions are no longer clearly distinguishable along the z -axis. Note, however, that a coarse structuring is still clearly visible and that the water molecules still penetrate the lipid volume no further than the region of the glycerol atoms. The effect of elevated temperature is probably somewhat limited in the present simulations because they are performed at constant lateral area and normal pressure ($NA_{xy}P_zT$ ensemble). As a result, an expansion of the box dimension is only possible (and observed) along the z -axis (by about 20% in pure water or 15-17% in the disaccharide solutions; Table 3.1), while the lateral expansion of the bilayer is prohibited. Previous simulations^{48,49} using the same force-field have shown, however, that the same temperature increase has a much more pronounced disruptive effect (in the absence of sugars) when the bilayer is also allowed to expand laterally (NPT ensemble). The simulation involving a TRH solution at 1.6 m bulk concentration (AT_hH) is clearly distinct from the other simulations in terms of structure preservation. Here, the peaks corresponding to the glycerol and phosphate atoms remain well separated (at distances of about 2.3 and 2.8 nm, respectively), so that the headgroup and tail regions can still be clearly recognized. The densities associated with sugar and water molecules also reach much less close to the midplane of the lipid volume. The discussion of the local concentrations of the disaccharides (C_{SP} and C_{SG}) or of WAT (C_{WP} and C_{WG}) in these high-temperature simulations is in most cases not very relevant in view of the significant disruption of the membrane. The exception is simulation AT_hH , where the bilayer structure is well preserved. Comparing this simulation with the corresponding simulation at low temperature (AT_hL), it is seen that C_{SP} is essentially unaltered, C_{SG} increased by about a factor two, C_{WP} decreased by about a factor two and C_{WG} slightly decreased. As a result of these changes, the local molalities m_P and m_G are larger by about 70 and 50 %, respectively. In other words, the extent surface coating by TRH has not changed significantly (although the depth of penetration of the sugar has increased), but the extent of water replacement (and membrane dehydration) has noticeably increased.

3.4.2 Order parameters

The order parameters $S_{CH}(C_n)$ corresponding to the 14 methylene groups C_n ($n = 2 \dots 15$) in the sn_1 and sn_2 chains of the DPPC molecules (calculated considering the final 4 ns of the ten simulations) are displayed in Figure 3.3. For the sn_2 chain, the calculated values are compared to the experimental carbon-deuterium order parameters (S_{CD}) measured¹³⁶ for DPPC in the liquid-crystalline phase and at full hydration at 317.15 K.

At 323K, the simulation in pure water (AWL) shows reasonable agreement with the experimental values for the sn_2 chain. The calculated values are slightly too low (less order; except for the methylene group C_2) in the present simulations at constant lateral area and normal pres-

sure ($NA_{xy}P_zT$ ensemble; A_{xy} based on the experimental¹²⁶ area per lipid of 0.64 nm^2), while the opposite is usually observed for the present lipid force field in simulations at constant pressure^{36,48,49,119} (NPT ensemble; average area per lipid at equilibrium of the order of 0.56 nm^2). The simulation results suggest comparable values for the order parameters of the two chains, except for the methylene group C_2 (significantly higher value for the sn_1 chain). Although experimental carbon-deuterium order parameters are (to our knowledge) not available for all atoms of the sn_1 chain, a higher value is indeed expected for the latter methylene group^{140,144}. When comparing the simulation in pure water (AWL) to the other simulations involving disaccharide solutions (AG_lL , AG_hL , AT_lL , and AT_hL), only limited differences are observed. At the two concentrations considered, both sugars noticeably increase the ordering of the sn_1 chain, the effect becoming decreasingly pronounced along the chain (*i.e.* towards the tail region). In contrast, they tend to rather slightly decrease the ordering of the sn_2 chain. Interestingly, the latter decrease is most pronounced towards the tail region for GNT and towards the headgroup region for TRH (at both sugar concentrations). However, the small magnitude of these changes suggests that the presence of the sugars does not drastically alter the ordering of the lipid chains. Qualitatively similar results were reported previously for sugar-membrane simulations performed at constant pressure^{48,49} (NPT ensemble).

At 475 K, the ordering of the chains decreases systematically for all systems compared to the corresponding simulations at 323 K. The effect is most pronounced for the bilayer in the presence of pure water, where the order parameters are very close to zero. At 0.8 m sugar concentration, the order parameters are higher compared to the case of pure water (especially for the sn_2 chain). The simulation involving GNT shows slightly more ordering compared to that involving TRH, but the difference remains limited. In contrast, at 1.6 m sugar concentration, the difference between the simulations involving GNT and TRH is much more pronounced. In this case, the simulation with TRH shows significantly more ordering compared to that with GNT. Accordingly, the comparison of the simulations involving a given sugar at the two different concentrations shows that the effect of increasing the sugar concentration is large for TRH, but essentially inexistent for GNT.

3.4.3 Bilayer structure

The final configurations (after 10 ns) of the ten simulations are displayed in Figure 3.4. A simplified graphical illustration of the average leaflet (headgroup) surfaces and individual headgroup trajectories (monitored over the last 0.4 ns of the simulations) is also provided in Figure 3.5.

At 323 K and irrespective of the solution environment, the bilayer presents a similar structure and is clearly found in the liquid-crystalline phase, in which the hydrophobic tails are flexible and partially disordered. At 0.8 m sugar concentration (AG_lL and AT_lL), most of the carbohydrate molecules interact with the bilayer headgroups and only few of them are totally exposed to the bulk solvent. At 1.6 m sugar concentration (AG_hL and AT_hL), the bilayer surfaces are more densely coated, and several sugar molecules are found to interact only with other sugar molecules or/and with the solvent (but not with the lipid headgroups). Considering the simplified representation of Figure 3.5, the average headgroup surfaces are found to be essentially planar and parallel for the two leaflets, the lateral headgroup diffusion is slow (no significant displacement within the 0.4 ns period considered), and the corresponding longitudinal displacements (away from the average planes) are very small.

At 475K and irrespective of the solution environment, the bilayer structure is found to be

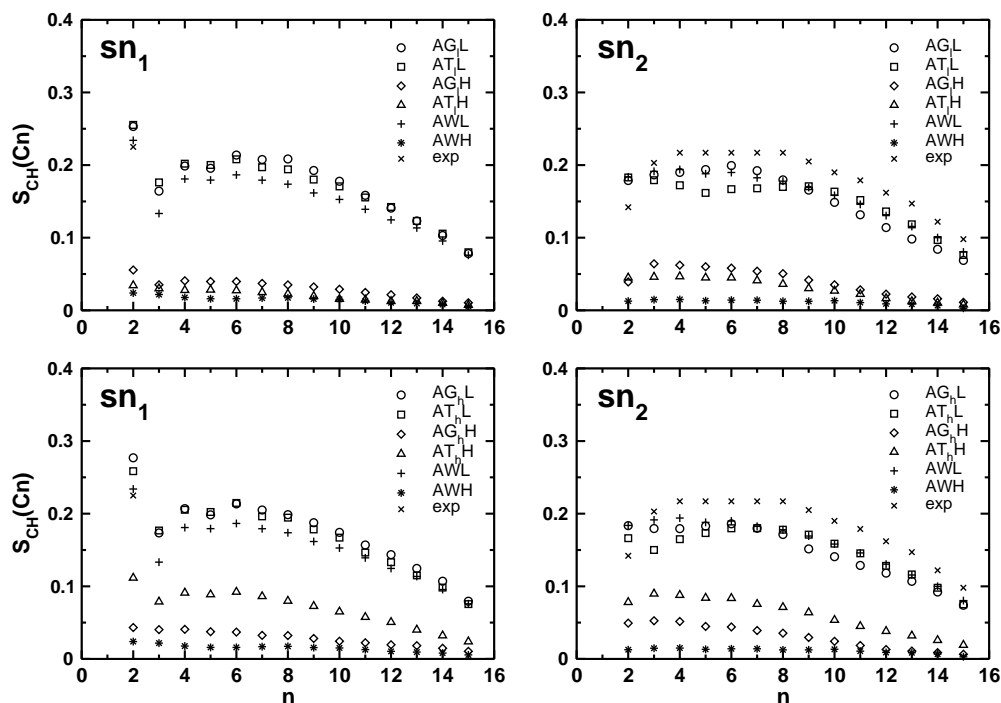


Figure 3.3: Order parameters corresponding to the 14 methylene groups in the sn_1 (left) and sn_2 (right) chains of the DPPC molecules, as calculated from the ten simulations performed at constant bilayer area ($NA_{xy}P_zT$ ensemble). The sn_2 chain is the one connected to the glycerol CH_1 group (i.e. the chain vicinal to the phosphate group), while the sn_1 chain is the one located on the second CH_2 group (i.e. most distant from the phosphate group). The experimental carbon-deuterium order parameters S_{CD} for the sn_2 chain measured¹²⁶ in the liquid-crystalline phase at full hydration and at 317.15 K are also reported for comparison. A single experimental value available for the methylene group C_2 of the sn_1 chain¹⁴⁰ is also reported. The data was calculated (Section 3.3.3) based on the final 4 ns of each simulation. See Table 3.1 for the simulations codes.

largely perturbed as a result of the increased thermal motion (with the possible exception of simulation AT_hH ; see below). The largest perturbation is observed for the simulation in pure water (AWL), which evidences a nearly complete randomization of the tail orientations and a tendency towards the clustering of the headgroups. At 0.8 m sugar concentration (AG_7H and AT_7H), the perturbation is still very strong, although a bilayer-like structure (tail region surrounded by two sugar-coated headgroup layers) can now be identified. There is, however, no obvious difference between the extent of structure preservation by GNT or TRH. At 1.6 m sugar concentration (AG_hH and AT_hH), the results for GNT are similar. However, the simulation in the presence of the TRH solution evidences a much lower extent of disorder, and the structure clearly remains closest to that of a bilayer in the liquid-crystalline phase. Comparison with *e.g.* the final structure of the corresponding simulation at 323 K (AT_hL) shows that the headgroups still arrange in two well defined surfaces (although deviating more significantly from planarity), that the lipid tails are preferentially ordered along the bilayer normal (although with a lower extent of ordering) and

that the lipid tails from the two leaflets do not interdigitate significantly (even close to the bilayer midplane). For TRH, qualitatively similar results were reported previously for sugar-membrane simulations performed at constant pressure^{48,49}, at the same temperatures and at 1 or 2 m sugar concentration (*NPT* ensemble). In particular, in these independent simulations (using the same force-field) TRH also only promoted a significant extent of structure preservation at high temperature for the highest bulk concentration considered. Considering the simplified representation of Figure 3.5, the average headgroup surfaces are found to deviate significantly from planarity, the headgroup lateral diffusion is fast (significant displacements within the 0.4 ns period considered), and the corresponding longitudinal displacements (away from the average planes) are large. Here also, the simulation in the presence of TRH at 1.6 m concentration appears closest to the low-temperature case (average leaflet surfaces closest to planarity, more homogeneous lateral headgroup distribution, more limited headgroup longitudinal displacements).

3.4.4 Hydrogen-bonding

The occurrences of hydrogen bonds (average over the last 4 ns of the simulations) between all the species present in the different systems are shown in Table 3.3 for the ten simulations.

At 323 K, both sugars are found to form a significant number of H-bonds to the lipid headgroups, with an average of about 2.0-2.1 (0.8 m solutions) or 1.1-1.2 (1.6 m solutions) H-bonds per sugar molecule. The observation that this number decreases by about a factor of two upon doubling the bulk sugar concentration is again indicative that the saturation of the bilayer surface in sugar molecules has already been reached at the lowest bulk concentration considered. The number of sugar-headgroup H-bonds is slightly lower for TRH compared to GNT (at both bulk concentrations), although the difference is small (about 2-3, *i.e.* 3-4%). On the other hand, the number of headgroup-water H-bonds is significantly higher in the TRH solutions compared to the GNT solutions (at identical bulk concentration), the difference being larger (about 10-30, *i.e.* 5-15%). Consistent with previous studies,^{36,48-50} the total number of H-bonds between the lipid headgroups and their solution environment (sugar and water molecules) is only weakly affected by the possible presence (and concentration) of the cosolutes (suggesting a mere exchange of H-bonding partners for the lipid headgroups, the total number of H-bonding sites provided by the membrane remaining about constant). This observation is accurately valid in the case of TRH (AWL *vs* AT_lL *vs* AT_hL), but a decrease (by about 25-30, *i.e.* 7-9%) in this total number of H-bonds (relative to the situation of the bilayer in pure water) is apparent for GNT (AWL *vs* AG_lL *vs* AG_hL). Finally, a systematic difference between GNT and TRH is also observed in the number of sugar-sugar and sugar-water H-bonds (at identical sugar concentration). The former number is significantly lower for GNT (by about 15-20, *i.e.* 8-38%), while the latter is significantly lower for TRH (by about 15-20, *i.e.* 2-5%).

At 475 K, the number of sugar-headgroup H-bonds is reduced (by about 15-20, *i.e.* 22-25%) for the systems at 0.8 m sugar concentration and increased (GNT; by about 20, *i.e.* 23%) or unaltered (TRH) for the systems at 1.6 m sugar concentration (compared to the corresponding simulations at 323 K). In all cases, the number of headgroup-water H-bonds is dramatically reduced (by a factor 1.6-1.8) upon increasing the temperature. In spite of these changes, the number of sugar-headgroup H-bonds is still lower for TRH compared to GNT (by about 5 and 20 at 0.8 and 1.6 m concentration, *i.e.* 9 and 20%, respectively) while the number of headgroup-water H-bonds is still higher in the TRH solutions compared to the GNT solutions (by about 5 and 20 at 0.8 and

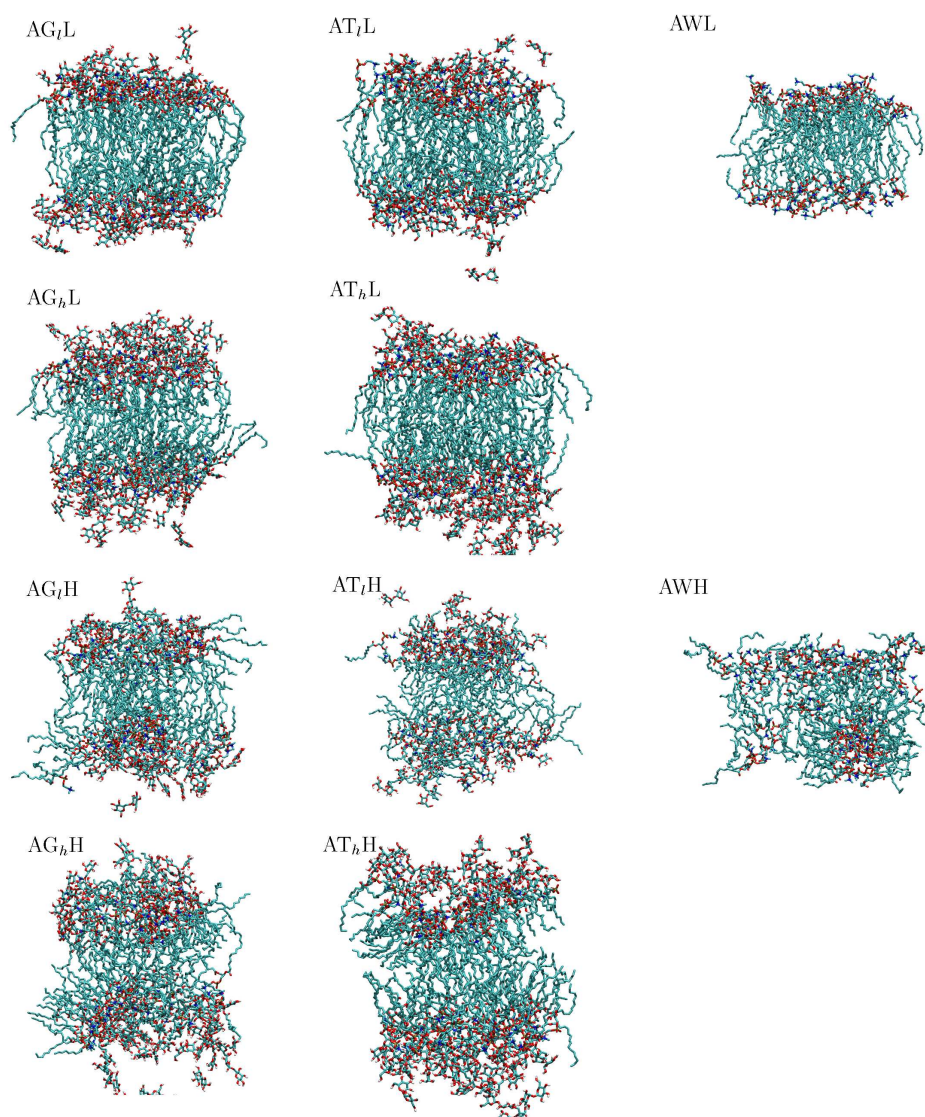


Figure 3.4: Final configurations of the simulations performed at constant bilayer area ($NA_{xy}P_zT$ ensemble). The configurations correspond to the end of the 10 ns simulations. DPPC and sugar molecules are represented using a stick model. WAT molecules have been removed for clarity. Carbon atoms are represented in gray, oxygen atoms in red, phosphorus atoms in yellow, nitrogen atoms in blue and hydrogen atoms in white. See Table 3.1 for the definition of the simulation codes.

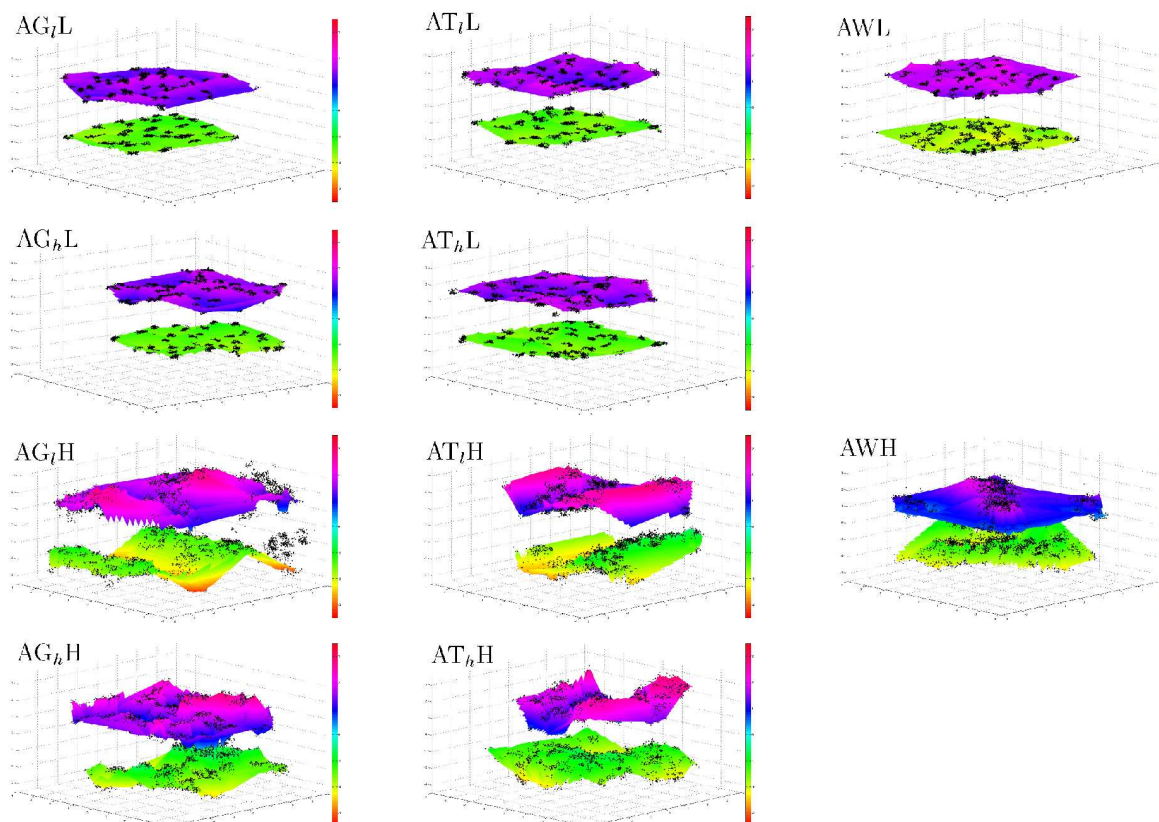


Figure 3.5: Leaflet average surfaces and headgroup trajectories corresponding to the simulations performed at constant bilayer area ($NA_{xy}P_zT$ ensemble). The average leaflet surfaces are fitted to the average positions of the phosphorus atoms in the DPPC headgroups, and colored according to the distance from the bilayer midplane. The headgroup trajectories are represented by individual dots corresponding to successive instantaneous positions of the headgroup phosphorous atoms sampled at 0.01 ns intervals. The bilayer midplane (average of the coordinates of all bilayer atoms along the z -axis) is placed at $z = 0$ in all drawings. The figures are generated based on the final 0.4 ns of the trajectories. See Table 3.1 for the definition of the simulation codes.

| System | n_{LS} | n_{LW} | $n_{LS} + n_{LW}$ | n_{SS} | n_{SW} |
|-------------------|----------|----------|-------------------|----------|----------|
| AWL | - | 336.6 | 336.6 | - | - |
| AG _l L | 74.7 | 231.2 | 305.9 | 35.4 | 373.7 |
| AT _l L | 71.6 | 263.5 | 335.1 | 56.9 | 353.7 |
| AG _h L | 83.4 | 228.8 | 312.2 | 146.3 | 774.4 |
| AT _h L | 81.1 | 240.6 | 321.7 | 159.4 | 760.4 |
| AWH | - | 163.8 | 163.8 | - | - |
| AG _l H | 58.2 | 144.1 | 202.3 | 31.4 | 257.2 |
| AT _l H | 53.6 | 147.2 | 200.8 | 49.4 | 245.6 |
| AG _h H | 102.3 | 125.7 | 228.0 | 94.3 | 478.8 |
| AT _h H | 81.0 | 148.2 | 229.2 | 119.9 | 515.3 |

Table 3.3: Occurrences of intermolecular H-bonds between the different species as calculated from the ten simulations performed at constant bilayer area ($NA_{xy}P_zT$ ensemble). The occurrences n_{xy} are reported in the form of average number of H-bonds of the given type in the system, distinguishing DPPC lipid (L), sugar (S) and water (W) molecules. The data was calculated (Section 3.3.3) considering the final 4 ns of each simulation. See Table 3.1 for the definition of the simulation codes.

number is also higher for TRH by about 35, *i.e.* 8%).

The (absolute and relative) contributions of H-bonds involving the phosphate or the ester groups to the total number of sugar-lipid H-bonds are reported in the Table 3.4.

| System | n_{LS} | $n_{LS}(P)$ | $n_{LS}(E)$ | $n_{LS}(P)/n_{LS}$ [%] | $n_{LS}(E)/n_{LS}$ [%] |
|-------------------|----------|-------------|-------------|------------------------|------------------------|
| AG _l L | 74.7 | 45.1 | 29.6 | 60.4 | 39.6 |
| AT _l L | 71.6 | 49.3 | 22.3 | 68.9 | 31.1 |
| AG _h L | 83.4 | 55.5 | 28.0 | 66.5 | 33.5 |
| AT _h L | 81.1 | 58.3 | 22.8 | 71.9 | 28.1 |
| AG _l H | 58.2 | 30.9 | 27.3 | 53.1 | 46.9 |
| AT _l H | 53.6 | 29.5 | 24.1 | 55.0 | 45.0 |
| AG _h H | 102.3 | 48.4 | 53.9 | 47.3 | 52.7 |
| AT _h H | 81.0 | 49.5 | 31.5 | 61.0 | 39.0 |

Table 3.4: Occurrences and relative contributions of H-bonds involving the phosphate ($n_{LS}(P)$) or the ester ($n_{LS}(E)$) groups to the total number of H-bonds formed between sugar and lipid molecules (n_{LS} in Table 3.3). The data was calculated (Section 3.3.3) considering the final 4 ns of each simulation. See Table 3.1 for the definition of the simulation codes.

At 323 K and irrespective of the sugar concentration, both sugars form a lower number of H-bonds with the ester groups (30-40%) compared to the phosphate groups of the lipids. However, in addition to the slightly higher total number of sugar-lipid H-bonds (see above), the systems with GNT also evidence a higher proportion of H-bonds to the ester groups compared to those with TRH, suggesting a somewhat deeper penetration of GNT inside the membrane.

At 475 K, the proportion of H-bonds to the ester groups (40-50%) is significantly increased compared to 323 K. This change is probably related in large part to the perturbation of the bilayer structure, leading to an increased exposure of the (usually more buried) ester groups to the solution environment. For the 0.8 m solutions, the decrease observed in the total number of sugar-headgroup H-bonds (see above) nearly exclusively involves H-bonds to the phosphate groups. For the 1.6 solutions, the corresponding increase observed for GNT (see above) nearly exclusively involves H-bonds to the ester groups, while for TRH, the absence of net change (see above) still hides an increase in the proportion of H-bonds to the ester groups.

The distribution of the sugar molecules along different H-bonding patterns and associated degrees of bridging are reported in Table 3.5. At 323 K, most of the sugar molecules that are H-bonded to the lipids belong to the patterns 1, 11, 111, 2, 21 and 211, *i.e.* are (possibly multiple times) H-bonded to 1, 2 or 3 lipid molecules, although many other patterns (involving the bridging of up to 6 lipid molecules) do occur in the simulations. Upon increasing the sugar concentration, most of the additional sugar molecules do not bind to the lipids (saturation), although the numbers of sugar molecules H-bonded to 1 or 2 distinct lipid headgroups still increases significantly. The comparison of GNT and TRH solutions (at both bulk concentrations) does not reveal clear systematic differences in terms of H-bonding patterns. At 475 K, the same H-bonding patterns

are significantly populated. Comparison with the simulations at 323 K suggests a trend towards a decrease in the number of sugar molecules H-bonded to 2 or 3 distinct lipid molecules (except for simulation AG_hL). Upon increasing the sugar concentration, most of the additional sugar molecules again do not bind to the lipids, although the numbers of sugar molecules H-bonded to 1, 2 or 3 distinct lipid headgroups noticeably increases.

Finally, the increase in the number of sugar-lipid H-bonds for GNT at 1.6 m concentration (AG_hL vs AG_hH) is predominantly associated with the patterns 2, 21, 211 and 2111. For TRH, the above observations contrast again with the results of previous sugar-membrane simulations performed at constant pressure^{48,49} (NPT ensemble). In these simulations, the temperature increase was accompanied by a nearly systematic enhancement of all patterns involving the bridging of multiple lipid molecules. Here also, the difference certainly results from the absence of lateral expansion of the membrane in the present simulations (which prevents the intercalation of TRH molecules between the lipid headgroups). In this respect, it is interesting to observe that this intercalation occurs to some extent for GNT at 1.6 m concentration, but is responsible in the present case for a destabilization (rather than stabilization) of the bilayer structure.

3.5 Conclusion

In the present work, explicit-solvent MD simulations of a dipalmitoyl-phosphatidylcholine (DPPC) bilayer patch in the presence of pure water or of aqueous solutions of the two disaccharides trehalose (TRH) and gentiobiose (GNT) were reported and compared, in order to investigate the effect of the difference in flexibility between the two sugars on their interaction with the bilayer (and the resulting alteration of the bilayer properties), and on their potential for bioprotection. To this purpose, a total of ten simulations (10 ns each) have been carried out (at constant lateral area and normal pressure), differing in the disaccharide identity and concentration (pure water, GNT or TRH at either 0.8 or 1.6 m concentration, “m” indicating the solution molality in $\text{mol}\cdot\text{kg}^{-1}$) and temperature (323 or 475 K).

When comparing the influence of the two disaccharides on the bilayer, it is important to emphasize the very close similarity between the two compounds. The two sugars are constitutional isomers (*i.e.* they share an identical atom content). They possess the same total number of hydroxyl or hydroxymethyl groups (8), all oriented equatorially around the two pyranose rings (with the slight difference that TRH has two hydroxymethyl groups compared to one only for GNT) and the same number of ring (2) and glycosidic (1) oxygen atoms. Furthermore, the non-bonded interaction parameters (atomic partial charges and van der Waals parameters) within the 45A4 force field¹¹⁶ are very similarly distributed in the two compounds (so that significant cancellation of possible force field errors can be expected). As a result, the only significant difference resides in the presence of an extra atom (*i.e.* an extra dihedral angle) within the glycosidic linkage of GNT compared to TRH. As discussed in Section 3.2, however, this difference has a considerable impact on the flexibility and conformational entropy of the molecule (also considering that neither of the two molecules presents persistent intramolecular H-bonds in aqueous solution). The differences observed in the present simulations concerning the influence of TRH and GNT on a lipid bilayer are thus particularly interesting, because they can be predominantly attributed to this difference in intrinsic flexibility.

The main observations resulting from the analysis of the ten simulations (Section 3.4) are sum-

| pattern | AG _l L | AT _l L | AG _h L | AT _h L | AG _l H | AT _l H | AG _h H | AT _h H |
|--------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| 0 | 7.1 | 6.4 | 34.4 | 34.6 | 11.8 | 12.8 | 31.7 | 35.50 |
| 1 | 7.0 | 8.8 | 13.3 | 13.5 | 6.2 | 7.4 | 10.5 | 13.0 |
| 11 | 6.0 | 5.6 | 7.7 | 8.6 | 3.4 | 4.2 | 4.6 | 6.7 |
| 111 | 2.7 | 3.6 | 2.4 | 2.3 | 1.2 | 1.7 | 1.9 | 2.5 |
| 1111 | 0.2 | 0.4 | 0.1 | 0.2 | 0.4 | 0.5 | 0.6 | 0.5 |
| 11111 | 0.1 | - | 0.0 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| 2 | 2.6 | 2.5 | 4.4 | 3.3 | 4.9 | 2.8 | 7.4 | 4.1 |
| 21 | 4.6 | 2.6 | 4.8 | 4.5 | 3.9 | 2.6 | 6.0 | 4.0 |
| 211 | 2.7 | 2.2 | 1.8 | 2.7 | 1.9 | 1.7 | 4.1 | 2.4 |
| 2111 | 0.6 | 0.8 | 0.3 | 0.7 | 0.7 | 0.6 | 1.9 | 0.8 |
| 21111 | 0.3 | 0.1 | 0.1 | 0.2 | 0.1 | 0.1 | 0.4 | 0.1 |
| 211111 | 0.2 | - | 0.0 | - | 0.0 | - | 0.0 | 0.0 |
| 22 | 0.1 | 0.0 | 0.3 | 0.2 | 0.3 | 0.1 | 0.3 | 0.2 |
| 221 | 0.1 | 0.1 | 1.0 | 0.4 | 0.2 | 0.2 | 0.3 | 0.3 |
| 2211 | 0.1 | 0.2 | 0.1 | 0.1 | 0.1 | 0.1 | 0.3 | 0.1 |
| 22111 | 0.4 | 0.0 | 0.4 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 |
| 221111 | 0.2 | - | - | - | 0.0 | - | 0.0 | - |
| 3 | 0.2 | 0.5 | 0.4 | 0.1 | 0.4 | 0.4 | 0.6 | 0.6 |
| 31 | 0.4 | 1.4 | 0.2 | 0.2 | 0.3 | 0.4 | 0.5 | 0.5 |
| 311 | 0.3 | 0.4 | 0.1 | 0.4 | 0.1 | 0.2 | 0.3 | 0.3 |
| 3111 | 0.1 | 0.1 | - | 0.1 | 0.0 | 0.1 | 0.1 | 0.1 |
| 31111 | 0.0 | 0.1 | - | 0.0 | 0.0 | - | 0.0 | 0.0 |
| 32 | 0.0 | - | 0.0 | - | 0.0 | 0.0 | 0.0 | 0.0 |
| 321 | - | - | - | - | 0.0 | 0.0 | 0.0 | 0.0 |
| 3211 | - | - | - | - | 0.0 | 0.0 | 0.0 | 0.0 |
| 4 | 0.0 | 0.0 | 0.0 | - | 0.0 | 0.0 | 0.0 | 0.1 |
| 41 | 0.0 | 0.0 | - | - | 0.0 | 0.0 | 0.0 | 0.1 |
| total | 36 | 36 | 72 | 72 | 36 | 36 | 72 | 72 |
| degree of bridging | | | | | | | | |
| 0 | 7.1 | 6.4 | 34.3 | 34.6 | 11.8 | 12.8 | 31.7 | 35.5 |
| 1 | 9.8 | 11.8 | 18.0 | 16.9 | 11.5 | 10.7 | 18.6 | 17.7 |
| 2 | 11.0 | 9.7 | 13.1 | 13.5 | 7.8 | 7.3 | 11.3 | 11.6 |
| 3 | 5.9 | 6.3 | 5.3 | 5.8 | 3.5 | 3.9 | 6.7 | 5.5 |
| 4 | 1.0 | 1.5 | 0.5 | 1.1 | 1.2 | 1.2 | 1.2 | 1.5 |
| other | 1.3 | 0.3 | 0.9 | 0.3 | 0.3 | 0.2 | 0.7 | 0.3 |
| total | 36 | 36 | 72 | 72 | 36 | 36 | 72 | 72 |

Table 3.5: Average numbers of sugar molecules forming a specific H-bonding pattern (top) and characterized by a specific degree of bridging (bottom) with lipid molecules as calculated from the ten simulations performed at constant bilayer area ($NA_{xy}P_zT$ ensemble). A pattern is encoded by a series of integers (listed in decreasing order) indicating the number of H-bonds formed between a sugar molecule and a number of distinct DPPC molecules in a given system configuration. The degree of bridging (number of integers in the pattern code), represents the number of distinct DPPC molecules forming at least one H-bond with the sugar molecule. The data was calculated (Section 3.3.3) considering the final 4 ns of each simulation. See Table 3.1 for the definition of the simulation codes.

marized qualitatively in Table 3.6. According to the sugar-like mechanism proposed by Pereira

and Hünenberger³⁶, both sugars should evidence preferential affinity (compared to water) for the membrane surface owing to the large entropy gain resulting from the release of water molecules from the surface region to the bulk, offset by a limited entropy penalty for immobilizing and rigidifying fewer (already inherently rigid) sugar molecules. However, the total number of H-bonds between the bilayer and its solution environment should remain approximately constant during this process (the lipid headgroups undergoing a mere exchange of H-bonding partners), leading to a negligible enthalpic change. According to this mechanism, TRH, due to its lower conformational entropy (reduced entropic penalty of rigidification), is expected to present a stronger preferential affinity compared to GNT. In agreement with these ideas, the present results at 323 K support a preferential affinity of the sugars for the bilayer surfaces, a significant extent of sugar-lipid H-bonding, an essentially identical number of H-bonds between the lipid headgroups and their solution environment irrespective of the nature and concentration of the sugar, and a stronger preferential affinity for TRH compared to GNT (as judged by the local sugar molalities in the headgroup region). However, these simulations also reveal three other important differences between the two sugars : (i) a higher extent of residual bilayer hydration for TRH compared to GNT (Tables 3.2 and 3.3); (ii) a higher extent of sugar-sugar H-bonding (sugar clustering) for TRH compared to GNT (Table 3.3); (iii) a more limited extent of sugar-lipid H-bonding for TRH compared to GNT (Table 3.3); (iv) a more limited depth of penetration into the bilayer (H-bonds to the ester *vs.* phosphate groups) for TRH compared to GNT (Table 3.4). These observations may also be indirectly related to the particular conformational properties of the TRH molecule (high propensity to self-association, strong hydrophilicity, lack of flexibility). Finally, both sugars present a significant extent of H-bonded bridging of the lipid molecules.

The simulations at 475 K clearly show that at sufficiently high bulk concentration, TRH is indeed superior to GNT in preserving the integrity of the bilayer structure. It seems therefore logical to correlate the observed (and experimentally supported) higher bioprotective ability of TRH with one or more of the above-mentioned features. However, it appears that these features encompass elements of at least four of the hypotheses formulated in Section 3.2 for the bioprotective action of sugars. The partial replacement of lipid-water with sugar-lipid H-bonds is compatible with the WRH hypothesis. However, the idea that such a replacement is a key element of the bioprotection mechanism seems contradicted by the observation that the extent of sugar-lipid H-bonding is higher for GNT compared to TRH in the present simulations (at both bulk concentrations and temperatures considered). The presence of a significant extent of H-bonded bridging of the lipid headgroups by sugar molecules is compatible with the HBH hypothesis. However, the idea that such a replacement is a key element of the bioprotection mechanism is also somewhat at odds with the observation that the extent of bridging is about identical for GNT compared to TRH at 323 K, and even higher at 475 K. The more significant extent of residual hydration of the bilayer surface for TRH compared to GNT is also clearly reminiscent of the WEH hypothesis. Finally, the higher extent of sugar clustering for TRH compared to GNT suggests an increased viscosity, closely related to the VIH hypothesis. On the other hand, the results of the present simulations (as well as of a number of previous simulation studies) definitely support preferential affinity (rather than preferential exclusion) of the sugar molecules for the bilayer surface, in disagreement with the HFH hypothesis.

The compatibility of the present (and previous experimental^{5, 7, 9, 26–33, 37–47, 51–65, 65–67} and computational^{48–50, 69–78}) observations with the WRH, HBH, VIH and WEH hypotheses appears somewhat surprising at first sight. A first possible interpretation is that multiple mechanisms are

| Property | AWL | AG _l L | AT _l L | AG _h L | AT _h L | AWH | AG _l H | AT _l H | AG _h H | AT _h H |
|-----------------------|-----|-------------------|-------------------|-------------------|-------------------|-----|-------------------|-------------------|-------------------|-------------------|
| Bilayer structure | ++ | ++ | ++ | ++ | ++ | -- | - | - | - | + |
| Order parameters | +++ | +++ | +++ | +++ | +++ | -- | - | - | + | ++ |
| Sugar clustering | o | - | + | ++(+) | +++ | o | - | + | ++(+) | ++ |
| Preferential affinity | o | ++ | +++ | ++ | +++ | o | / | / | / | / |
| Sugar-lipid H-bonds | o | ++(+) | ++ | ++(+) | ++ | o | + | + | +++ | ++ |
| Sugar-ester H-bonds | o | + | -- | - | -- | o | ++ | ++(+) | ++(+) | + |
| Lipid-water H-bonds | +++ | + | ++(+) | ++(+) | ++ | - | - | (+) | -- | + |
| Lipid bridging | o | ++ | ++ | ++ | ++ | o | + | + | ++(+) | + |

Table 3.6: *Qualitative summary of the main observations made in the present study based on the ten simulations performed. Bilayer structure refers to Sections 3.4.1 and 3.4.3, Figures 3.2, 3.4 and 3.5 and Table 3.2; order parameters refers to Section 3.4.2 and Figure 3.3; sugar coating/clustering refers to Sections 3.4.1 and 3.4.4, Tables 3.2, 3.3 and Figure 3.2; preferential affinity refers to Section 3.4.1, Figure 3.2 and Table 3.2; sugar-lipid H-bonds refers to Section 3.4.4, Tables 3.3 and 3.4, sugar-ester H-bond fraction refers to Section 3.4.4 and Table 3.4; hydration/lipid-water H-bonds refers to Section 3.4.4 and Table 3.3; H-bonded lipid bridging refers to Section 3.4.4 and Table 3.5. See Table 3.1 for the definition of the simulations codes.*

simultaneously active. A second one is that multiple mechanisms are potentially active, their relevance depending on the nature and intensity of the environmental stress as well as on the type of protected biostructure. A third one is that some of the observed features are mere consequences of the others, and do not bear relevance as driving forces for bioprotection. And a fourth one is simply that the proposed mechanisms are not defined precisely enough for an unambiguous assessment based on experimental or theoretical data. We believe that the fourth point is actually very relevant : although the five mechanisms represent valid qualitative (pictorial) interpretations, they clearly lack a clear quantitative (measurable) definition. In particular, it seems essential to include into the formulation of each hypothesis : (i) a clear mention of the type of biomolecule considered; (ii) a clear specification of the experimental conditions (temperature, pressure, concentration regime); (iii) a quantitative definition for intuitive concepts such as “preferential affinity”, “water replacement”, “headgroup bridging” or “water entrapment”. The suggestion of revised definitions for these hypotheses, along with an overview of the experimental and theoretical data supporting each of them, will be the scope of forthcoming work.

Acknowledgments

The author would like to thank Bruno Horta for his active participation in the present project, Cristina Pereira and Maria Reif for fruitful discussions, and Anna-Pitschna Kunz for her help with Matlab.

3.6 References

- [1] Feovilova, E.P. *Appl. Biochem. Microbiol.* **39** 1-18 (2003)
- [2] Zhmakin, A.I. *Physics-Uspekhi* **51** 231-252 (2008)
- [3] Keilin, D. *Proc. R. Soc. Lond. B Biol. Sci.* **150** 149-191 (1959)
- [4] Clegg, J.S. *Comp. Biochem. Physiol.* **128B** 613-624 (2001)
- [5] Crowe, J.H., Hoekstra, F.A. & Crowe, L.M. *Annu. Rev. Physiol.* **54** 579-599 (1992)
- [6] Guppy, M. & Withers, P. *Biol. Rev.* **74** 1-40 (1999)
- [7] Crowe, J.H., Crowe, L.M., Wolkers, W.F., Oliver, A.E., Ma, X., Auh, J.-H., Tang, M., Zhu, S., Norris, J. & Tablin, F. *Integr. Comp. Biol.* **45** 810-820 (2005)
- [8] Hengherr, S., Heyer, A.G., Koehler, H., *FEBS J.* **275** 281-288 (2008)
- [9] Crowe, L.M. *Comp. Biochem. Physiol.* **131A** 505-513 (2002)
- [10] Crowe, J.H., Oliver, A.E. & Tablin, F. *Integr. Comp. Biol.* **42** 497-503 (2002)
- [11] Crowe, J.H. Trehalose as a "Chemical chaperone": Fact and fantasy In: *Molecular Aspects of the stress response: Chaperones, membranes and networks Volume 13*. Csermely, P. & Vigh, L., Eds.; Landes Bioscience and Springer Science+ Business Media pp 143-158 (2007)
- [12] Guo, N., Puhlev, I., Brown, D.R., Mansbridge, J. & Levine, F. *Nature Biotechnology* **18** 168-171 (2000)
- [13] Valluru, R. & Van den Ende, W. *J. Exp. Botan.* **59** 2905-2916 (2008)
- [14] Valluru, R., Lammens, W., Claupein, W. & van den Ende, W. *Trends Plant. Sci.* **8** 409-414 (2008)
- [15] Van den Ende, W. & Valluru, R. *J. Exp. Botan.* **60** 9-18 (2008)
- [16] Fernandes, P.M.B. *J. Med. Biol. Res.* **38** 1239-1245 (2005)
- [17] Iwahashi, H. et al. *Appl. Environ. Microbiol.* **64** 4614-4617 (1998)
- [18] Purvis, J. E., Yomano, L.P. & Ingram, L.O. *Appl. Environ. Microbiol.* **71** 3761-3769 (2005)
- [19] Ragoonan, V. & Aksan, A. *Biophys. J.* **94** 2212-2227 (2008)
- [20] D'Alfonso, L. et al. *Biophys. J.* **93** 284-293 (2007)
- [21] Attanasio, F. *Biochem. Biophys. Res. Comm.* **354** 899-905 (2007)
- [22] Baptista, R.P. *Biopolymers* **89** 538-547 (2008)

- [23] Best, R.B., Buchete, N.-V. & Hummer, G. *Biophys. J.* **95** L07-L09 (2008)
- [24] Bellavia, G., Cordone, L. & Cupane, A. *J. Therm. Anal. Calorim.* **95** 699-702 (2009)
- [25] Jain, N.K. & Roy, I. *Protein Sci.* **18** 24-36 (2009)
- [26] Crowe, J.H., Crowe, L.M., Carpenter, J.F. & Wistrom, C.A. *Biochem. J.* **242** 1-10 (1987)
- [27] Crowe, J.H., Crowe, L.M., Oliver, A.E., Tsvetkova, N., Wolkers, W. & Tablin, F. *Cryobiology* **43** 89-105 (2001)
- [28] Cacela, C. & Hinch, D.K. *Biophys. Biochim. Acta* **1758** 680-691 (2006)
- [29] Cacela, C. & Hinch, D.K. *Biophys. J.* **90** 2831-2842 (2006)
- [30] Holovati, J.L. & Acker, J. P. *Cryobiology* **55** 98-107 (2007)
- [31] Ohtake, S., Schebor, C., Palecek, S.P. & de Pablo, J.J. *Cryobiology* **48** 81-89 (2004)
- [32] Ohtake, S., Schebor, C. & de Pablo, J.J. *Biochem. Biophys. Acta* **1758** 65-73 (2006)
- [33] Oliver, A.E. et al. *Lab on a Chip.* **8** 892-897 (2008)
- [34] Zhu, B. Furuki, T., Okuda, T., Sakurai, J. *Phys. Chem.* **111** 5542-5544 (2007)
- [35] Santivarangkna, C., Higl, B. & Foerst, P. *Food Microbiol.* **25** 429-441 (2008)
- [36] Pereira, C.S. & Hünenberger, P.H. *Molecular Simulation* **4** 403-420 (2008)
- [37] Crowe, J.H., Crowe, L.M. & Chapman, D. *Science* **223** 701-703 (1984)
- [38] Lee, C.W.B., Waugh, J.S. & Griffin, R.G. *Biochemistry* **25** 3737-3742 (1986)
- [39] Nakagaki, M., Nagase, H. & Ueda, H. *J. Membr. Sci.* **73** 173-180 (1992)
- [40] Tsvetkova, N.M., Phillips, B.L., Crowe, L.M., Crowe, J.H. & Risbud, S.H. *Biophys. J.* **75** 2947-2955 (1998)
- [41] Crowe, J.H., Carpenter, J.F. & Crowe, L.M. *Annu. Rev. Physiol.* **60** 73-103 (1998)
- [42] Luzardo, M.C., Amalfa, F., Nunez, A.M., Diaz, S., Biondii de Lopez, A.C. & Disalvo, E.A. *Biophys. J.* **78** 2452-2458 (2000)
- [43] Lambruschini, C., Relini, A., Ridi, A., Cordone, L. & Gliozzi, A. *Langmuir* **16** 5467-5470 (2000)
- [44] Hoekstra, F.A., Golovina, E.A. & Buitink, J. *J. Trends Plant Sci* **6** 431- (2001)
- [45] Hinch, D.K., Zuther, E., Hellwege, E.M. & Heyer, A.G. *Glycobiology* **12** 103-110 (2002)
- [46] Ricker, J., Tsvetkova, N., Wolkers, W., Leidy, C., Tablin, F., Longo, M. & Crowe, J. *Biophys. J.* **84** 3045-3051 (2003)

- [47] Hinch, D.K. & Hagemann, M. *Biochem. J.* **383** 277-283 (2004)
- [48] Pereira, C.S., Lins, R.D., Chandrasekhar, I., Freitas, L.C.G & Hünenberger, P.H. *Biophys. J.* **86** 2273-2285 (2004)
- [49] Pereira, C.S. & Hünenberger, P.H. *J. Phys. Chem. B* **110** 15572-15581 (2006)
- [50] Pereira, C.S. & Hünenberger, P.H. *Biophys. J.* **95** 3525-3534 (2008)
- [51] Williams, R.J. & Leopold, A.C. *Plant. Physiol.* **89** 977-981 (1989)
- [52] Green, J.L. & Angell, C.A. *J. Phys. Chem.* **93** 2880-2882 (1989)
- [53] Sun, W.Q. & Leopold, A.C. *Ann. Bot.* **74** 601-604 (1994)
- [54] Koster, K.L., Webb, M.S., Bryant, G. & Lynch, D.V. *Biochim. Biophys. Acta* **1193** 143-150 (1994)
- [55] Sun, W.Q., Leopold, A.C., Crowe, L.M. & Crowe, J.H. *Biophys. J.* **70** 1769-1776 (1996)
- [56] Sun, W.Q. & Leopold, A.C. *Comp. Biochem. Physiol. A* **117** 327-333 (1997)
- [57] Koster, K.L., Lei, Y.P., Anderson, M., Martin, S. & Bryant, G. *Biophys. J.* **78** 1932-1946 (2000)
- [58] Belton, P.S. & Gil, A.M. *Biopolymers* **34** 957-961 (1994)
- [59] Tironi, I.G., Luty, B.A. & van Gunsteren, W.F. *J. Chem. Phys.* **106** 6068-6075 (1997)
- [60] Cottone, G., Ciccotti, G. & Cordone, L. *J. Chem. Phys.* **117** 9862-9866 (2002)
- [61] Lins, R.D., Pereira, S.C. & Hünenberger, P.H. *Proteins: Struct. Funct. Bioinformatics* **55** 177-186 (2004)
- [62] Giuffrida, S., Cottone, G. & Cordone, L. *J. Phys. Chem. B* **108** 15415-15421 (2004)
- [63] Massari, A.M., Finkelstein I.J. McClain, B.L., Goj, A., Wen, X., Bren, K.L., Loring, R.F. & Fayer, M.D. *J. Am. Chem. Soc.* **127** 14279-14289 (2005)
- [64] Cordone, L., Cottone, G., Giuffrida, S., Palazzo, G., Venturoli, G. & Viappiani, C. *Biochim. Biophys. Acta* **1749** 252-281 (2005)
- [65] Lenne, T., Garvey, C.J., Koster, K.L. & Bryant, G. *J. Phys. Chem. B* **113** 2486-2491 (2009)
- [66] Westh, P. *Phys. Chem. Chem. Phys.* **10** 4110-4112 (2008)
- [67] Rand, R.P. & Parsegian, V.A. *Biochimica et Biophysica Acta* **988** 351-376 (1989)
- [68] Sakurai, M. et al. *PNAS* **105** 5093-5098 (2008)
- [69] Sum, A.K., Faller, R. & de Pablo, J.J. *Biophys. J.* **85** 2830-2844 (2003)

- [70] Sum, A.K. *Chem. Biodiv.* **2** 1503-1516 (2005)
- [71] Doxastakis, M., Sum, A.K. & de Pablo, J.J. *J. Phys. Chem. B* **109** 24173-24181 (2005)
- [72] Skibinsky, A., Venable, R.M. & Pastor, R.W. *Biophys. J.* **89** 4111-4121 (2005)
- [73] Leekumjorn, S. & Sum, A.K. *Biophys. J.* **90** 3951-3965 (2006)
- [74] Leekumjorn, S. & Sum, A.K. *Molecular Simulation* **32** 219-230 (2006)
- [75] Róg, T., Vattulainen, I., Bunker, A. & Karttunen, M. *J. Phys. Chem. B* **111** 10146-10154 (2007)
- [76] van den Bogaart, G., Hermans, N., Krasnikov, V., de Vries, A.H. & Poolman, B. *Biophys. J.* **92** 1598-1605 (2007)
- [77] Leekumjorn, S., Wu, Y., Sum, A.K. & Chan, C. *Biophys. J.* **94** 2869-2883 (2008)
- [78] Leekumjorn, S. & Sum, A.K. *J. Phys. Chem. B* **112** 10732-10740 (2008)
- [79] Xie, G. & Timasheff, S.N. *Biophys. Chem.* **64** 25-43 (1997)
- [80] Curtis, J.E., Dirama, T.E., Carri, G.A. & Tobias, D.J. *J. Phys. Chem. Lett.* **110** 22953-22956 (2006)
- [81] Liu, F. et al. *J. Mol. Graph. Model.* **27** 421-429 (2008)
- [82] Cottone, G. *J. Phys. Chem. B* **111** 3563-3569 (2007)
- [83] Smolin, N. & Winter, R. *J. Phys. Chem. B* 112- (2008)
- [84] Peric, L., Hansen, H.S., Baron, R. & Hünenberger, P.H. *J. Phys. Chem. B* to be submitted- (2009)
- [85] Brown, G.M. et al. *Acta Cryst.* **B28** 3145- (1972)
- [86] Albertorio, F., Chapa, V.A., Chen, X., Diaz, A.J. & Cremer, P.S. *J. Am. Chem. Soc.* **129** 10567-10574 (2007)
- [87] Donnamaria, M.C., Howard, E.I. & Grigera, J.G. *J. Chem. Soc. Faraday Trans.* **90** 2731-2735 (1994)
- [88] Richards, A.B. et al. *Food Chem. Toxicol.* **40** 871-898 (2002)
- [89] Chen, T., Fowler, A. & Toner, M. *Cryobiology* **40** 277-282 (2000)
- [90] Lerbret, A. et al. *Carbohydr. Res.* **340** 881-887 (2005)
- [91] Dowd, M.K., French, A.D. & Reilly, P.J. *Carbohydr. Res.* **264** 1-19 (1994)
- [92] Liu, Q., Schmidt, R.K., Teo, B., Karplus, P.A. & Brady, J.W. *J. Am. Chem. Soc.* **119** 7851-7862 (1997)

- [93] Bonanno, G., Noto, R. & Fornili, S.L. *J. Chem.Soc. Faraday Trans.* **94** 2755-2762 (1998)
- [94] Conrad, P.B. & de Pablo, J.J. *J. Phys. Chem. A* **103** 4049-4055 (1999)
- [95] Engelsen, S.B. & Pérez, S. *J. Phys. Chem. B* **104** 9301-9311 (2000)
- [96] Bordat, P., Lerbret, A., Demaret, J.-P., Affouard, F. & Descamps, M. *Europhys. Lett.* **65** 41-47 (2004)
- [97] Lerbret, A., Bordat, P., Affouard, F., Descamps, M. & Migliardo, F. *J. Phys. Chem. B* **109** 11046-11057 (2005)
- [98] Choi, Y., Cho, K.W., Jeong, K. & Jung, S. *Carbohydr. Res.* **341** 1020-1028 (2006)
- [99] Luo, Y., Li, W. & Wang, W. *Env. Exp. Bot.* **63** 378-384 (2008)
- [100] Béranger, F., Crozet, C. Goldsborough A. & Lehmann, S. *Biochem. Biophys. Res. Commun.* **374** 44-48 (2008)
- [101] Schiraldi, C., Di Lernia, I. & De Rosa, M. *TRENDS Biotech.* **20** 420-425 (2002)
- [102] Higashiyama, T. *Pure Appl. Chem.* **74** 1263-1269 (2002)
- [103] Cerdeira, M., Martini, S. & Herrera, M.L. *J. Food Sci. E* **70** 401-408 (2005)
- [104] Vega, C., Goff, H.D. & Roos, Y.H. *Int. Dairy J.* **17** 683-695 (2007)
- [105] Cerdeira, M., Puppo, M.C. & Herrera, M.L. *J. Food Sci.* **68** 2644-2650 (2003)
- [106] Whelan, A.P., Regand, A., Vega, C., Kerry, J.P. & Goff, H.D. *Int. J. Food. Sci. Technol.* **43** 510-516 (2008)
- [107] Nie, Y., de Pablo, J.J. & Palecek, S.P. *Biotechnol. Bioeng.* **92** 79-90 (2005)
- [108] Willart, J.F. & Descamps, M. *Mol. Pharmaceutics* **5** 905-920 (2008)
- [109] Buléon, A. & Tran, V. *Int. J. Biol. Macromol.* **12** 345- (1990)
- [110] Imberty, A. & Pérez, S. *Int. J. Biol.Macromol.* **11** 177- (1989)
- [111] Pereira, C.S., Kony, D., Baron, R., Müller, M., van Gunsteren, W.F. & Hünenberger, P.H. *Biophys. J.* **90** 4337-4344 (2006)
- [112] Pereira, C.S., Kony, D., Baron, R., Müller, M., van Gunsteren, W.F. & Hünenberger, P.H. *Biophys. J.* **93** 706- (2007)
- [113] Kony, D., Damm, W., Stoll, S., van Gunsteren, W.F. & Hünenberger, P.H. *Biophys. J.* **93** 442-455 (2007)
- [114] van Gunsteren, W.F., Billeter, S.R., Eising, A.A., Hünenberger, P.H., Krüger, P., Mark, A.E., Scott, W.R.P. & Tironi, I.G. Biomolecular simulation: The GROMOS96 manual and user guide. Verlag der Fachvereine, Zürich, Switzerland. (1996)

- [115] Scott, W.R.P., Hünenberger, P.H., Tironi, I.G., Mark, A.E., Billeter, S.R., Fennen, J., Torda, A.E., Huber, T., Krüger, P. & van Gunsteren, W.F. *J. Phys. Chem. A* **103** 3596-3607 (1999)
- [116] Lins, R.D. & Hünenberger, P.H. *J. Comput. Chem.* **26** 1400-1412 (2005)
- [117] Schuler, L.D. & van Gunsteren, W.F. *Mol. Simul.* **25** 301-319 (2000)
- [118] Schuler, L.D., Daura, X. & van Gunsteren, W.F. *J. Comput. Chem.* **22** 1205-1218 (2001)
- [119] Chandrasekhar, I., Kastenholz, M.A., Lins, R.D., Oostenbrink, C., Schüler, L.D., Tieleman, D.P. & van Gunsteren, W.F. *Eur. Biophys. J.* **32** 67-77 (2003)
- [120] Soares, T.A., Hünenberger, P.H., Kastenholz, M.A., Kräutler, V., Lenz, T., Lins, R.D., Oostenbrink, C. & van Gunsteren, W.F. *J. Comput. Chem.* **26** 725-737 (2005)
- [121] Kräutler, V., Müller, M. & Hünenberger, P.H. *Carbohydr. Res.* **342** 2097-2124 (2007)
- [122] Hansen, H.S. & Hünenberger, P.H. *J. Comput. Chem.*, in press (2009)
- [123] Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F. & Hermans, J. Interaction models for water in relation to protein hydration. In: *Intermolecular Forces*. Pullman, B., Ed. Reidel, Dordrecht, The Netherlands. pp 331-342 (1981)
- [124] Hockney, R.W. *Methods Comput. Phys.* **9** 136-211 (1970)
- [125] Ryckaert, J.-P., Ciccotti, G. & Berendsen, H.J.C. *J. Comput. Phys.* **23** 327-341 (1977)
- [126] Nagle, J.F. & Tristram-Nagle, S. *Biochim. Biophys. Acta* **1469** 159-195 (2000)
- [127] Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., Di Nola, A. & Haak, J.R. *J. Chem. Phys.* **81** 3684-3690 (1984)
- [128] van Gunsteren, W.F. & Berendsen, H.J.C. *Angew. Chem. Int. Ed.* **29** 992-1023 (1990)
- [129] Barker, J.A. & Watts, R.O. *Mol. Phys.* **26** 789-792 (1973)
- [130] Tironi, I.G., Sperb, R., Smith, P.E. & van Gunsteren, W.F. *J. Chem. Phys.* **102** 5451-5459 (1995)
- [131] Heinz, T.N., van Gunsteren, W.F. & Hünenberger, P.H. *J. Chem. Phys.* **115** 1125-1136 (2001)
- [132] Zhu, Y., Zajicek, J. & Serianni, A.S. *J. Org. Chem.* **66** 6244-6251 (2001)
- [133] Takaoka, Y., Pasenkiewicz-Gierula, M., Miyagawa, H., Kitamura, K., Tamura, Y. & Kusumi, A. *Biophys. J.* **79** 3118-3138 (2000)
- [134] Davis, J.H. *Biochim. Biophys. Acta* **737** 117-171 (1983)
- [135] Dong, R.Y. *Can. J. Phys.* **56** 678-680 (1978)

-
- [136] Douliez, J.-P., Léonard, A. & Dufourc, E.J. *Biophys. J.* **68** 1727-1739 (1995)
- [137] Gally, H.U., Pluschke, G., Overath, P. & Seelig, J. *Biochemistry* **20** 1826-1831 (1981)
- [138] Seelig, J. & Seelig, A. *Biochem. Biophys. Res. Comm.* **57** 406-411 (1974)
- [139] Seelig, A. & Seelig, J. *Biochemistry* **13** 4839-4844 (1974)
- [140] Seelig, A. & Seelig, J. *Biochim. Biophys. Acta* **406** 1-5 (1975)
- [141] Seelig, J., Gally, H.-U. & Wohlgemuth, R. *Biochim. Biophys. Acta* **467** 109-119 (1977)
- [142] Seelig, J. *Quarterly Reviews of Biophysics* **10** 353-418 (1977)
- [143] Seelig, J. & Seelig, A. *Quarterly Reviews of Biophysics* **13** 19-61 (1980)
- [144] Klauda, J.B., Venable, R.M., MacKerell Jr., A.D. & Pastor, R.W. *Current Topics in Membranes* **60** 1-48 (2008)
- [145] Tieleman, D.P., Marrink, S.J. & Berendsen, H.J.C. *Biochim. Biophys. Acta* **1331** 235-270 (1997)
- [146] Lenné, T., Bryant, G., Holcomb, R. & Koster, K.L. *Biochim. Biophys. Acta* **1768** 1019-1022 (2007)

Chapter 4

Conformation, dynamics and ion-binding properties of single-chain polyuronates

4.1 Summary

In the present study, we report and compare sixteen explicit-solvent molecular dynamics simulations (10 ns) of homopolyuronate single chains corresponding to poly- $\beta(1\rightarrow4)$ -D-glucuronate (pGlcU), poly- $\beta(1\rightarrow4)$ -D-mannuronate (pManU), poly- $\alpha(1\rightarrow4)$ -D-galacturonate (pGalU) and poly- $\alpha(1\rightarrow4)$ -L-guluronate (pGlcU). Eight main simulations are performed at 300 K in the presence of Ca^{2+} counter-ions (neutralizing amount), and starting from alternative regular two- and three-fold helical structures. Eight simulation variants probe the effect of an elevated temperature or of a different counter-ion environment. The chains are made formally infinite by application of artificial periodicity along the chain axis (octameric or nonameric repeat unit). The main observations made in these simulations are: (i) the glycosidic linkages (and local helical parameters) show an important flexibility (in time) and variability (along the chains), and regular helical structures only account for a limited fraction of the conformational ensembles populated at 300 K; (ii) for all system considered, the binding of Ca^{2+} counter-ions is essentially non-specific, with the formation of a dense counter-ion atmosphere around the chains (condensation), but no specific (tight-binding) interactions at well-defined coordination sites; (iii) the 3_2 -, 9_5 -, 2_1 - and 2_1 -helices appear to be the preferential regular helical forms for single chain pGlcU, pManU, pGalU and pGulU, respectively, in aqueous solution, with the possibility of a 3_1 -helix for pGalU (these forms should be viewed in the sense of helical propensities). Taken together, these observations suggest that if chain dimers are appropriately described by the egg-box model (or any other structural model with similar qualitative features) chain-chain association within junction zones in gels must be accompanied by a substantial chain stiffening and a dramatic change in the ion-binding mode.

4.2 Introduction

Polyuronates¹⁻⁴ are (predominantly) polymers of uronic acids in (1 \rightarrow 4)-linkage (Figure 4.1 and 4.2). In aqueous solution and at neutral pH, they are polyelectrolytes due to the ionization of their acidic groups. These polymers have diverse biological functions in plants, such as the preservation of

the structure, texture and flexibility, as well as the prevention of desiccation⁵⁻⁹. These functions are related to their ability to form gels²⁻⁴, typically in the presence of divalent metal cations. The three most important natural derivatives of polyuronates are pectins, alginates and glucuronans.

Pectins^{2-4,7-10} present a complex sequence alternating homogeneous segments of α -D-galacturonate (α -D-GalU) residues in (1 \rightarrow 4)-linkage ("smooth regions") and highly-branched segments ("hairy regions") comprising neutral sugar molecules (dominantly rhamnose, galactose, and arabinose). They are present in the cell walls of earth plants (1-2% of the dry weight), and are particularly abundant in fruits and vegetables (10-30%). Pectins may be (methoxy-)esterified to different extents, which strongly influences their physico-chemical properties¹¹⁻¹⁴ (*e.g.* solid state structure, solubility and gel formation characteristics).

Alginates^{2-4,6,15} are linear copolymers of β -D-mannuronate (β -D-ManU) and α -L-guluronate (α -L-GulU) residues in (1 \rightarrow 4)-linkage, predominantly alternating homogeneous segments of either types of residues and regular repeats of their dimer¹⁶. The relative amounts of the two acids and their distribution along the polymer chain varies widely depending on the natural source^{6,15}, and strongly influences the physico-chemical properties of alginates^{6,17}. They are present in large amounts in the extracellular medium of brown seaweeds (20-40% of the dry weight) and, for this reason, constitute the most abundant marine biopolymer.

Finally, glucuronans¹⁸⁻²⁰ are linear polymers of β -D-glucuronate (β -D-GlcU) in (1 \rightarrow 4)-linkage. They are produced by some molds (as fragments of the more complex heteropolysaccharide mucoran²¹) and strains of bacteria (in a partially acetylated form^{18,22}).

The properties of natural polyuronate-based materials are largely used in the food industry, where these polysaccharides serve as stabilizers, thickener, gelling or emulsifying agents^{7,15,23}. In addition, they have found many applications in the fields of medicine (encapsulation, prints), biochemistry (chromatography, culture media), biotechnology (immobilization of enzymes) and environmental protection (detoxification of industrial wastes)^{5,15,24-28}. Most of these industrial and technological applications also rest on the ability of polyuronates to form gels²⁻⁴.

An aqueous gel is formed when the polysaccharide chains associate through non-covalent interactions at the level of junction zones. These zones are separated by disordered segments consisting in general of only one chain. The framework of associated and disordered segments leaves large-sized cavities filled by water molecules. In the context of natural pectins and alginates, the size of the junction zones is typically limited by inhomogeneities of two types in the polymer sequence: (i) the presence of destructuring residues, with a substitution (*e.g.* methylation or glycosylation) preventing them to enter into a regular structure; (ii) the presence, in a copolymer, of segments with heterogeneous sequences and reduced propensities to chain-chain association. This size limitation of the junction zones prevents the formation of crystalline structures (precipitation) and, consequently, stabilizes the polymer-water system in the gel state. The gelation of polyuronates is generally induced by addition of divalent metal cations (typically Ca^{2+}) at a pH close to neutrality. In this case, the leading intermolecular interactions responsible for chain-chain association in the junction zones are ionic interactions between the negatively-charged carboxylate groups and bridging divalent cations^{2-4,29-31}. The most popular structural model for calcium-induced chain-chain association between homoguluronate (pGulU) segments in alginates and homogalacturonate (pGalU) segments in pectins is the so-called egg-box model^{29,36}. In this model, pairs of segments in a 2_1 -helical conformation are associated in an antiparallel fashion, leaving "niches" for the calcium ions that are coordinated by (two) carboxylate and (eight) hydroxyl groups from the two chains^{29,36-39}.

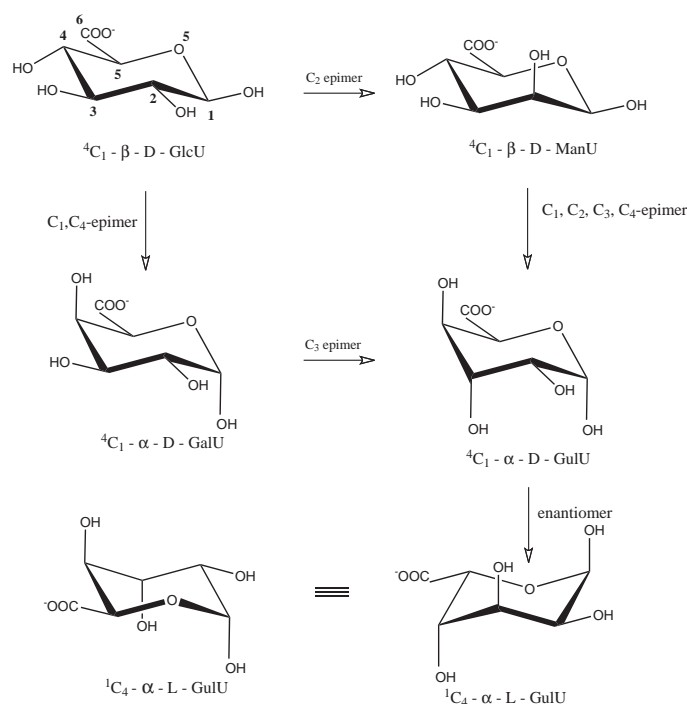


Figure 4.1: Chemical structures of the four uronate building blocks (β -D-GlcU, β -D-ManU, α -D-GalU and α -L-GulU; anomery representative of the linkage in the corresponding homopolyuronate; Figure 4.2). All compounds are represented in their most stable (4C_1 or 1C_4) chair conformation. The (non-natural) D-enantiomer α -D-GulU is also represented (center right drawing), as well as a non-standard representation of the 1C_4 - α -L-GulU chair conformation (bottom left drawing). The atom numbering is also indicated (top left drawing).

For pGulU and pGalU, this model is compatible with the selectivity for calcium over monovalent (*e.g.* sodium) ions^{13,29} (*i.e.* once bound within the gel, the Ca²⁺ ions cannot be displaced by Na⁺ ions) and the stoichiometry of calcium binding^{13,29,40,41} (one Ca²⁺ ion for four uronate residues), as well with the 2_1 -helical conformation observed in the solid state (pGulU; acidic form^{42–46}, Na⁺ salt⁴⁷ and, presumably, Ca²⁺ form⁴⁵) or inferred to dominate in the gel (pGulU^{29,36} and pGalU^{13,37,48}). However, the egg-box model may not be the only model compatible with the available experimental data and it has been questioned in a number of recent studies, on the basis of the following arguments: (*i*) this model suggests a coordination number of ten for the calcium ions, while calcium-carbohydrate complexes typically rather exhibit eight-fold coordination⁴⁹; (*ii*) a molecular modeling study of chain-chain association in homopolyuronates suggested two alternative models with more favorable binding energies³⁹; (*iii*) a model of the Ca²⁺ form of pGulU in the solid state proposed on the basis of fiber diffraction data for the acidic form^{42–44} suggested yet another Ca²⁺ coordination mode⁴⁶; (*iv*) experimental evidence suggests that not only homogeneous pGulU segments, but also regular repeats of the α -[D-ManU- β (1 \rightarrow 4)-L-GulU] dimer may participate in junction zones⁵⁰; (*v*) small-angle X-ray scattering (SAXS) measurements suggest a larger width for the junction zones (compared to that expected for the egg-box dimer), even at the onset of gel formation^{38,51}; (*vi*) X-ray scattering measurements on (slowly-formed) gels appear incompatible with junction zones exclusively associating 2_1 -helices⁵²

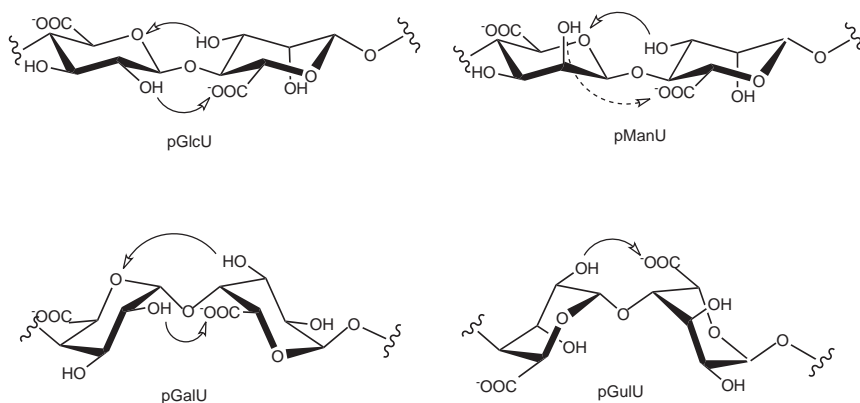


Figure 4.2: Schematic representations of the linkages found in the four homopolyuronates considered in the present study (poly- β (1 \rightarrow 4)-D-GlcU, poly- β (1 \rightarrow 4)-D-ManU, poly- α (1 \rightarrow 4)-D-GalU and poly- α (1 \rightarrow 4)-L-GulU; abbreviated as pGlcU, pManU, pGalU and pGulU, respectively). The drawings are oriented (left to right) from the non-reducing end to the reducing end. The linkage geometry is chosen to be approximately representative of the one that would be found in a regular a 2_1 -helix. The β (1 \rightarrow 4) linkage (pGlcU and pManU) is diequatorial (similar to cellulose^{32,33}) while the α (1 \rightarrow 4)-linkage (pGalU and pGulU) is diaxial (similar to amylose^{34,35}). Possible inter-residue hydrogen bonds across the glycosidic linkage are indicated by arrows: $H_2 \rightarrow O'_6$ (left to right arrow; O'_6 stands for either or both of the carboxylate oxygen atoms, O'_{61} and O'_{62} ; not possible for pManU in the represented conformation) and $O_5 \leftarrow H'_3$ (right to left arrow; not possible for pGulU), where the prime indicates the successor of a residue in the direction of the reducing end.

(the latter study has in turn been questioned in view of the low-resolution data employed⁴⁵). In summary, it appears that the original egg-box model^{29,36} is excessively idealized in view of all recent data. In addition, although the model, initially proposed for pGulU in alginates^{29,36}, has often been assumed applicable to pGalU in pectins^{13,37,48}, the latter polyuronate presents a 3_1 -helical conformation in the solid state (acidic form and Na^+ salt⁵³). Although a conformational transition is indeed observed for this polymer upon dehydration³⁷, the question of its preferential conformation in junction zones remains a matter of debate^{10,39}. Finally, polyglucuronate (pGlcU) also forms stable gels in the presence of calcium ions^{20,22}. However, the selectivity for calcium against sodium is much lower, and the mechanism involved in chain-chain association may be different.

Although a primary dimerization mechanism is generally accepted for the cation-induced gelation of polyuronates (irrespective of the detailed structural features of chain-chain association), further (presumably weaker; see however⁴¹) aggregation of the dimeric junction zones may occur when Ca^{2+} is the predominant counter-ion in solution^{38,41,45} (but not in the presence of a large excess of univalent cations^{12,13,48}) or upon dehydration⁵². It should also be mentioned that highly-methoxylated pectins may form gels in the absence of divalent cations^{11,54} (at low pH and in the presence of cosolutes like sucrose), but the mechanism of chain-chain association is different and the junction zones typically associate more than two chains.

Fiber diffraction experiments indicate a right-handed 3_1 -helical conformation for pGalU (acidic form and Na^+ salt⁵³) and a 2_1 -helical conformation for pGulU (acidic form⁴²⁻⁴⁶, Na^+ salt⁴⁷ and, presumably, Ca^{2+} form⁴⁵), pGlcU (Ca^{2+} form¹⁸) and pManU (acidic form^{42-44,46,47}) in the solid

state. Calculated energy differences between the regular two- and three-fold helical conformations of homopolyuronates suggest that two types of helical conformations (the 2_1 -helix and one type of three-fold helix) are almost equally stable⁵⁵, with the exception of pGulU (significantly more stable 2_1 -helix). Thus, depending on the environment (*e.g.* solvation, packing forces, nature of the counter-ions), one or the other structure may predominate. This conformational flexibility is exploited by pManU, which adopts in the solid state a 2_1 -helical conformation in its acidic form^{42–44,46,47} but a three-fold helical conformation (presumably a left-handed 3_2 -helix⁵⁵) as a salt of monovalent cations⁴⁷. A similar feature holds for pGalU in pectins, where a conformational transition between the gel (presumably^{12,13,37} 2_1 -helix) and the solid state (3_1 -helix⁵³) has been evidenced upon drying¹². In contrast, pGulU maintains a 2_1 -helical structure in the solid state^{42–47} and presumably in the gel⁴⁵ (see however⁵²). Note that this more limited flexibility of pGulU relative to pManU seems to correlate with the distribution of Gul-rich and Man-rich alginates among more rigid and more flexible tissues in plant⁵⁶.

In spite of the considerable amount of experimental data on polyuronate systems, the lack of high resolution information represents a serious obstacle in the understanding of their conformational and aggregation properties. For this reason, modeling studies represent a promising approach to shed some light on these properties^{10,19,39,54,55,57–64}.

A number of recent studies (in particular by Braccini *et al.*^{39,55}) have investigated the conformational flexibility^{19,55,57}, energetically favored regular helical forms^{19,55} and ion-binding properties⁵⁵, as well as possible chain-chain association modes³⁹, for the four most relevant homopolyuronates, namely poly- $\beta(1 \rightarrow 4)$ -D-GlcU, poly- $\beta(1 \rightarrow 4)$ -D-ManU, poly- $\alpha(1 \rightarrow 4)$ -D-GalU and poly- $\alpha(1 \rightarrow 4)$ -L-GulU (Figure 4.2; abbreviated here as pGlcU, pManU, pGalU and pGulU, respectively). The modeling approach employed involved a combination of: (i) determination of relaxed-residue energy maps at the disaccharide level using molecular mechanics (MM3^{65–67}) energy minimization; (ii) conformational sampling at the polymer level using a Monte Carlo procedure⁵⁷; (iii) refinement of energetically favored regular helical models; (iv) surface probing for cation binding sites around rigid regular helices using the GRID procedure⁶⁸; (v) rigid-body docking of pairs of helices in various relative orientations³⁹. These studies suggested in particular: (i) a more important stiffness for the diaxially-linked polyuronates (pGalU, pGulU) compared to the diequatorially-linked ones (pGlcU, pManU); (ii) the existence of alternative two-fold (2_1 -helix) and three-fold (3_1 -helix for pGulU; 3_2 -helix for pGlcU, pManU and pGulU) regular helical conformations with similar stabilities, except for pGulU (significantly higher stability for the 2_1 -helix); (iii) a higher calcium binding specificity (*i.e.* better defined binding sites) for the two diaxially-linked polyuronates⁵⁵; (iv) plausible chain-chain association modes (alternative to the egg-box model) for pGalU and pGulU³⁹. However, three major shortcomings of this modeling strategy are that: (i) it entirely relies on an implicit description of solvation effects (continuum electrostatics); (ii) it also relies, for the first three steps, on an implicit representation of the counter-ions (*i.e.* their mean effects was generally approximated by considering the acidic rather than ionized form of the uronate residues); (iii) it relies, for the last two steps, on a non-dynamical (rigid) representation of the polymer chain.

Over the last few years, other types of models have been developed with the goal of providing a more mesoscopic (and thermodynamically-based) account of polyuronate-cation interactions, described as a combination of (non-specific) counter-ion condensation and (specific) tight binding effects^{17,69–71}. These models go beyond a purely atomistic description of the phenomenon, and have proved quite successful at reproducing experimental data.

A well suited theoretical method for bridging the gap between structural^{39,55} (rigid and implicit solvent) and mesoscopic^{17,69–71} models is explicit-solvent molecular dynamics (MD) simulation. To our knowledge, this technique has only been applied to polyuronates by four groups, to investigate calcium-bound complexes of pGalU oligomers^{60–62,72} (6–24 residues; 2–4 oligomers) or gellan oligomers⁵⁸ (4 residue; 2 oligomers), single-chain pGalU oligomers⁷³ (21 residues; charged or uncharged) or the interaction of single-chain alginates with calcium ions or calcite surfaces⁶⁴ (various Gul-Man ratios; 20 residues).

The goal of the present study is to expand the work of Braccini *et al.*⁵⁵ on single-chain polyuronates by taking solvent molecules, counter-ions and chain dynamics fully into account through the use of explicit-solvent MD simulation. To this purpose, we report the results of sixteen simulations (10 ns) of the same four homopolyuronates (pGlcU, pManU, pGalU and pGulU), modeled as formally infinite chains (by application of periodic boundary conditions along the chain axis with octameric or nonameric repeat units). These simulations are initiated from different starting structures, corresponding to the regular two- and three-fold helical models proposed by Braccini *et al.*⁵⁵ Most simulations are carried out at 300 K in the presence of calcium counter-ions (neutralizing amount), but some simulations also investigate the effect of an elevated temperature or of a different counter-ion environment.

4.3 Computational details

4.3.1 Molecular Dynamics Simulation

All MD simulations were performed using the GROMOS96 program^{74,75}, together with GROMOS the 45A4 force field^{76–81} (including a recently reoptimized parameter set for hexopyranose-based carbohydrates^{81–83}) and the SPC water model⁸⁴. Force-field parameters for the uronate residues were adapted from those of the corresponding hexopyranoses (substitution of the hydroxymethyl group by a carboxylate group), and are summarized in Table 4.1. They were validated by MD simulations of the four corresponding diuronates in water as illustrated in Figure 4.13.

A total of sixteen explicit-solvent simulations (10 ns) were carried out (Table 4.2). Eight main simulations involved the four (single-chain) homopolyuronates (pGlcU, pManU, pGalU and pGulU; Figure 4.1 and 4.2) in the presence of Ca^{2+} counter-ions at 300 K starting from two different regular helical conformations. Eight simulation variants involved a higher simulation temperature (500 K) or a different counter-ion environment (Na^+ ions or Ca^{2+} -like ions with modified Lennard-Jones interaction parameters). The equations of motion were integrated using the leapfrog scheme^{74,85} with a timestep of 2 fs. The SHAKE algorithm⁸⁶ was applied to constrain all bond lengths with a relative geometric tolerance of 10^{-4} . The simulations were carried out under periodic boundary conditions based on rectangular computational boxes (dimensions L_x , L_y , L_z) and in the canonical (NVT) ensemble. The temperature was maintained close to its reference value (300 or 500 K) by weakly coupling the solute and solvent degrees of freedom separately to a heat bath⁸⁷, with a relaxation time of 0.1 ps. Whenever required (equilibration), the pressure was also maintained close to a reference value of 1 bar by weakly coupling the particle coordinates and box dimensions (isotropic coordinate scaling) to a pressure bath⁸⁷, with a relaxation time of 0.5 ps. The center of mass motion was removed every 200 ps. Nonbonded interactions were

| Atom | Atom type | q [e] | |
|------------------------------|------------------|---|--|
| C6 | 11 | 0.36 | |
| O61 | 2 | -0.68 | |
| O62 | 2 | -0.68 | |
| Bond | b_0 [nm] | k_b [$10^{-6}\text{kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-4}$] | |
| C5-C6 | 0.153 | 7.15 | |
| C6-O61 | 0.125 | 1.34 | |
| C6-O62 | 0.125 | 1.34 | |
| Bond-angle | θ_0 [deg] | k_θ [$\text{kJ}\cdot\text{mol}^{-1}$] | |
| C5-C6-O61 | 117.0 | 635.00 | |
| C5-C6-O62 | 117.0 | 635.00 | |
| O61-C6-O62 | 126.0 | 770.00 | |
| Improper-dihedral distortion | ξ_0 [deg] | k_ξ [$\text{kJ}\cdot\text{mol}^{-1}\cdot\text{deg}^{-2}$] | |
| C6-O61-O62-C5 | 0.0 | 0.051 | |
| Torsional dihedral torsion | m | $\cos \delta$ | k_ϕ [$\text{kJ}\cdot\text{mol}^{-1}$] |
| O5-C5-C6-O61 | 6 | +1 | 1.00 |

Table 4.1: Force-field parameters employed for the uronate residues. The modifications applied to the GROMOS 45A4 force field⁸¹ to convert a hexopyranose into the corresponding uronate are listed. The hydroxymethyl group (C6, O6, HO6) is removed and replaced by a carboxylate group (C6, O61, O62; the corresponding GROMOS van der Waals atom types and atomic partial charges q are indicated). The carboxylate group represents one charge group (net charge $-1 e$). Following the GROMOS rules^{74,75}, first and second covalent neighbors are excluded from any non-bonded interaction, while third neighbor interactions are described by a special set of Lennard-Jones parameters (together with normal electrostatic interactions). The covalent interaction parameters^{74,75,81} (quartic bond-stretching, cosine-harmonic bond-angle bending and harmonic improper-dihedral distortion, with corresponding reference values b_0 , θ_0 and ξ_0 , and force constant k_b , k_θ and k_ξ , respectively; dihedral-angle torsion, with corresponding multiplicity m , phase shift cosine $\cos \delta$ and force constant k_ϕ) are indicated. The terms involving atoms C4, C5 or O5 replace the corresponding terms in the hexopyranose. In particular, the dihedral-angle torsional terms (one or two) associated with the orientation of the hydroxymethyl group are removed and replaced by the indicated six-fold potential associated with the orientation of the carboxylate group.

computed using a twin-range scheme^{74,88}, with short- and long-range cutoff distances of 0.8 and 1.4 nm respectively, and a frequency of 5 timesteps for the update of the short-range pairlist and intermediate-range interactions. A reaction-field correction was applied to account for the mean effect of omitted electrostatic interactions beyond the long-range cutoff distance, using a relative dielectric permittivity of 61 as appropriate for the SPC water model⁸⁹. All simulations were carried out for a duration of 10 ns, and coordinates were saved to file every 1 ps for subsequent analysis.

To avoid the presence of chain-end effects and the reduction of cooperativity associated with the simulation of finite (short) oligomers^{58,60–62,64,72,73}, the choice was made to simulate infinite chains by taking advantage of the periodic boundary conditions (note that an alternative approach has been recently suggested⁷³). To this purpose, a polymer segment consisting of N_{res} residues

| Code | System | Helix | N_{res} | Ions | N_w | L_x [nm] | L_y [nm] | L_z [nm] | T_{sim} [K] |
|--------------------|--------|----------------|-----------|-------------------------------------|-------|------------|------------|------------|--------------------|
| Glc2 | pGlcU | 2 ₁ | 8 | 4Ca ²⁺ | 3509 | 5.03 | 5.03 | 4.29 | 300 ^a |
| Man2 | pManU | 2 ₁ | 8 | 4Ca ²⁺ | 3473 | 5.04 | 5.04 | 4.26 | 300 ^a |
| Gal2 | pGalU | 2 ₁ | 8 | 4Ca ²⁺ | 2983 | 5.02 | 5.02 | 3.67 | 300 ^a |
| Gul2 | pGulU | 2 ₁ | 8 | 4Ca ²⁺ | 2986 | 5.01 | 5.01 | 3.67 | 300 ^{a,b} |
| Glc3 | pGlcU | 2 ₂ | 9 | 5Ca ²⁺ +1Cl ⁻ | 3777 | 5.06 | 5.06 | 4.58 | 300 |
| Man3 | pManU | 3 ₂ | 9 | 5Ca ²⁺ +1Cl ⁻ | 3858 | 5.06 | 5.06 | 4.70 | 300 |
| Gal3 | pGalU | 3 ₁ | 9 | 5Ca ²⁺ +1Cl ⁻ | 3475 | 5.06 | 5.06 | 4.20 | 300 |
| Gul3 | pGulU | 3 ₂ | 9 | 5Ca ²⁺ +1Cl ⁻ | 3515 | 5.06 | 5.06 | 4.25 | 300 |
| Gul2 _{Na} | pGulU | 2 ₁ | 8 | 8Na ⁺ | 2982 | 5.03 | 5.03 | 3.68 | 300 ^a |

Table 4.2: Simulated systems and simulation conditions. The simulation code, type of homopolyuronate considered (system; Figure 4.1 and 4.2), initial chain conformation (helix; 2₁-, 3₁- or 3₂-helix), number of residues in the computational box (N_{res} ; imposed periodic repeat unit along a formally infinite chain), types and numbers of counter-ions (ions), number of water molecules in the computational box (N_w), box dimensions (L_x , L_y , L_z ; after 0.2 ns equilibration at constant pressure) and simulation temperature (T_{sim}) are reported for the eight main simulations at 300 K performed with Ca²⁺ counter-ions and one of the eight simulation variants at 500 K or with a different counter-ion environment (Glc2H, Man2H, Gal2H, Gul2H, Gul2_{Na}, Gul2_{Na}H, Gul2_{LJ-} and Gul2_{LJ+}; see Table footnote). Note that the L_z dimensions for the simulations at 500 K are identical to those at 300 K, because the temperature raise and the subsequent simulation were performed at constant volume. All simulations were carried out for 10 ns duration (after 0.2 ns equilibration).

within the computational box (with the chain axis aligned along the z -direction) was made infinite by linking the two terminal residues via a glycosidic bond across the periodic boundary (resulting in a formally cyclic topology⁹⁰). In this setup, the N_{res} residues are thus associated through N_{res} glycosidic linkages connecting residues i to $i + 1$ with $i=1\dots N_{res}-1$, the last linkage connecting residue N_{res} to residue 1. This approach is expected to account more appropriately for the cooperativity effects present within longer polymer chains and to entirely eliminate chain-end effects. However, one should be aware that the inclusion of an artificial periodic constraint with a repeat unit of N_{res} residues (along with a fixed end-to-end vector for this N_{res} residue repeat, as imposed by the constant-volume simulations) results in a restriction of the chain longitudinal and torsional flexibility. In particular: (i) the chain extension is artificially constrained to L_z for the repeat unit of N_{res} residues (*i.e.* the average raise h per residue is constrained to $N_{res}^{-1} \cdot L_z$); (ii) the formation of a regular n -fold helix is only possible when N_{res} is multiple of n ; (iii) the chain is artificially maintained close to linearity (as long as N_{res} is significantly smaller than the persistence length of the polymer, expressed on a per residue basis). Experimental estimates for the persistence lengths of the polyuronates considered (approximate, because the probed samples are seldom rigorously homopolymeric) are: 11.4 (pGlcU²⁰), 4.5-9 (pManU⁹¹⁻⁹³), 7-5 (pGalU^{94,95}) and 5-17 nm (pGulU^{96,97}), while corresponding theoretical estimates for the acidic forms⁵⁵ are 10.5, 11.9, 16.6 and 21.0 nm. Considering a distance of 0.56 and 0.48 nm between glycosidic oxygen atoms for ideal chair conformations involved in diequatorial and diaxial linkages, respectively, the latter values correspond to about 20 (pGlcU, pManU) or 40 (pGalU, pGulU) residues. For each of the

| Code | ϕ [deg] | ϕ_{ref} [deg] | ψ [deg] | ψ_{ref} [deg] | γ [deg] | γ_{ref} [deg] | τ_γ [ns] |
|----------------------|--------------|--------------------|--------------|--------------------|----------------|----------------------|--------------------|
| Glc2 | 252.6 (24.1) | 268.7 | 109.3 (13.8) | 84.7 | 115.9 (21.6) | 129 | >4 |
| Man2 | 260.2 (35.2) | 274.5 | 105.7 (16.0) | 83.6 | 108.4 (17.5) | 126 | >6 |
| Gal2 | 103.0 (16.2) | 94.5 | 254.5 (16.5) | 267.9 | 142.4 (30.2) | 136 | 0.16 |
| Gul2 | 261.4 (15.1) | 275.8 | 98.3 (16.1) | 84.7 | 218.0 (31.2) | 225 | 0.25 |
| Glc3 | 269.3 (17.3) | 293.0 | 129.7 (15.3) | 132.8 | 111.6 (17.4) | 134 | >8 |
| Man3 | 279.6 (17.9) | 295.8 | 108.1 (12.9) | 135.4 | 109.4 (17.0) | 127 | >9 |
| Gal3 | 73.7 (14.3) | 74.6 | 215.6 (18.3) | 214.8 | 151.7 (21.4) | 154 | 0.74 |
| Gul3 | 289.8 (14.1) | 295.9 | 137.5 (19.2) | 135.6 | 211.1 (21.5) | 217 | 0.55 |
| Gul2 _{Na} | 261.6 (15.3) | 275.8 | 98.1 (16.4) | 84.7 | 211.1 (21.5) | 217 | 0.16 |
| Gul2 _{LJ-} | 261.6 (15.2) | 275.8 | 98.1 (16.7) | 84.7 | 211.1 (21.5) | 217 | 0.14 |
| Gul2 _{LJ+} | 261.6 (15.1) | 275.8 | 98.1 (15.9) | 84.7 | 211.1 (21.5) | 217 | 0.14 |
| Glc2H | 251.2 (26.4) | 268.7 | 110.6 (17.6) | 84.7 | 115.9 (21.6) | 129 | 0.25 |
| Man2H | 251.7 (39.2) | 274.5 | 113.6 (22.1) | 83.6 | 108.4 (17.5) | 126 | 0.34 |
| Gal2H | 102.3 (19.5) | 94.5 | 255.0 (19.2) | 267.9 | 142.4 (30.2) | 136 | 0.05 |
| Gul2H | 260.2 (19.9) | 275.8 | 99.4 (19.4) | 84.7 | 211.1 (21.5) | 217 | 0.05 |
| Gul2 _{Na} H | 260.2 (19.7) | 275.8 | 99.4 (19.2) | 84.7 | 211.1 (21.5) | 217 | 0.05 |

Table 4.3: Glycosidic linkage conformation and carboxylate group orientation. Mean values and standard deviations (along the time series; between parentheses) of the glycosidic dihedral angles ϕ and ψ , and of the dihedral angle γ characterizing the orientation of the carboxylate group, averaged over the N_{res} linkages for the different simulations (Table 4.2). The γ values were refolded to a single period of 180° prior to averaging (these periods were $30\text{-}210^\circ$ for pGlcU and pManU, $60\text{-}240^\circ$ for pGalU and $120\text{-}300^\circ$ for pGulU; Figure 4.3). The values ϕ_{ref} , ψ_{ref} and γ_{ref} correspond to the dihedral-angle values reported for the different helix types by Braccini *et al.*⁵⁵, based on MM3 relaxed-residue energy maps for the corresponding diuronic acids (ψ_{ref} was converted from the $\tilde{\psi}$ to the ψ definition). The timescale τ_γ associated with the rotational isomerization of the carboxylate groups is also reported. See Section 4.3.2 for the definitions of ϕ , ψ , $\tilde{\psi}$ and γ . Averaging was performed over the final 9 ns of the simulations.

four polymers considered, two alternative types of regular helices were selected as starting conformations for the simulations⁵⁵: 2_1 - and (left-handed) 3_2 -helix for pGlcU, pManU and pGulU, or 2_1 - and (right-handed) 3_1 -helix for pGalU. The 2_1 -helices were investigated for systems with an octameric ($N_{res}=8$) repeat unit, and the 3_1 - and 3_2 -helices for systems with a nonameric ($N_{res}=9$) repeat unit. Initial coordinates for these eight chain configurations were generated using the program InsightII⁹⁸, assuming ideal 4C_1 (pGlcU, pManU, pGalU) or 1C_4 (pGulU) ring geometries and using the reference ϕ and ψ glycosidic dihedral angle values (ϕ_{ref} and ψ_{ref} ; Table 4.3) reported for the different helix types by Braccini *et al.*⁵⁵ (based on MM3 relaxed-residues energy maps for the corresponding diuronic acids). The initial box dimension L_z was set to $N_{res} \cdot h_{ref}$, where h_{ref} is the raise per residue associated with the selected ϕ_{ref} and ψ_{ref} values in the specific chain (Table 4.4). The two other box dimensions L_x and L_y were set to 5 nm. The computational boxes were brought to overall neutrality by inclusion of counter-ions (4 Ca^{2+} or 8 Na^+ for $N_{res}=8$; 5 Ca^{2+} + 1 Cl^- for $N_{res}=9$), and filled to a water density of about $1 \text{ g}\cdot\text{cm}^{-3}$. After energy mini-

| Code | h [nm] | h_{ref} [nm] | h_{exp} [nm] | M | θ [deg] |
|----------------------|---------------|----------------|--------------------|--------------------------|----------------|
| Glc2 | 0.537 (0.008) | 0.516 | 0.516 ^a | 4 | 180.0 (29.8) |
| Man2 | 0.532 (0.008) | 0.514 | 0.518 ^b | 4 | 180.0 (39.6) |
| Gal2 | 0.459 (0.011) | 0.425 | — | 4 | 180.0 (23.1) |
| Gul2 | 0.459 (0.012) | 0.427 | 0.436 ^c | 4 | 180.0 (20.9) |
| Glc3 | 0.509 (0.013) | 0.506 | — | 6 (99.4), 5 | 239.7 (20.3) |
| Man3 | 0.523 (0.008) | 0.506 | — | 5 (99.7), 6 ^e | 200.1 (22.1) |
| Gal3 | 0.467 (0.011) | 0.443 | 0.447 ^d | 3 | 120.0 (24.4) |
| Gul3 | 0.473 (0.010) | 0.442 | — | 6 | 240.0 (24.4) |
| Gul2 _{Na} | 0.459 (0.012) | 0.427 | — | 4 | 180.0 (21.4) |
| Gul2 _{LJ-} | 0.459 (0.012) | 0.427 | — | 4 | 180.0 (21.5) |
| Gul2 _{LJ+} | 0.459 (0.011) | 0.427 | — | 4 | 180.0 (20.8) |
| Glc2H | 0.537 (0.010) | 0.516 | — | 4 | 180.0 (34.1) |
| Man2H | 0.532 (0.009) | 0.514 | — | 4 (99.9), 3 | 179.9 (48.3) |
| Gal2H | 0.459 (0.013) | 0.425 | — | 4 (99.9), 3 | 179.9 (27.5) |
| Gul2H | 0.459 (0.012) | 0.427 | — | 4 | 180.0 (27.0) |
| Gul2 _{Na} H | 0.459 (0.014) | 0.427 | — | 4 | 180.0 (26.9) |

Table 4.4: Helical parameters. Mean values and standard deviations (along the time series; between parentheses) of the turn angle θ and helix raise h , averaged over the N_{res} linkages, together with the number M of full right-handed turns per N_{res} residues, for the different simulations (Table 4.2). The values h_{ref} correspond to the regular model helices proposed by Braccini *et al.*⁵⁵ based on MM3 relaxed-residue energy maps for the corresponding diuronic acids, and the values h_{exp} to experimental (fiber diffraction; solid state) data (see footnotes below). For the M entries, the possible occurrence of two distinct values along a given simulation is indicated, together with the corresponding occurrence for the major component (between parentheses; in percent). See Section 4.3.2 for the definitions of θ , h and M . Averaging was performed over the final 9 ns of the simulations.

Experimental data from: ^a Heyraud *et al.*¹⁸; ^b Atkins *et al.*^{42–44}; ^c Atkins *et al.*^{42–44} and Sikorski *et al.*⁴⁵ (Li *et al.*⁵² report a lower value of 0.39 nm, which they attribute to a 3₁-helical conformation); ^d Walkinshaw *et al.*⁵³ ^e The corresponding distribution of M for Man3 when calculated over the entire (10 ns) trajectory is 5 (89.7%) and 6 (10.3%).

mization, equilibration was carried out by performing 0.2 ns MD simulation at constant pressure (1 bar; allowing for a small variation of L_z from its initial value $N_{res} \cdot h_{ref}$), increasing the temperature from 50 to 300 K in six successive steps, and the resulting configurations used for the production simulations (constant volume; fixed L_z).

The eight main simulations were carried out in the presence of Ca²⁺ counter-ions (and one Cl⁻ for $N_{res}=9$) at a temperature of 300 K (constant volume). In addition, eight simulation variants were carried out for a subset of systems with either or both of the following modifications: (i) an artificially elevated temperature of 500 K; (ii) the inclusion of Na⁺ counter-ions (instead of Ca²⁺) or of Ca²⁺-like counter-ions with their effective radius (distance R_{min}^* at the minimum of the Lennard-Jones curve for ion-water interactions) decreased or increased by 20% relative to $R_{min}^*=0.335$ nm (for the GROMOS Ca²⁺ ion in SPC water). A summary of the simulated systems

and simulation conditions is provided in Table 4.2.

4.3.2 Analysis of the trajectories

The analysis of the trajectories was performed in terms pyranose ring conformation, carboxylate group orientation, glycosidic dihedral-angle conformation, helical parameters, intramolecular hydrogen bonding, ion distribution and binding, chain translational and rotational diffusion, and ionic diffusion. All dihedral-angle values were defined using the IUPAC sign convention⁹⁹, the corresponding numerical values were selected within the range $[0; 360^\circ[$, and the corresponding distributions were computed with a bin width of 1° .

The ring conformation was assessed for all N_{res} residues by monitoring the dihedral angles α_k , $k=1\dots6$ ($C_1-C_2-C_3-C_4$, $C_2-C_3-C_4-C_5$, $C_3-C_4-C_5-O_5$, $C_4-C_5-O_5-C_1$, $C_5-O_5-C_1-C_2$ and $O_5-C_1-C_2-C_3$). The resulting ring shape (chair, boat or other) was defined by assigning the successive dihedral angles to wells of width 60° centered at the six values corresponding to eclipsed or staggered rotamers. The orientation of the corresponding carboxylate group was assessed by monitoring the dihedral angles γ_1 ($O_5-C_5-C_6-O_{61}$) and γ_2 ($O_5-C_5-C_6-O_{62}$). Due to symmetry, these two dihedral angles are expected to present identical properties (given sufficient sampling) and are collectively referred to as γ .

The glycosidic dihedral angles ϕ and ψ associated with the N_{res} glycosidic linkages were defined (according to IUPAC recommendations⁹⁹) as $O_5-C_1-O_1-C'_4$ and $C_1-O_1-C'_4-C'_3$, respectively. An alternative definition $\tilde{\psi}$ is sometimes used for the latter angle⁵⁵ as $C_1-O_1-C'_4-C'_5$. The two definitions are related by the approximate relationships $\tilde{\psi} = \psi + 120^\circ$ (S-stereochemistry at C'_4 ; D-GlcU, D-ManU, L-GulU) or $\tilde{\psi} = \psi - 120^\circ$ (R-stereochemistry at C'_4 ; D-GalU). The corresponding local helical parameters were defined in the following way¹⁰⁰. The helix raise h was measured by the distance between the "reference points" of two successive residues, this point being defined here as the center of geometry of the ring atoms C_2 , C_3 , C_5 and O_5 . The helix turn angle θ was measured by the dihedral angle between the "reference vectors" of two successive residues, this vector being defined here as the projection of the vector connecting the midpoint of the C_2-C_3 bond to the midpoint of the C_5-O_5 bond onto the xy -plane (perpendicular to the helix axis). Values of θ within ranges of $\pm 30^\circ$ around 120 , 180 or 240° were associated to local 3_1 -, 2_1 - or 3_2 -helical conformations, respectively. Due to the (artificial) periodicity constraint along the z -axis of the computational box, the sum of θ values is always an integer multiple of 360° . However, the associated integer multiplicative factor M , representing the number of full right-handed turns per N_{res} residues, may vary in time in the range $0..N_{res}-1$. This parameter was also monitored along the different simulations. A regular n_m helix compatible with the periodic repeat unit of N_{res} residues artificially imposed in the present simulations (*i.e.* for which n is a divisor of N_{res}) will be characterized by $M=n^{-1} \cdot N_{res} \cdot m$. For example, a regular 2_1 -helix with $N_{res}=8$ will be associated with $M=4$. Similarly, regular 3_1 - and 3_2 -helices with $N_{res}=9$ will be associated with $M=3$ and $M=6$, respectively. It is important to realize that although M values can be associated to both simulated configurations and regular helices, it would be incorrect to assume that a simulated configuration with a given M value is necessarily a regular helix of the corresponding type.

Intramolecular hydrogen bonding was analyzed considering all hydroxyl groups as potential hydrogen-bond donors, and all hydroxyl, ring or carboxylate oxygen atoms as potential acceptors. The presence of a hydrogen bond was defined by a maximal hydrogen-oxygen distance of 0.25 nm and a minimal oxygen-hydrogen-oxygen angle of 135° . A hydrogen bond between a hydroxyl

and a carboxylate group was assumed to be present if the hydroxyl group formed a hydrogen bond with either (or both) of the carboxylate oxygen atoms.

The distribution of the N_{ion} positive (Ca^{2+} or Na^+) counter-ions along the chain was visualized by superimposing successive trajectory frames (chain and counter-ions) onto the initial (equilibrated) chain configuration (rotational and translational fit based on the C_2 , C_3 , C_5 and O_5 atoms), and displaying the positions of all positive ions at 10 ps intervals (excluding the first 1 ns of simulation; these figures are complemented by two-dimensional radial distribution functions $g_{2D}(r)$ for the ionic coordinates in the xy -plane relative to the chain axis). The counter-ion distribution was further analyzed in terms of the radial distribution function $g(\rho)$ corresponding to the smallest (minimum-image) distance ρ between any of the two carboxylate oxygen atoms and the closest counter-ion (excluding the first 1 ns of simulation). This function is defined in such a way that the integral of $4\pi\rho^2g(\rho)$ over the entire simulation box (or any larger volume) is equal to $V = L_x \cdot L_y \cdot L_z$. In addition, ion-binding events were monitored for the N_{res} residues, by assigning two cutoff values ρ_t and ρ_l (with $\rho_t < \rho_l$) to the distance ρ . The condition $\rho \leq \rho_t$ was associated to a tight binding event (TB; contact ion pair), the condition $\rho_t < \rho \leq \rho_l$ to a loose binding event (LB; solvent-separated ion pair), while $\rho > \rho_l$ corresponded to the absence of ion binding. The values of ρ_t and ρ_l were chosen as the locations of the first and the second minima in $g(\rho)$ for a given type of ion, *i.e.* ρ_t was set to 0.33 (Ca^{2+}), 0.30 (Na^+), 0.31 (Ca^{2+} -like ion with R_{min}^* decreased by 20%) or 0.35 nm (Ca^{2+} -like ion with R_{min}^* increased by 20%), while the corresponding values of ρ_l were set to 0.57, 0.54, 0.52 or 0.59 nm. The binding (LB or TB) of the carboxylate group of a given residue to a counter-ion that is simultaneously bound (LB or TB) to the carboxylate group of a different residue was referred to as a double binding (DB; ion bridging) event. Two different types of DB events were further distinguished depending on whether they involved the carboxylate groups of two neighbor residues (DB_n) or of two second-neighbor residues (DB_s ; *i.e.* separated by single residue along the chain). The coordination of counter-ions with the other sugar oxygen atoms (O_1 , O_2 , O_3 or O_5) was also monitored in an analogous way (radial distribution functions and binding events, using $\rho_t=0.34$ and $\rho_l=0.56$ nm). All radial distribution functions were computed with a bin width of 0.01 nm.

Diffusion coefficients were calculated by monitoring the mean-square displacements $Q_{\mathbf{q}}(t)$ of various time dependent quantities $\mathbf{q}(t)$, with averaging over all possible time origins (excluding the first 1 ns of simulation). The corresponding diffusion coefficients $D_{\mathbf{q}}$ were calculated based on the Einstein relation, taking into account the dimension n (1, 2 or 3) of the quantity \mathbf{q} . In practice, $D_{\mathbf{q}}$ was estimated from the slope of a regression line fitting $Q_{\mathbf{q}}(t)$ over the interval from 0 to 3 ns where these functions were found to be approximately linear. The quantities \mathbf{q} considered and the corresponding diffusion coefficients $D_{\mathbf{q}}$ were: (i) the z -coordinate of the chain center (center of geometry of the "reference points" of all residues; with $n=1$), to evaluate the chain longitudinal diffusion constant D_z ; (ii) the coordinate vector of the chain center in the xy -plane (with $n=2$), to evaluate the chain transverse diffusion constant D_{xy} ; (iii) the chain orientation angle Θ (mean of the "reference angles" of all residues; with $n=1$), to evaluate the chain rotational diffusion constant D_{Θ} ; (iv) the z -coordinate of a given counter-ion (with $n=1$), to evaluate longitudinal diffusion constants $D_{i,z}$ for all (positive) counter-ions $i=1\dots N_{ion}$; (v) the coordinate vector of a given counter-ion in the xy -plane (with $n=2$), to evaluate the transverse diffusion constants $D_{i,xy}$ for all (positive) counter-ions $i=1\dots N_{ion}$. While evaluating the corresponding time series $\mathbf{q}(t)$, care was taken to follow periodic coordinates as they diffused across different periods, without "refolding" to the reference interval.

Life-times of specific properties (dihedral angle in a specific well, local helical conformation of a given type, formation of inter-residue hydrogen bonds and ion binding events) were estimated by averaging the durations of all uninterrupted occurrences of the specific property over the trajectories. However, to avoid an excessive influence of occurrences or interruptions with very short durations on the overall statistics, an allowed excursion time τ_{exc} (set to 10 ps) was applied, *i.e.* occurrences or interruptions were only accepted as such when their durations exceeded τ_{exc} .

4.4 Results and discussion

4.4.1 Ring conformation and carboxylate group orientation

The mean values and standard deviations of the ring dihedral angles α_k , $k=1\dots6$, averaged over the N_{res} residues, are reported in Table 4.8 for all simulations. The reported values are all consistent with leading 4C_1 (pGlcU, pManU, pGalU) or 1C_4 (pGulU) chair conformations (Figure 4.1), with deviations from the ideal values (60 or 300°) no larger than 10° . The leading chair conformation accounts for 99.7-100.0% of the sampled configurations at 300 K, or (98.1-99.8%) at 500 K. The remaining conformations are slightly distorted chair forms, with very short lifetimes (typically ≤ 1 ps). Note that glucuronate appears to be slightly more flexible compared to the other uronates. A similar observation was reported previously⁸³ in the context of hexopyranoses.

The probability distributions of the dihedral angle γ characterizing the orientation of the carboxylate group, averaged over the N_{res} residues, are displayed in Figure 4.3 for the eight main simulations (300 K, Ca^{2+} counter-ions; Table 4.2). The corresponding probability distributions for the eight simulation variants (500 K or different counter-ion environment), provided in Figure 4.14, are qualitatively similar to the former ones. The mean values and standard deviations associated with these distributions are reported in Table 4.3 for all simulations, together with the corresponding timescales τ_γ for rotational isomerization between the two main (equivalent) carboxylate group orientations, and the values γ_{ref} suggested for the different types of regular helices by Braccini *et al.*⁵⁵ As expected, the γ distributions are essentially periodic of period 180° , the conformations corresponding γ and $\gamma+180^\circ$ representing equivalent carboxylate group orientations. Minor deviations from this expected periodicity, especially visible for simulation Glc2, only arise from incomplete sampling. Over a single 180° period, all distributions are essentially unimodal except for simulation Glc2. Focusing on the eight main simulations, the corresponding average values (Table 4.3) are found to be about 115° (pGlcU, pManU), 145° (pGalU) and 215° (pGulU). These values are in good agreement with the corresponding reference values γ_{ref} , with deviations of at most 25° . A possible explanation for the dominance of these specific rotamers is that they involve approximately equal distances between the two carboxylate oxygen atoms and the glycosidic oxygen, thereby minimizing the corresponding Coulombic repulsion. However, other electronic effects, solvation, inter-residue hydrogen bonding and cation binding may *a priori* also influence these conformational preferences. For simulation Glc2, a second peak in the γ distribution is seen at about 130 - 140° (Figure 4.3), which is not present in simulation Glc3. Besides this change, the distributions are only sensitive to the initial helical conformation for pGalU and pGulU, where the distribution becomes narrower and the peak position is shifted by about $+10^\circ$ or -10° , respectively, when changing from the two- to the three-fold helix.

At 300 K, the timescales τ_γ associated with the rotation of the carboxylate group are sig-

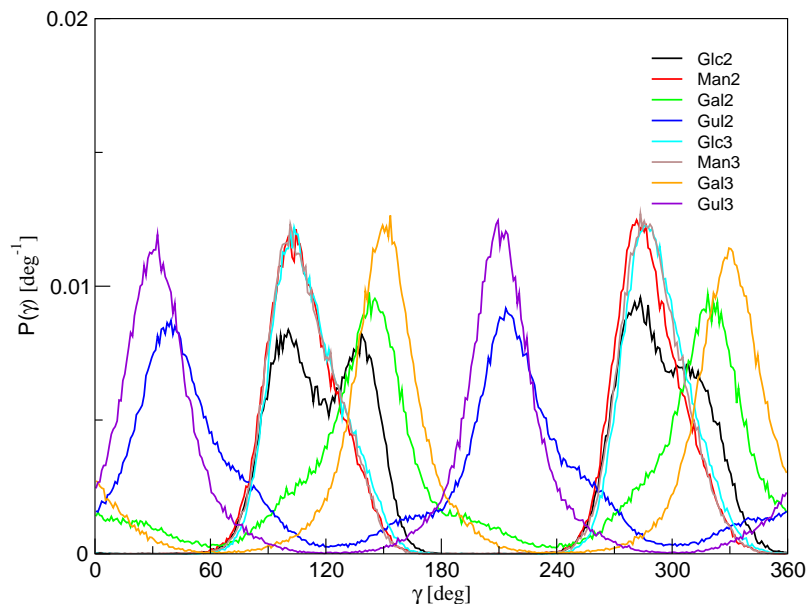


Figure 4.3: Normalized probability distributions of the γ dihedral angle characterizing the orientation of the carboxylate group, averaged over the N_{res} residues, for the eight main simulations (300K, Ca^{2+} counter-ions; Table 4.2). The γ distribution accounts for the combined distributions of the two symmetry-related dihedral angles γ_1 and γ_2 . See Section 4.3.2 for the definitions of γ_1 , γ_2 and γ . Averaging was performed over the final 9 ns of the simulations.

nificantly longer for pGlcU and pManU (≥ 4 ns) compared to pGalU and pGulU (160-740 ps). Expectedly, rotational isomerization is much faster at 500 K (in particular for pGlcU and pManU). For pGulU at 300 K, a change in the counter-ion environment has essentially no influence on τ_γ .

The origin of the above differences between diequatorially- and diaxially-linked homopolyuronates in terms of the dependence of the γ distribution and of the τ_γ value on the helix type at 300 K remains unclear. In particular, these differences can not be easily rationalized on the basis of the hydrogen bonds involving the carboxylate groups (Section 4.4.4) or of counter-ion binding or bridging (Section 4.4.5).

4.4.2 Glycosidic linkage conformation

The probability distributions of the glycosidic dihedral angles ϕ and ψ , averaged over the N_{res} linkages, are displayed in Figure 4.4 for the eight main simulations. The corresponding probability distributions for the eight simulation variants, provided in Figure 4.15, are qualitatively similar to the former ones. The mean values and standard deviations associated with these distributions are reported in Table 4.3 for all simulations, and compared with the values ϕ_{ref} and ψ_{ref} suggested for the different types of regular helices by Braccini *et al.*⁵⁵. A representation of the trajectories in the form of configurations sampled on the Ramachandran (ϕ , ψ) map is also provided in Figures 4.16

and 4.17 for all simulations.

The conformational preferences around the glycosidic dihedral angle ϕ are expected to be dominated by the *exo*-anomeric effect^{101–105} which stabilizes (assuming the chair conformations of Figure 4.1) conformers with ϕ values of about 300° (D-enantiomer, β -linkage; pGlcU, pManU), 60° (D-enantiomer, α -linkage; pGalU) and 300° (L-enantiomer, α -linkage; pGulU). These preferences are essentially respected in the reference values ϕ_{ref} suggested by Braccini *et al.*⁵⁵. These values are comprised between 268.7 and 295.9° for pGlcU, pManU and pGulU, while they are 74.6 or 94.5° for pGalU. The associated ψ_{ref} values are not influenced by such a large stereoelectronic effect, and probably correspond to conformations affording a good compromise between steric effects, inter-residue hydrogen bonding and solvation given the above ϕ_{ref} preferences. Focusing on the eight main simulations, the average ϕ and ψ values agree reasonably well with

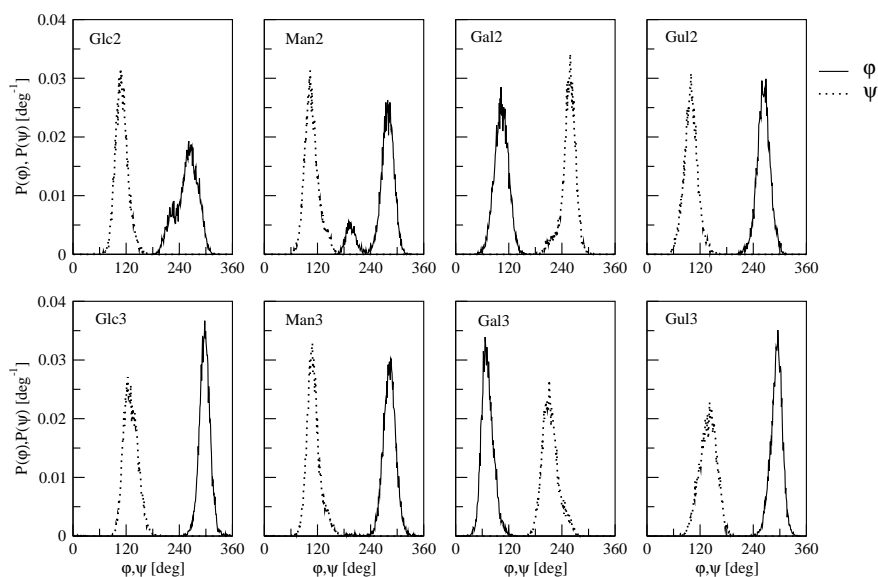


Figure 4.4: Normalized probability distributions of the glycosidic dihedral angles ϕ and ψ , averaged over the N_{res} linkages, for the eight main simulations (300K, Ca^{2+} counter-ions; Table 4.2). See Section 4.3.2 for the definitions of ϕ and ψ . Averaging was performed over the final 9 ns of the simulations.

the corresponding ϕ_{ref} and ψ_{ref} values for the same helix type. However, the deviations of the simulation averages from the reference values (ranging between -16.2 and $+8.5^\circ$ for ϕ or between -27.3 and $+24.6^\circ$ for ψ), as well as the fluctuations observed during the simulations (ranging between 14.1 and 35.2° for ϕ or between 12.9 and 19.2° for ψ), remain significant. The largest deviations relative to the reference values are found for simulations Glc2, Man2 and Man3. Note, however, that perfect agreement should probably not be expected because: (i) the reference values are associated with rigid (idealized) configurations, while the present averages characterize conformational ensembles at 300 K; (ii) the reference values are based on relaxed-residue energy maps for uronic acids (protonated form) at the disaccharide level, while the present averages are representative for formally infinite polyuronate chains (deprotonated form); (iii) the reference values are based on implicit-solvent calculations with the MM3 force field^{65–67} in the absence of

counter-ions, while the present averages are based on explicit-solvent MD simulations with the GROMOS force field^{74,75,81} including counter-ions; (*iv*) the reference values were explicitly adjusted to be compatible with the formation of regular two- or three-fold helices, while no such constraint is applied in the present simulations. Nevertheless, as can be seen from Figure 4.16, the reference conformation of Braccini *et al.* is always included (and usually centered; except for Glc2 and Man2) in the area of the Ramachandran map visited by the simulations.

The probability distributions of the ϕ and ψ dihedral angles in the eight main simulations (Figure 4.4) are all essentially unimodal, with the exception of the ϕ distributions for simulations Glc2 and Man2. This feature explains the particularly high fluctuations of ϕ (24.1 and 32.2°, respectively), and possibly also the more important deviations from ϕ_{ref} (-16.1 and -14.3°, respectively), observed in these two simulations (Table 4.3). The corresponding ϕ distributions can be deconvoluted as the sum of two quasi-Gaussian peaks, centered at 219 and 261° (relative populations 18 and 82%, respectively) for Glc2, or at 195 and 279° (relative populations 21 and 79%) for Man2. The corresponding ψ values (evaluated separately for the subset of conformers with $\phi < 233^\circ$ and $\phi > 233^\circ$) are 100.0 and 112.1° for Glc2, or 97.1 and 108.1° for Man2. In both systems, the second (major) peak is in the expected range for a local 2₁-helical conformation. However, the first (minor) peak is not in the appropriate range for a local 3₂-helical conformation. As will be seen below (Section 4.4.3), it corresponds to a local 3₁-helical conformation. The simulations initiated from the corresponding three-fold helices (Glc3 and Man3) are characterized by an essentially unimodal probability distribution for ϕ , although the minor conformer can also be detected in simulation Man3 (with a marginal occurrence). Note that these minor conformers are the only ones sampled in the present simulations (with a non-negligible population) that are not in the expected range for stabilization by the *exo*-anomeric effect. The time series of the ϕ and ψ dihedral angles for the individual N_{res} linkages are displayed in Figure 4.5 and 4.6 for simulations Glc2, Man2, Glc3 and Man3. Hopping between the minor and major conformers is evident in the time series of the ϕ dihedral angle for the individual linkages during simulations Glc2 and Man2 (Figure 4.5) and, to some extent, Man3 (Figure 4.6). The transitions are local and asynchronous along the chain, *i.e.* only a few linkages adopt the minor conformation simultaneously. The average residence time of a dihedral angle in the latter conformation ($\phi < 233^\circ$) is 70 ps for Glc2 and 220 ps for Man2. For simulations Glc3 and Man3, very brief excursions can be observed (of a few ps duration) toward ϕ values comprised between 0 and 120° (Figure 4.6). Note, finally, that none of the transitions described above for ϕ are associated with significant changes in ψ .

The average ϕ and ψ values at 500 K do not differ significantly from those at 300 K, while the corresponding fluctuations are somewhat enhanced (Table 4.3). For pGulU at 300 K, the nature of the counter-ions has essentially no influence on the corresponding average values and fluctuations.

4.4.3 Helical parameters

The probability distributions of the helix raise h and turn angle θ , averaged over the N_{res} linkages, are displayed in Figure 4.7 for the eight main simulations, and in Figure 4.18 for the eight simulation variants. The latter distributions are qualitatively similar to the former ones. The mean values and standard deviations associated with these distributions are reported in Table 4.4 for all simulations, together with the corresponding values h_{ref} for the regular helical models proposed by Braccini *et al.*⁵⁵ and with experimental values h_{exp} from fiber diffraction measurements in the solid state^{18,42–44,44,45,52,53}. Due to the (artificial) periodicity constraint imposed to the chains

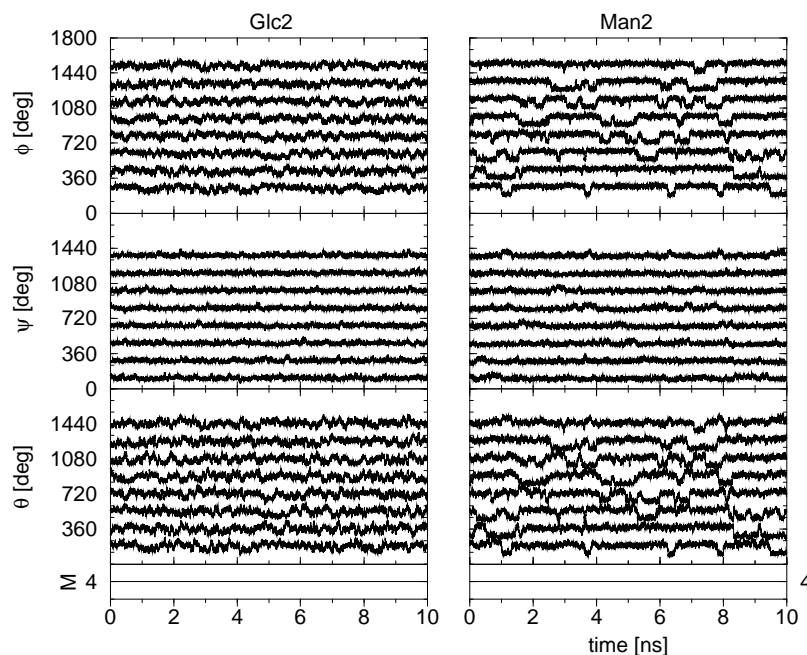


Figure 4.5: Time series of the glycosidic dihedral angle ϕ and ψ , and of the turn angle θ , for the N_{res} linkages, together with that of the number M of full right-handed turns per N_{res} residues, as monitored during simulations Glc2 and Man2 (Table 4.2). See Section 4.3.2 for the definitions of ϕ , ψ , θ and M . Successive curves (ϕ , ψ and θ) corresponding to the different residues along the chain are shifted upwards by 180° for readability.

along the z -axis of the computational box (and the fact that L_z is fixed in the present constant-volume simulations), the instantaneous value of h averaged over all linkages along the chains (and thus, obviously, its time-average) is always equal to $N_{res}^{-1} \cdot L_z$ (compare Tables II and IV). Focusing on the eight main simulations, diequatorially-linked chains (pGlcU, pManU) are more extended than diaxially-linked ones (pGalU, pGulU), by about 16-17% and 8-12% for the simulations initiated from two- and three-fold, helices respectively. This difference is expected since a rigid chair conformation places diequatorial O_4 and O_1 atoms at a distance of about 0.56 nm, compared to only about 0.48 nm for the corresponding diaxial configuration. For the diequatorially-linked chains, h is higher by about 2-6 % for the simulations initiated from a two-fold helix compared to those initiated from the three-fold helix. For the diaxially-linked chains, the opposite is true and the values are higher by about 2-3 % for the three-fold relative to the two-fold helices. The distributions of h values are all unimodal (Figure 4.7). The average values of h are somewhat higher (by 0.5-7%) compared the h_{ref} values for the corresponding model regular helices, themselves being very close to the experimental values h_{exp} in the solid state (Table 4.4). This results from a slight expansion of the chains during the 0.2 ns equilibration period at constant pressure (Section 4.3.1). Such a difference probably results from the introduction of thermally-induced conformational disorder (deviation from rigid regular helices in the simulated ensemble at 300 K; $h \neq h_{ref}$) and of solvation forces (replacing the packing forces present in the solid state; $h \neq h_{exp}$).

Assuming fixed bond lengths and angles as well as rigid pyranose ring conformations, the helix turn angle θ defining the local helicity is a unique function of the glycosidic dihedral angles

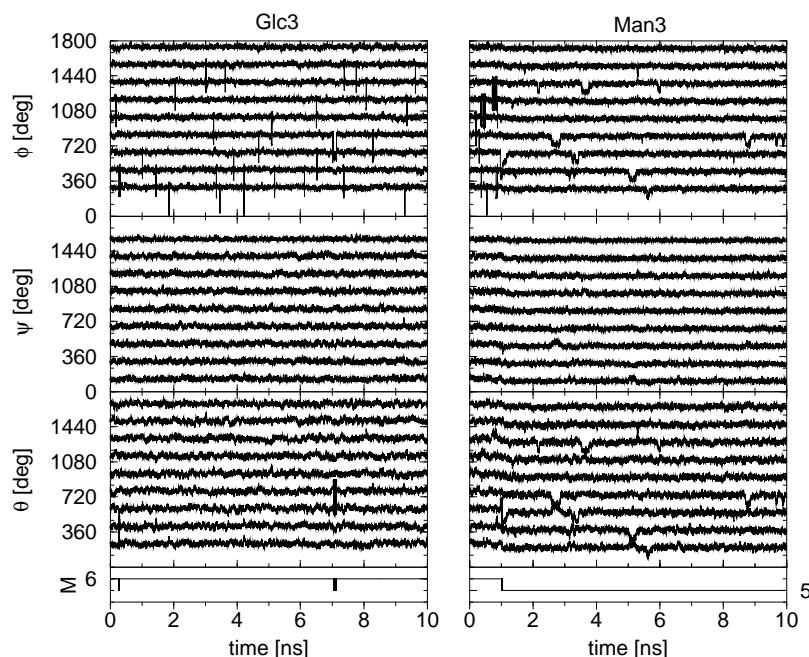


Figure 4.6: Time series of the glycosidic dihedral angle ϕ and ψ , and of the turn angle θ , for the N_{res} linkages, together with that of the number M of full right-handed turns per N_{res} residues, as monitored during simulations Glc3 and Man3 (Table 4.2). See Section 4.3.2 for the definitions of ϕ , ψ , θ and M . Successive curves (ϕ , ψ and θ) corresponding to the different residues along the chain are shifted upwards by 180° for readability.

ϕ and ψ . A representation of the trajectories correlating the (ϕ, θ) and (ψ, θ) dihedral angle pairs is provided in Figures 4.19–4.22 for all simulations. These maps indicate that θ is about equally sensitive to variations in both ϕ and ψ . The (ϕ, θ) maps corresponding to simulations Glc2 and Man2 also show that the minor conformers observed in the corresponding ϕ distributions (Section 4.4.2) can be associated to local 3_1 -helical conformations. The average θ values for these minor conformers ($\phi < 233^\circ$) are 140.5 and 115.7° for Glc2 and Man2, respectively (local 3_1 -helical conformation), to be compared with 191.8 and 197.7° for the major conformers (local 2_1 -helical conformation).

Due to the (artificial) periodicity constraint imposed to the chains in the present simulations, the instantaneous value of the θ , averaged over all linkages along the chain, is always an integer multiple of $N_{res}^{-1} \cdot 360^\circ$. However, because the corresponding integer multiplicative factor M (number of full right-handed turns for the repeat unit of N_{res} residues) may vary in time, the same property need not be satisfied by the corresponding time average (Table 4.4). In practice, however, the value of M remains identical or nearly identical (Glc3, Man2H, Gal2H) to the corresponding value for the initial model helix throughout all but one simulations (Man3), *i.e.* M is 4 for the simulations initiated from 2_1 -helices ($N_{res}=8$), 3 for those initiated from 3_1 -helices ($N_{res}=9$) and 6 for those initiated from 3_2 -helices ($N_{res}=9$). Simulation Man3 represents the exception with a dominant M value of 5 (over the final 9 ns of the simulation) and an average θ value close to 200° .

The time series of θ for the individual N_{res} linkages, together with the corresponding time series of M , are displayed in Figure 4.5 and 4.6 for simulations Glc2, Man2, Glc3 and Man3. The

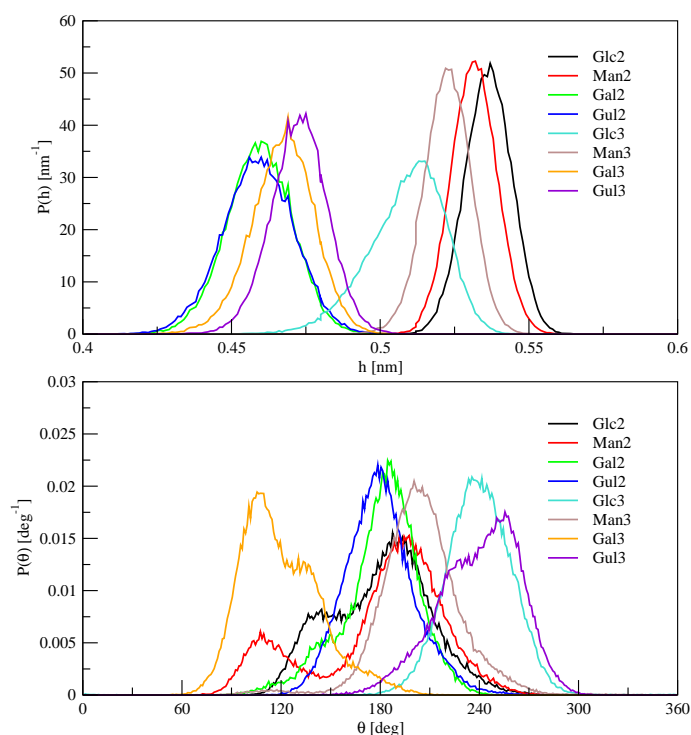


Figure 4.7: Normalized probability distributions of the helix raise h and turn angle θ , averaged over the N_{res} linkages, for the eight main simulations (300K, Ca^{2+} counter-ions; Table 4.2). See Section 4.3.2 for the definition of h and θ . Averaging was performed over the final 9 ns of the simulations.

transitions observed in ϕ between major and minor conformers for simulations Glc2, Man2 and, to some extent, Man3 (Section 4.4.2) clearly correspond to simultaneous transitions in θ . The M value is always 4 in simulations Glc2 and Man2. It is nearly always 6 in simulation Glc3, (besides two short transitions to a value of 5). The situation is different for simulation Man3. During the first 1.02 ns of this simulation, the M value is 6 (as appropriate for a 3_2 -helix with $N_{res}=9$, *i.e.* for the starting structure). However, at this time point, the value abruptly changes to 5 (as appropriate for a 9_5 -helix) and remains constant for the rest of the simulation.

The occurrences of regular or quasi-regular helices are reported in Table 4.5 for all simulations (final 9 ns), together with the five most frequently encountered distributions of linkages among different local helical conformations. Although the value of M essentially stays identical to that of the initial two- or three-fold helical model in all simulations except Man3, the corresponding regular helix only constitutes a limited fraction of the configurations sampled along these simulations. Focusing on the eight main simulations, the corresponding occurrences range from 2.0 to 33.8% (0.0% for Man3). These occurrences increase if one includes quasi-regular helices (*i.e.* presenting one or a few linkages with θ values outside the expected range for a regular helix of the given type). For example, quasi-regular helices matching the initial helical structure except for at most one defect already account for 10.0 to 69.8% (0.1% for Man3) of the sampled configurations. Simulation Man3 evidences a marginal occurrence of (quasi-)regular 3_2 -helix (over the final 9 ns of the simulation), which is not surprising given that its dominant M value of 5 over this time period is incompatible with such a helix. However, when the comparison is performed relative to

| Code | H_0 [%] | H_1 [%] | H_2 [%] | H_3 [%] | P_1 | P_2 | P_3 | P_4 | P_5 |
|----------------------|-----------|-----------|-----------|-------------------|-------------|-------------|-------------|-------------|------------|
| Glc2 | 5.0 | 20.7 | 45.8 | 71.0 | 161 (20.9) | 251 (18.9) | 242 (15.4) | 170 (11.9) | 152 (6.1) |
| Man2 | 2.0 | 10.0 | 28.5 | 51.2 | 242 (25.3) | 161 (15.4) | 251 (15.1) | 233 (10.1) | 170 (7.3) |
| Gal2 | 20.5 | 53.9 | 82.5 | 95.0 | 170 (27.6) | 161 (25.0) | 080 (20.5) | 251 (8.8) | 071 (5.7) |
| Gul2 | 30.2 | 65.9 | 88.3 | 96.5 | 080 (30.2) | 071 (23.2) | 161 (19.9) | 170 (12.5) | 152 (4.4) |
| Glc3 | 33.8 | 69.8 | 90.8 | 97.8 | 009 (33.8) | 018 (18.2) | 008' (17.9) | 017' (17.5) | 026' (3.5) |
| Man3 | 0.0 | 0.1 | 0.2 | 1.1 | 063 (30.2) | 072 (29.2) | 054 (11.3) | 081 (8.4) | 144 (7.5) |
| | 0.5 | 9.0 | 38.2 | 69.6 ^a | | | | | |
| | 37.4 | 68.0 | 89.6 | 97.4 ^b | | | | | |
| Gal3 | 14.8 | 46.4 | 76.3 | 92.9 | 810 (23.7) | 7'10 (23.4) | 900 (14.8) | 6'20 (10.8) | 8'00 (7.8) |
| Gul3 | 12.9 | 40.9 | 70.8 | 89.9 | 017' (24.1) | 018 (20.1) | 009 (12.9) | 026' (11.5) | 008' (7.9) |
| Gul2 _{Na} | 25.4 | 59.3 | 87.0 | 96.8 | 080 (25.4) | 161 (25.1) | 071 (20.6) | 170 (13.4) | 251 (4.9) |
| Gul2 _{LJ-} | 31.1 | 63.6 | 87.4 | 96.6 | 080 (31.1) | 161 (21.5) | 071 (18.8) | 170 (13.6) | 152 (5.2) |
| Gul2 _{LJ+} | 26.5 | 60.5 | 87.1 | 96.8 | 080 (26.5) | 161 (24.2) | 071 (20.5) | 170 (13.5) | 251 (4.9) |
| Glc2H | 2.5 | 11.2 | 32.1 | 57.8 | 242 (20.4) | 161 (18.5) | 251 (14.9) | 152 (10.0) | 332 (7.3) |
| Man2H | 0.1 | 0.7 | 4.7 | 14.8 | 242 (17.3) | 323 (17.0) | 233 (10.2) | 332 (9.2) | 314 (6.2) |
| Gal2H | 9.8 | 31.6 | 61.3 | 83.6 | 161 (26.5) | 170 (13.0) | 251 (12.0) | 242 (10.3) | 080 (9.8) |
| Gul2H | 9.6 | 31.1 | 61.3 | 84.0 | 161 (26.8) | 071 (12.3) | 251 (11.2) | 152 (11.2) | 242 (10.5) |
| Gul2 _{Na} H | 9.6 | 31.6 | 63.0 | 84.6 | 161 (27.7) | 071 (12.9) | 152 (11.1) | 251 (10.2) | 242 (10.0) |

Table 4.5: Occurrences of regular (or quasi-regular) helices and most frequent distributions of the linkages among the different types of local helical conformations. The results are reported for the different simulations (Table 4.2). The successive entries H_k , $k=0..3$, denote the occurrences of trajectory configurations where at least $N_{res} - k$ linkages are characterized by a turn angle θ in the appropriate range for the helical conformation of the regular helix used to initiate the corresponding simulation (Table 4.2). The successive P_k , $k=1..5$, entries encode the five most frequent distributions of the N_{res} residues among the different types of local helical conformations, together with the corresponding occurrences (between parentheses; in percent). The first, second and third digits of the code denote the number of residues in a local 3_1 -, 2_1 - and 3_2 -helical conformation, respectively. A prime after the first or third digit indicates the presence of one linkage with a θ value in the range $[0; 90^\circ[$ or $]270; 360^\circ[$, respectively. The sum of the three digits increased by the number of primes is equal to N_{res} . See Section 4.3.2 for the definition of θ and of the associated local helical conformations. Averaging was performed over the final 9 ns of the simulations.

^a with reference to a regular 2_1 -helix (instead of the 3_2 -helical starting structure); ^b with reference to a regular 9_5 -helix (instead of the 3_2 -helical structure; a local 9_5 -helical conformation is defined here by a turn angle θ within a range of $\pm 30^\circ$ around 200°).

a 9_5 -helix instead, the occurrences of the regular helix and the quasi-regular helix with at most one defect raise to 37.4% and 68.0%, respectively.

The comparison of the different systems shows that simulations Gul2, Glc3 and (to a lesser extent) Gal2 sample configurations that are closer to their initial helical structures and simulations Glc2, Man2 and Man3 sample very different configurations, while simulations Gal3 and Gul3 show intermediate behavior. These differences are also evidenced by the most frequently occurring distributions of linkages among local helical conformations. This distribution is encoded in Table 4.5 in the form of a pattern $a^{[]}bc^{[]}$ indicating how many linkages along the chain adopt a local 3_1 -, 2_1 - or 3_2 -helical conformation, respectively (the prime corresponding to a linkage with θ outside any of the three corresponding intervals, its position indicating the closest recognized configuration). The leading patterns for Gul2 and Glc3 are the regular 2_1 - and 3_2 - helices, respec-

tively, while the second and fourth or second and third most frequent patterns, respectively, are quasi-regular helices with a single defect. Simulations Gal2, Gal3 and Gul3 all have the regular helix (2_1 -, 3_1 - and 3_2 -helix, respectively) as the third most frequent pattern, and the same helix with a single defect as one of the two most frequent ones. In contrast, the leading patterns for Glc2 and Man2 are all irregular. For Man3, the analysis in terms of such patterns is probably somewhat misleading, since these do not account for the local 9_5 -helical conformation.

Taken together, the above observations provide a hint that conformational ensembles in the neighborhood of the regular 3_2 -, 9_5 -, 2_1 - and 2_1 -helices may represent the expected helical propensities for single chain pGlcU, pManU, pGalU and pGulU, respectively, in solution, with the possibility of an alternative 3_1 -helical form for pGalU.

The time evolution of the local helical conformation at the level of the N_{res} glycosidic linkages is illustrated in Figure 4.8 and 4.9 for the eight main simulations. The corresponding occurrences and lifetimes, averaged over the N_{res} linkages (together with the associated standard deviations), are reported in Table 4.6 for all simulations. Note that local conformations that cannot be classified as 2_1 -, 3_1 - or 3_2 - helical conformations (*i.e.* with θ outside the range $[90; 270^\circ]$) have rather limited occurrences, remaining below 1% except for simulations Glc3 (5.9%), Gal3 (7.8%), Gul3 (9.1%) and Man2H (2.8%).

Focusing on the eight main simulations, those initiated from a regular 2_1 -helix expectedly show a predominance of local 2_1 -helical conformations (66.1 and 57.5% for Glc2 and Man2; 81.4 and 85.0% for Gal2 and Gul2). However, local 3_1 - and 3_2 -helical conformations (which require a change of about $\pm 60^\circ$ in θ) are also frequently encountered (about 15-20% each for Glc2 and Man2; about 5-10% each for Gal2 and Gul2). These alternative local conformations occur asynchronously and non-cooperatively along the chain (Figure 4.8), with average lifetimes of the order of 25-55 ps, except for the local 3_1 -helical conformation in Man2 (181 ps; distinct formation events of about 1 ns duration are clearly visible in Figure 4.8, which coincide with the ϕ and θ transitions observed in Figure 4.5). The latter observation is probably related to the fact that the associated transition from the 2_1 -helical conformation involves hopping between two non-overlapping wells of the ϕ distribution (Figure 4.4). The occurrences of about 20% for the local 3_1 -helical conformation in simulations Glc2 and Man2 agree very well with the populations associated with the minor peaks in the ϕ distributions (Section 4.4.2), the major peak being associated with all other local conformation (predominantly 2_1 - and 3_2 -helical).

Except for simulation Man3, the main simulations initiated from a regular three-fold (3_1 - or 3_2 -) helix show an even stronger predominance of the corresponding local helical conformation (87.9% for Glc3; 80.9 and 79.0% for Gal3 and Gul3, respectively). The alternative local 2_1 -helical conformation (which requires a change of about $\pm 60^\circ$ in θ) is also frequently encountered (about 5-10%), but the local three-fold conformation of opposite handedness is essentially absent (probably because it requires a larger change of about $\pm 120^\circ$ in θ). The alternative local 2_1 -helical conformations also occurs asynchronously and non-cooperatively along the chain (Figure 4.9), with an average lifetime of the order of 30-45 ps. For simulation Man3 the results of Figure 4.9 and Table 4.6 are probably somewhat misleading, since they do not account for the local 9_5 - helical conformation dominant after 1.02 ns.

The corresponding graphs describing the time-evolution of the local-helical conformation for the eight simulation variants are provided in Figures 4.23 and 4.24. Expectedly, the interconversion between the different local helical conformations is much more rapid at 500 K compared to 300 K (Figure 4.8). Except for simulation Man2H, the average θ values are essentially unaltered

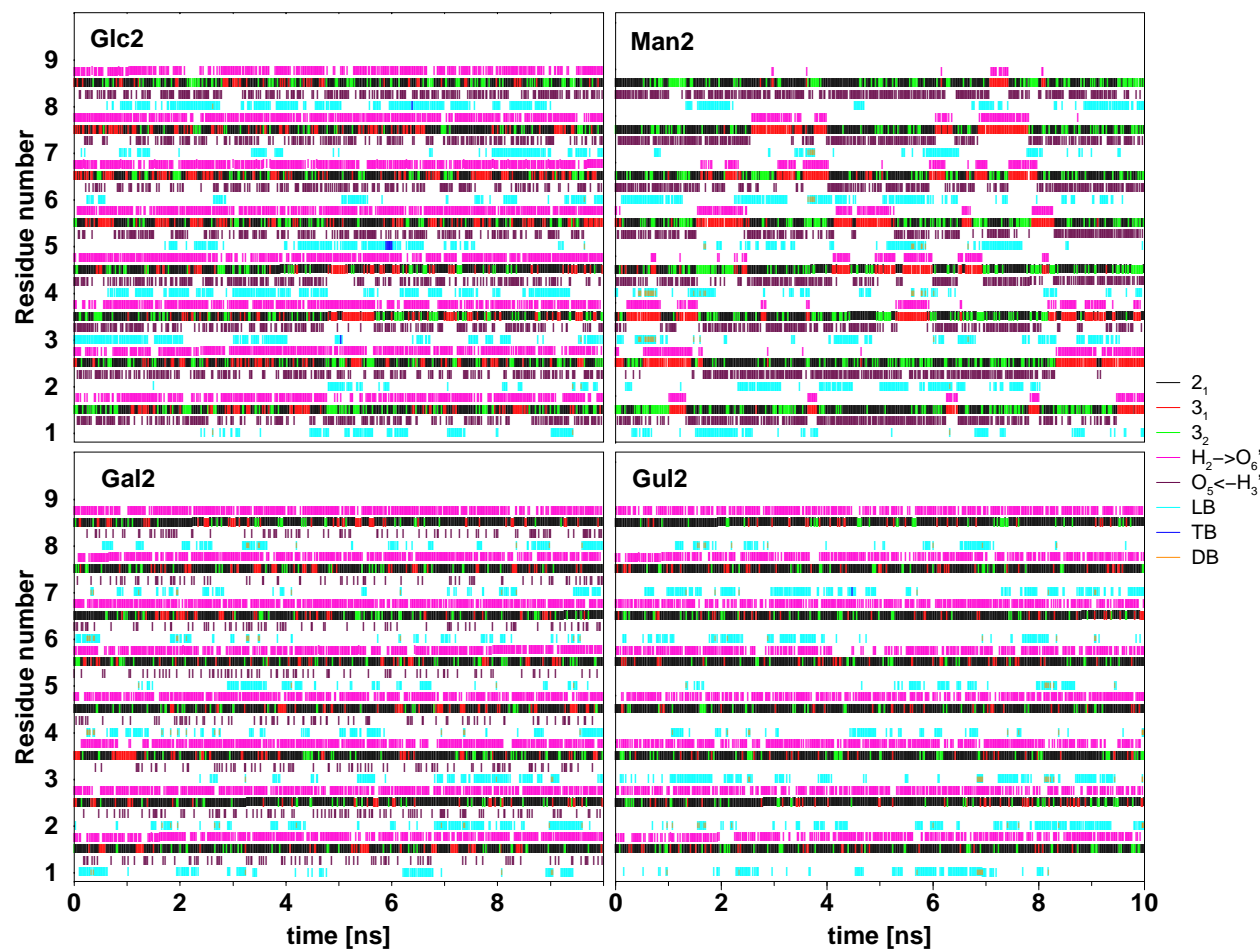


Figure 4.8: Time evolution of the local helical conformations and of the inter-residue hydrogen bonds for the N_{res} glycosidic linkages, as well as of ion binding events for the N_{res} residues, displayed for simulations Glc2, Man2, Gal2 and Gul2 (Table 4.2). The indicated events are the formation of a local 2_1 -, 3_1 - or 3_2 -helical conformation, of ($H_2 \rightarrow O_6'$ or $O_5 \leftarrow H_3'$) hydrogen bonds, of a solvent-separated (loose binding; LB) ion pair, of a contact (tight binding; TB) ion pair or of an ion pair (LB or TB) involving an ion also bound (LB or TB) to the carboxylate group of another residue (double binding; DB; smaller bars for the two involved carboxylate groups). See Section 4.3.2 for definitions of local helical conformations, hydrogen bonding and ion binding. The sampling period (individual bars) is 20 ps. Color code 2_1 (black), 3_1 (red), 3_2 (green), $H_2 \rightarrow O_6'$ (pink), $O_5 \leftarrow H_3'$ (brown), LB (cyan), TB (blue), DB (orange).

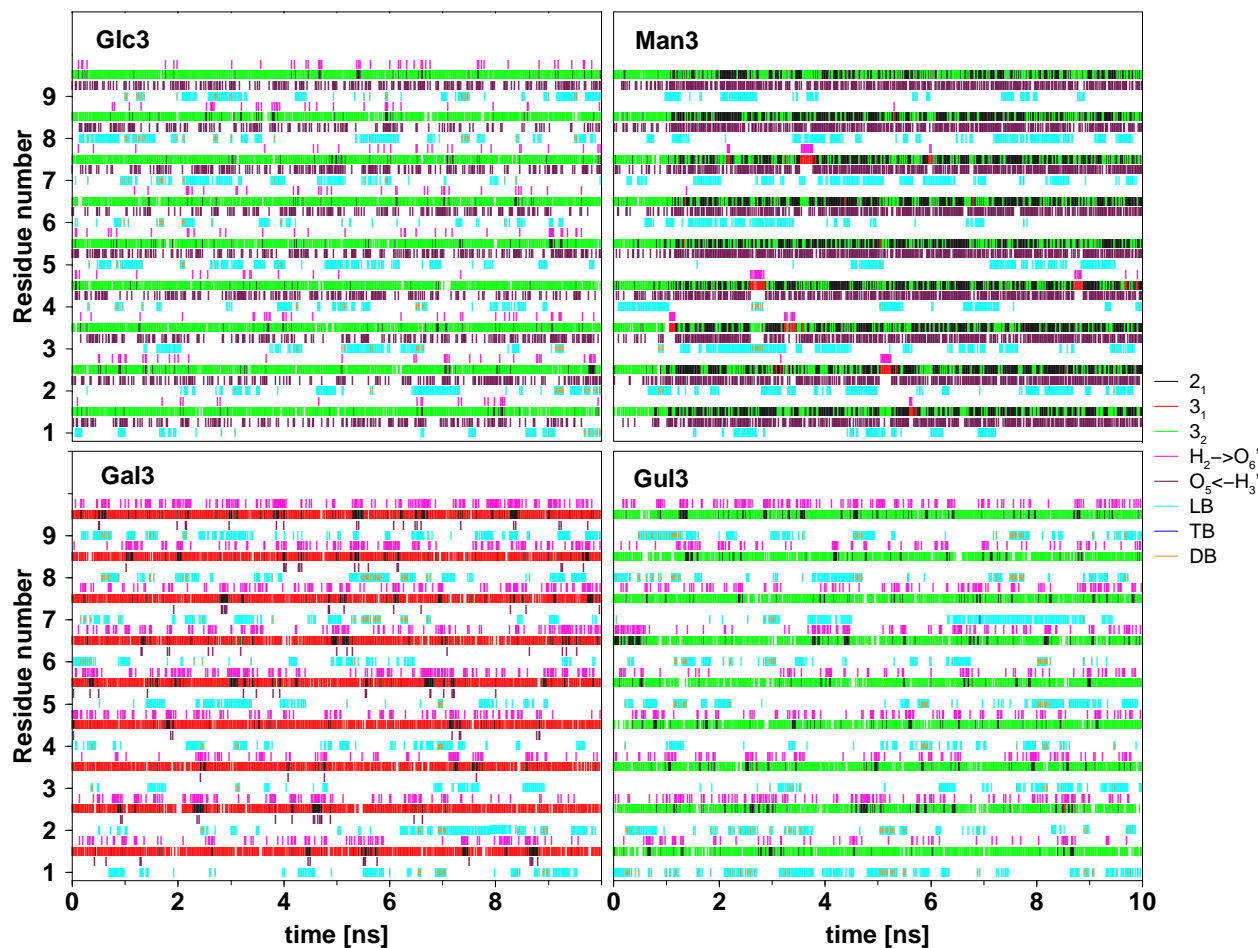


Figure 4.9: Time evolution of the local helical conformations and of the inter-residue hydrogen bonds for the N_{res} glycosidic linkages, as well as of ion binding events for the N_{res} residues, displayed for simulations *Glc3*, *Man3*, *Gal3* and *Gul3* (Table 4.2). The indicated events are the formation of a local 2_1 -, 3_1 - or 3_2 -helical conformation, of ($H_2 \rightarrow O_6'$ or $O_5 \leftarrow H_3'$) hydrogen bonds, of a solvent-separated (loose binding; LB) ion pair, of a contact (tight binding; TB) ion pair or of an ion pair (LB or TB) involving an ion also bound (LB or TB) to the carboxylate group of another residue (double binding; DB; smaller bars for the two involved carboxylate groups). See Section 4.3.2 for definitions of local helical conformations, hydrogen bonding and ion binding. The sampling period (individual bars) is 20 ps. Color code 2_1 (black), 3_1 (red), 3_2 (green), $H_2 \rightarrow O_6'$ (pink), $O_5 \leftarrow H_3'$ (brown), LB (cyan), TB (blue), DB (orange).

(Table 4.4), the occurrences of the three types of local helical conformations changes little (Table 4.6; by up to about 10%) and the populations of regular helices (Table 4.5) decrease by about a factor 2-3. The nature of the counter-ions for pGulU at 300 K has essentially no influence on any of the observables discussed in this section.

4.4.4 Intramolecular hydrogen bonding

The four homopolyuronates may present two types of inter-residue hydrogen bonds across the glycosidic linkage: $H_2 \rightarrow O'_6$ and $O_5 \leftarrow H'_3$ (Figure 4.2). Due to the axial orientation of the 2-OH group in pManU, the former hydrogen bond is incompatible with a regular 2_1 -helical conformation (although it may occur in other conformations). Similarly, the orientation of the 3-OH group in pGulU prevents the formation of the second type of hydrogen bond in regular two- and three-fold helices. In an aqueous environment, the actual occurrences of these inter-residue hydrogen bonds are expected to depend on the local chain conformation. In the solid state, where regular helices are observed, the reported crystallographic structures are compatible with $H_2 \rightarrow O'_6$ hydrogen bonds for pGlcU¹⁸ (2_1 -helix; presumed by analogy with the cellulose II structure³³), pGalU⁵³ (3_1 -helix) and pGulU⁴²⁻⁴⁶ (2_1 -helix), and with $O_5 \leftarrow H'_3$ hydrogen bonds for pGlcU¹⁸ (2_1 -helix; presumed) and pManU^{42-44,46} (2_1 -helix).

In the sixteen simulations, only the two above types of intramolecular hydrogen bonds were observed. The time evolution of these hydrogen bonds at the level of the N_{res} glycosidic linkages is illustrated in Figure 4.8 and 4.9 for the eight main simulations. The corresponding occurrences and lifetimes, averaged over the N_{res} linkages (together with the associated standard deviations), are reported in Table 4.6 for all simulations.

Focusing on the eight main simulations, the following qualitative observations can be made. The $H_2 \rightarrow O'_6$ hydrogen bond is only compatible (*i.e.* present with a significant occurrence) with the following local helical conformations: 2_1 and 3_1 for pGlcU; 3_1 for pManU; 2_1 , 3_2 and (to a limited extent) 3_1 for pGalU; 2_1 and (to a limited extent) 3_2 for pGulU. As a consequence of the populations of these local helical conformations in the different simulations, the occurrence of the $H_2 \rightarrow O'_6$ hydrogen bond is high for Glc2, Gal2 and Gul2 (about 70-80%), moderate for Man2, Gal3 and Gul3 (about 20-25%), and low for Glc3 and Man3 (about 6% or less). The $O_5 \leftarrow H'_3$ hydrogen bond is only compatible with the following local helical conformations: 2_1 and (to a limited extent) 3_2 for pGlcU; 2_1 and (to a limited extent) 3_2 for pManU; 2_1 (to a limited extent) for pGalU; none for pGulU. As a consequence of the populations of these local helical conformations in the different simulations, the occurrence of the $O_5 \leftarrow H'_3$ hydrogen bond is high for Glc2, Man2 and Man3 (about 45-75%), moderate for Gal2 and Glc3 (15-30%), and low for Gal3, Gul2 and Gul3 (about 3% or less). The lifetimes of the two types of hydrogen bonds range from 26 to 269 ps for the eight main simulation, the exact values being largely correlated with the corresponding occurrences. The occurrences of the two types of hydrogen bonds in simulations Glc2, Man2, Gal3 and Gul2 agree very well with the hydrogen bonding patterns suggested by the corresponding solid-state structures obtained from fiber diffraction measurements^{18,42-44,44-46,53}. Assuming that hydrogen bonding is an important conformational driving force⁸³, it is reasonable to expect that the formation of $H_2 \rightarrow O'_6$ inter-residue hydrogen bonds affects the rotameric distribution and isomerization timescale of the γ angle characterizing the orientation of the carboxylate groups (Section 4.4.1). However, the present simulations do not evidence clearly this presumed influence. In particular, the γ distribution is shifted by about 10° and becomes narrower for pGulU and

| Code | 2 ₁ | 3 ₁ | 3 ₂ | H ₂ →O' ₆ | O ₅ ←H' ₃ | LB | TB | DB |
|---------------------------------|----------------|----------------|----------------|---------------------------------|---------------------------------|-------------|------------|-----------|
| occurrences [%] | | | | | | | | |
| Glc2 | 66.1 (2.9) | 19.3 (4.6) | 14.6 (3.8) | 79.2 (4.9) | 45.9 (3.7) | 31.9 (15.2) | 0.3 (0.5) | 0.3 (0.2) |
| Man2 | 57.5 (5.4) | 20.7 (7.1) | 20.8 (2.9) | 19.5 (6.8) | 55.9 (5.7) | 26.4 (7.2) | - | 1.2 (0.7) |
| Gal2 | 81.4 (2.7) | 11.4 (2.1) | 7.2 (1.1) | 80.0 (1.4) | 17.9 (1.8) | 23.4 (5.1) | - | 1.5 (0.7) |
| Gul2 | 85.0 (1.4) | 6.7 (1.2) | 8.2 (1.4) | 69.5 (2.5) | - | 27.1 (9.1) | - | 2.4 (0.7) |
| Glc3 | 6.1 (1.2) | - | 87.9 (1.9) | 5.9 (1.5) | 28.3 (3.9) | 25.5 (7.3) | - | 4.1 (1.5) |
| Man3 | 66.2 (3.0) | 2.0 (2.1) | 31.6 (3.9) | 1.6 (1.8) | 74.7 (1.7) | 30.2 (9.8) | - | 0.7 (0.9) |
| Gal3 | 11.2 (3.6) | 80.9 (1.8) | - | 26.2 (5.5) | 2.6 (0.8) | 25.6 (6.8) | - | 3.9 (2.7) |
| Gul3 | 11.9 (3.8) | - | 79.0 (2.6) | 21.8 (5.7) | - | 29.3 (7.8) | - | 6.3 (2.3) |
| Gul2 _{N_a} | 83.5 (1.4) | 7.7 (1.1) | 8.7 (0.8) | 68.3 (3.1) | - | 28.5 (7.5) | 0.6 (1.0) | 4.1 (3.3) |
| Gul2 _{LJ-} | 84.8 (1.9) | 7.1 (1.4) | 8.1 (1.4) | 68.5 (1.3) | - | 28.5 (13.4) | - | 1.3 (0.8) |
| Gul2 _{LJ+} | 83.7 (1.9) | 7.6 (1.5) | 8.6 (1.7) | 68.0 (2.2) | - | 28.6 (3.7) | - | 4.3 (1.7) |
| Glc2H | 59.8 (1.2) | 20.8 (1.6) | 18.9 (1.2) | 71.9 (1.4) | 34.0 (0.9) | 34.8 (3.0) | 7.9 (2.3) | 1.9 (0.6) |
| Man2H | 37.7 (1.2) | 28.8 (2.3) | 30.7 (2.4) | 25.3 (2.1) | 24.8 (0.9) | 41.2 (4.6) | 13.2 (4.2) | 6.6 (0.8) |
| Gal2H | 72.6 (1.0) | 14.2 (0.8) | 13.0 (0.9) | 66.9 (0.8) | 17.3 (0.6) | 38.8 (2.3) | 7.9 (1.7) | 6.3 (0.8) |
| Gul2H | 72.6 (0.9) | 1.5 (0.9) | 13.8 (0.8) | 60.7 (1.1) | - | 37.4 (4.9) | 7.6 (2.7) | 5.3 (0.9) |
| Gul2 _{N_a} H | 73.0 (1.1) | 13.2 (0.7) | 13.7 (0.8) | 61.3 (1.1) | - | 35.6 (2.0) | 7.2 (0.9) | 5.5 (0.6) |
| lifetime [ps] | | | | | | | | |
| Glc2 | 143 (26) | 55 (9) | 38 (6) | 269 (56) | 100 (13) | 146 (28) | 21 (46) | 19 (8) |
| Man2 | 172 (44) | 181 (66) | 47 (6) | 177 (41) | 220 (50) | 122 (27) | - | 34 (18) |
| Gal2 | 318 (93) | 44 (8) | 26 (2) | 234 (26) | 34 (3) | 93 (23) | - | 29 (8) |
| Gul2 | 448 (64) | 25 (3) | 31 (7) | 167 (15) | - | 100 (19) | - | 32 (6) |
| Glc3 | 28 (4) | - | 1097 (418) | 25 (4) | 77 (12) | 99 (21) | - | 40 (10) |
| Man3 | 190 (26) | 51 (47) | 50 (8) | 56 (53) | 444 (107) | 129 (17) | - | 17 (14) |
| Gal3 | 46 (6) | 361 (89) | - | 54 (7) | 26 (3) | 101 (28) | - | 48 (13) |
| Gul3 | 39 (5) | - | 303 (57) | 43 (4) | - | 103 (34) | - | 46 (10) |
| Gul2 _{N_a} | 405 (76) | 29 (5) | 33 (3) | 175 (25) | - | 70 (14) | 25 (49) | 39 (12) |
| Gul2 _{LJ-} | 399 (88) | 28 (4) | 30 (5) | 164 (23) | - | 154 (58) | - | 23 (13) |
| Gul2 _{LJ+} | 428 (52) | 26 (2) | 31 (4) | 160 (12) | - | 93 (20) | - | 40 (13) |
| Glc2H | 211 (26) | 32 (1) | 30 (1) | 336 (98) | 60 (3) | 67 (10) | 50 (9) | 21 (5) |
| Man2H | 58 (3) | 48 (5) | 51 (7) | 47 (4) | 49 (4) | 76 (9) | 59 (16) | 27 (4) |
| Gal2H | 524 (141) | 24 (1) | 25 (2) | 522 (116) | 33 (3) | 64 (6) | 46 (5) | 23 (4) |
| Gul2H | 748 (267) | 25 (1) | 25 (2) | 322 (56) | - | 62 (8) | 48 (11) | 22 (3) |
| Gul2 _{N_a} H | 738 (161) | 24 (1) | 24 (1) | 314 (48) | - | 45 (3) | 24 (3) | 21 (2) |

Table 4.6: Local helical conformations, intramolecular hydrogen bonding and ion binding. The occurrences and lifetimes of local helical conformations, intramolecular hydrogen bonds (only observed to occur between adjacent residues) and ion binding, averaged over the N_{res} linkages (two former types of observables) or the N_{res} residues (binding events) are reported together with the corresponding standard deviations over the N_{res} individual values (indicated between parentheses) for the different simulations (Table 4.2). The indicated properties are the formation of a local 2₁-, 3₁- or 3₂-helical conformation, of (H₂→O'₆ and O₅←H'₃) hydrogen bonds, of a solvent-separated (loose binding; LB) ion pair, of a contact (tight binding; TB) ion pair, or of an ion pair (LB or TB) involving an ion also bound (LB or TB) to the carboxylate group of the another residue (double binding; DB). The percentages of the three local helical conformations do not add up to 100%, due to the (limited) occurrence of other conformations (with θ outside the range [90; 270°]). These other conformations also influence the lifetimes of the three conformations. See Section 4.3.2 for the definitions of local helical conformations, hydrogen bonding and ion binding. Averaging was performed over the final 9 ns of the simulations.

pGalU when changing from an initial two-fold to an initial three-fold helical model (Figure 4.3), although the occurrence of the H₂→O'₆ hydrogen bond concomitantly decreases (by about 50-55

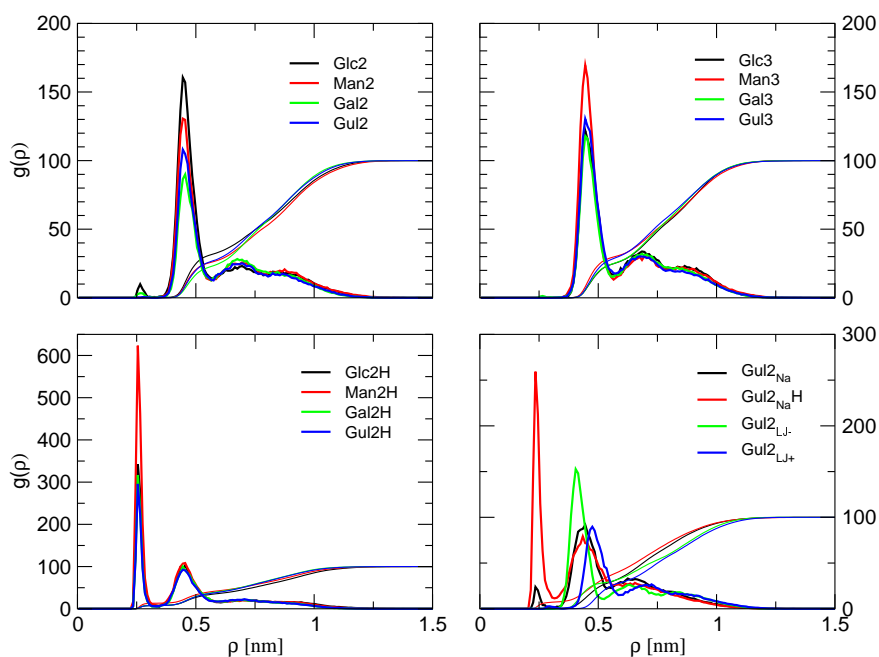


Figure 4.10: Radial distribution function $g(\rho)$ for the smallest (minimum-image) distance between any of the two carboxylate oxygen atoms of a given residue and the closest positive counter-ion, averaged over the N_{res} residues, for the different simulations (Table 4.2). The cumulative integral of $V^{-1}4\pi\rho^2g(\rho)$ (amplified by 100 for readability) is also displayed. See Section 4.3.2 for the definitions of $g(\rho)$ and V . Averaging was performed over the final 9 ns of the simulations.

%). However, no similar change is seen in the γ distribution for pGlcU and pManU, although the occurrence of this hydrogen bond also significantly decreases (by about 75 and 20%, respectively). Moreover, the rotational isomerisation timescale τ_γ is consistently at least an order of magnitude longer for pGlcU and pManU compared to pGalU and pGulU (Table 4.3), an observation that cannot be correlated with the occurrences of the $H_2 \rightarrow O'_6$ hydrogen bond in the eight simulations.

The time evolution of the inter-residue hydrogen bonds for the eight simulation variants is illustrated in Figures 4.23 and 4.24. Expectedly, the formation and disruption of hydrogen bonds is much more rapid at 500 K compared to 300 K (Figure 4.8). Surprisingly, however, the occurrences of the two types of hydrogen bonds (Table 4.6) are not dramatically affected by the increase in temperature. A moderate decrease (by about 10-15%) is observed for both types of hydrogen bonds, except in simulation Man2H. The nature of the counter-ions for pGulU at 300 K has essentially no influence on inter-residue hydrogen bonding.

4.4.5 Ion binding

The radial distribution functions $g(\rho)$ between carboxylate oxygen atoms and the closest counter-ion, averaged over the N_{res} residues, are displayed in Figure 4.10 for all simulations (together with the corresponding integrals). Focusing on the twelve simulations involving Ca^{2+} counter-ions (at

either 300 or 500 K), at least two density peaks can be seen in the neighborhood of the carboxylate groups, centered at calcium-oxygen distances of 0.26 and 0.45 nm (at 300 K, a third peak is visible at 0.68 nm). These two first peaks correspond to contact (tight binding; TB) and solvent-separated (loose binding; LB) ion pairs. The two types of events are illustrated in Figure 4.11 (top) based on single frames extracted from the Gal2 simulation. For the two simulations involving Na^+ counter-ions, the corresponding peak positions are 0.23 and 0.44 nm (at 300 K, a third peak is visible at 0.64 nm). For the simulations performed using Ca^{2+} -like counter-ions with their effective radii decreased (increased) by 20%, the positions of these first two peaks are, as expected, shifted to lower (higher) distances relative to the standard Ca^{2+} ion. To quantify the occurrences of LB and TB events, cutoff criteria were applied to the calcium-oxygen distance (Section 4.3.2), with cutoff values corresponding to the first and second minima in the $g(\rho)$ curves.

The occurrences and lifetimes of LB and TB events, averaged over the N_{res} residues, are reported in Table 4.6 for all simulations. Maybe the most unexpected result of this analysis (see also Figure 4.10) is the very low occurrences of TB events at 300 K (<1% for simulations Glc2 and Gul2_{Na}; none for the others). Although the sugar carboxylate groups are commonly found in the second solvation shells of the counter-ions at this temperature (LB; occurrences of about 25-30%), they almost never penetrate into their first solvation shells. The two simulations with Ca^{2+} -like counter-ions were undertaken to ascertain that this observation was not the consequence of an improper choice of ion-solvent Lennard-Jones interaction parameters, since there is room for ambiguity concerning the appropriate way to calibrate these parameters^{106,106}. However, the use of a Ca^{2+} -like ion with its effective radius decreased or increased by 20% did not change the above picture qualitatively. The simulations also failed to evidence any striking qualitative difference between the coordination of Ca^{2+} and Na^+ in terms of $g(\rho)$ and TB events, in spite of the different ionic charge. A second counter-intuitive observation is that an elevated temperature of 500 K enhances ionic binding, in spite of the increased thermal motion (the occurrences of LB and TB events increasing from about 25-30% and 0%, respectively, to about 35-40% and 8-13%, respectively). This observation is probably related to the decrease in the dielectric permittivity of water upon increasing the temperature, leading to a reduction of solvation effects. Since dielectric screening opposes favorable electrostatic interactions between carboxylate groups and counter-ions, this is expected (and observed) to induce a tighter binding of the counter-ions to the chain. Independent simulations of a pure SPC water system (data not shown) evidenced a decrease in the permittivity (evaluated using the appropriate dipole-moment fluctuation formula⁸⁹) from 61 to 34 upon increasing the temperature from 300 to 500 K. Experimentally, the permittivity of water decreases from 78.3 to 55.7 upon increasing the temperature from 298.15 to 398.15 K¹⁰⁷. Such an attenuation of dielectric screening upon increasing the temperature, leading to an enhanced influence of direct solute-solute or solute-ion electrostatic interactions, has been evidenced previously¹⁰⁸.

The lifetimes of LB events are of the order of 100-150 ps for Ca^{2+} (70 ps for Na^+) at 300 K, and somewhat shorter (60-70 ps) for Ca^{2+} at 500 K. The lifetimes of TB events are about 20 ps for both Ca^{2+} and Na^+ at 300 K, and somewhat longer for Ca^{2+} (about 50-60 ps) and Na^+ (about 25 ps) at 500 K.

Double binding (DB; ion bridging) events (*i.e.* carboxylate groups binding an ion that is simultaneously bound to the carboxylate group of another residue) were relatively infrequent during the simulations, the corresponding occurrences ranging from 0.3 to 6.3% (with lifetimes of the order of 20-50 ps) for the main simulations (Table 4.6). These occurrences also slightly increased

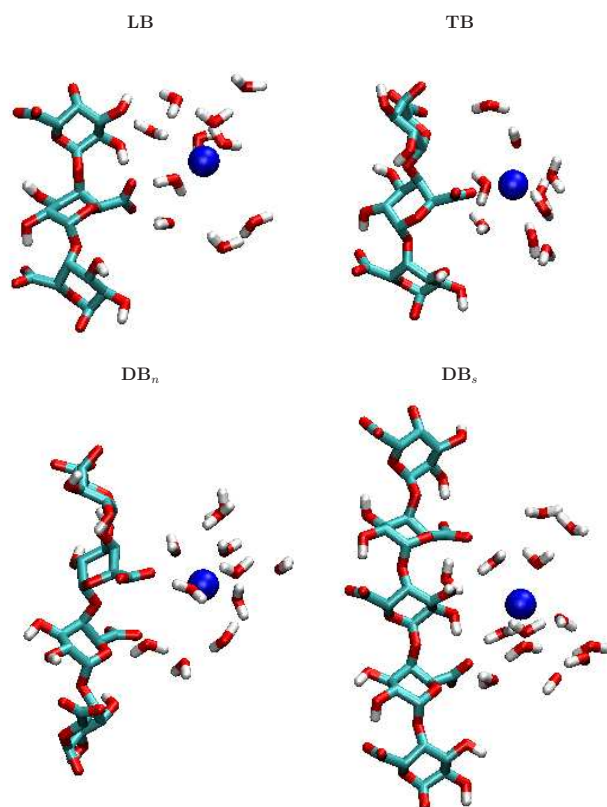


Figure 4.11: Snapshots from the Gal2 simulation. The configurations illustrate the following events: loose binding (LB; $t=1.664$ ns, residues 3-5, $\rho=0.43$ nm); tight binding (TB; $t=5.700$ ns, residues 2-4, $\rho=0.26$ nm); double binding involving neighbor residues (DB_n ; $t=3.859$ ns, residues 1-4, $\rho=0.46$ and 0.50 nm relative to the carboxylate groups of residues 2 and 3, respectively); double binding involving second-neighbor residues (DB_s ; $t=4.094$ ns; residues 4-8, $\rho=0.50$ and 0.48 nm relative to the carboxylate groups of residues 4 and 7, respectively). The orientation (top to bottom of each drawing) is from the non-reducing to the reducing end. Only water molecules with their oxygen atoms closer than 0.6 nm from the counter-ion are displayed.

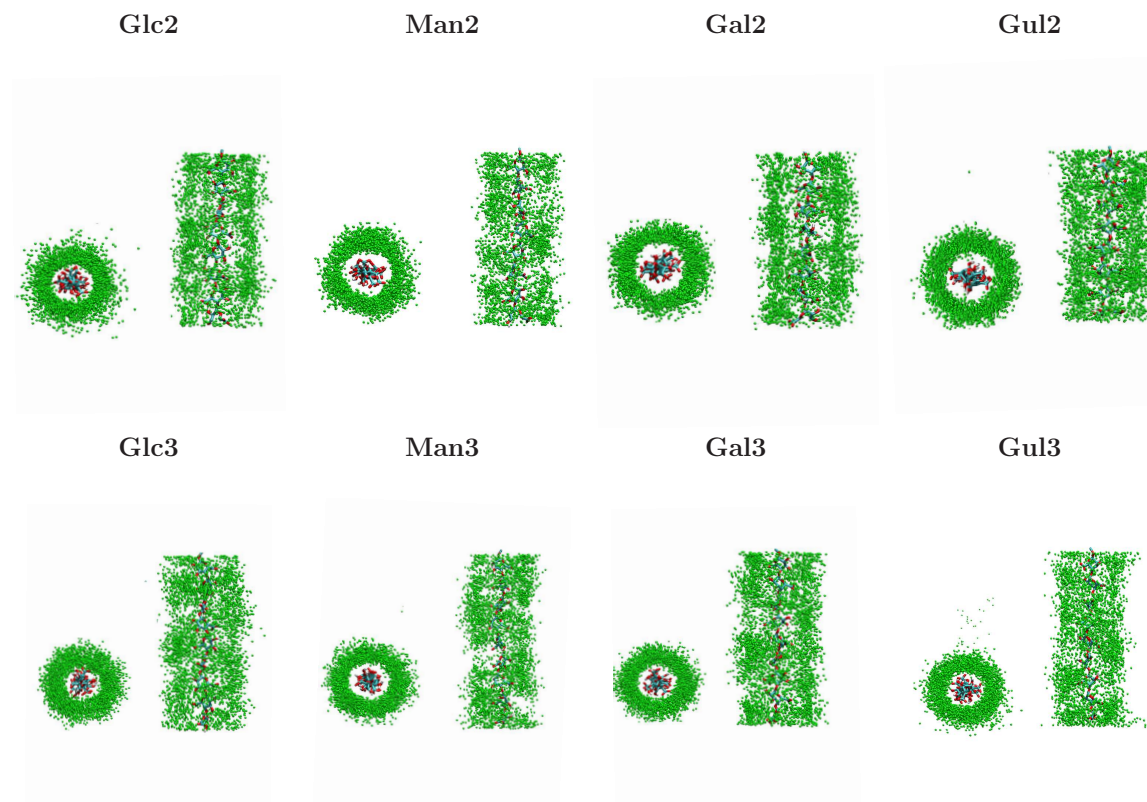


Figure 4.12: Positions of the counter-ions along the chain (top and side views), as sampled along the eight main simulations (300 K, Ca^{2+} counter-ions; Table 4.2). The ion coordinates correspond to successive configurations superimposed onto the initial (after 0.2 ns equilibration) one (rotational and translational fit based on the C_2 , C_3 , C_5 and O_5 atoms). The latter reference configuration is also displayed. The sampling period (between displayed ion coordinates) is 10 ps. The figures correspond to the final 9 ns of the simulations.

when raising the temperature from 300 to 500 K. Triple binding events (Ca^{2+} bound to three carboxylate groups) were observed but with negligible occurrences ($<0.3\%$ in all simulations). All DB events were found to involve either neighbor (DB_n) or second-neighbor (DB_s) residues along the chains. Illustrations for those types of events are provided in Figure 4.11 (bottom). At 300 K, DB_n events represented the leading contribution to DB events for simulations Glc2 and Man2 (80-90%), as well as for all simulations initiated from a three-fold helical model (95-99%). In contrast, DB_s events were more frequent in simulations Gal2 and Gul2 (80-85%). These differences are probably related to differences in the preferred values for the local turn angle and helix raise (determining the spatial proximity between neighbor and second-neighbor carboxylate groups) and in the glycosidic linkage flexibility (*e.g.* DB_n events in a 2_1 -helix must be associated with a significant distortion of the local helical conformation; Figure 4.11).

The time evolution of calcium binding events is illustrated in Figure 4.8 and 4.9 as well as 4.23 and 4.24 for all simulations. These graphs do not reveal any obvious correlation between ion binding events, local helical conformation and inter-residue hydrogen bonding. Note also that since TB and DB events have so limited occurrences at 300 K, it is unlikely that calcium binding has a significant influence on the conformational preferences and rotational isomerization timescale of the carboxylate groups (Section 4.4.1).

The distributions of the positive counter-ions around the chains (sampled at 10 ps intervals along the trajectories, excluding the initial 1 ns of the simulations), are illustrated in Figure 4.12 for the eight main simulations. Analogous illustrations for the simulation variants are provided in Figure 4.25. The associated two-dimensional radial distribution functions $g_{2D}(r)$ for the ion coordinates in the xy -plane relative to the helix axis are provided in Figure 4.26 for all simulations. These figures further point towards a lack of specific binding of the counter-ions to the chain, irrespective of the system considered (polyuronate type, initial helical model, temperature and nature of the counter-ions). In all cases, the counter-ion atmosphere is almost entirely confined to a cylindrical region around the chain axis. Although the counter-ion density within this region exhibits some broad patterns, the absence of well-defined binding sites remains evident. For the Ca^{2+} and Ca^{2+} -like ions, the cylinder radii accounting for 95% (or 99%) of the counter-ion density range from 1.0 to 1.1 (or 1.1 to 1.2) nm irrespective of the temperature. For the Na^+ ions, the counter-ion atmosphere is slightly more diffuse, but 90% of the counter-ion density is still contained within cylinders of 1.8 (300 K) and 2.2 (500 K) nm radii. Note that the observed confinement could possibly be enhanced by the artificial periodicity of the system in the xy -plane (*i.e.* the chains are surrounded in this plane by an infinite lattice of periodic replicas with a spacing of about 5 nm; Table 4.2). However, the absence of visible anisotropy in the counter-ion distributions (with the possible exception of simulations Gul2_{Na} and Gul2_{NaH}) and the nearly complete charge cancellation after a distance of about 1.2 nm (for Ca^{2+} and Ca^{2+} -like ions; effectively excluding the presence of long-range electrostatic interactions between the screened chain and its periodic copies) suggest that this effect is not significant. A change in temperature from 300 to 500 K promotes a slight contraction of the counter-ion atmosphere around the chain, in agreement with the previous observations concerning LB, TB and DB events. The corresponding peak positions in $g_{2D}(r)$ are shifted to lower values by about 0.15 nm for both Ca^{2+} and Na^+ counter-ions. Finally, it should be stressed that all the above observations are in principle restricted to the counter-ion concentration ranges and solution ionic strengths investigated in the present simulations (neutralizing amount; no other salt in solution).

The fact that the counter-ions do not show strong binding to the carboxylate groups renders

a tight interaction with the other sugar oxygen atoms unlikely. Radial distribution functions for the distances between the oxygen atoms O_1 (glycosidic oxygen), O_2 and O_3 (hydroxyl groups), or O_5 (ring oxygen) and the closest counter-ion are displayed in Figures 4.27–4.30 for all simulations. The corresponding occurrences of LB and TB events, based on appropriate cutoff criteria (Section 4.3.2), are reported in Table 4.9. Focusing on the eight main simulations, LB events are infrequent (at most about 15%) and TB events exceptional (<2% for simulations Glc3 and Gul3; none for the others). At 500 K, the occurrences of LB events increase to about 10-20% while those of TB events remain very low. The glycosidic oxygen and, to a lesser extent, the ring oxygen atoms appear to present a somewhat lower affinity for the counter-ions compared to the two hydroxyl oxygen atoms. Finally, Na^+ is seen to have a slightly higher propensity for coordination to the sugar oxygen atoms compared to Ca^{2+} .

4.4.6 Diffusion properties

The mean-square displacements Q_z , Q_{xy} and Q_Θ associated with the chain longitudinal, transverse and rotational rigid-body motions are displayed in Figure 4.31 in the form of averages over different sets of simulations. The mean-square displacements $Q_{i,z}$ and $Q_{i,xy}$ associated with the counter-ion coordinates along the z -axis and in the xy -plane, averaged over the N_{ion} positive ions, are displayed in Figure 4.32 in the same form. The corresponding diffusion coefficients D_z , D_{xy} , D_Θ , $D_{i,z}$ and $D_{i,xy}$ are reported in Table 4.7, while the values calculated for all individual simulations are provided in Table 4.10. However, because the individual mean-square-displacement curves significantly deviated from linearity the latter values may not be well converged. For this reason, the results are only discussed here in terms of averages over four sets of simulations: (i) all simulations at 300 K with Ca^{2+} counter-ions; (ii) the corresponding simulations initiated from a 2_1 -helical structure; (iii) the corresponding simulations initiated from a three-fold (3_1 - or 3_2 -) helical structure; (iv) all simulations at 500 K with Ca^{2+} counter-ions. The mean-square displacement curves averaged over the four above sets of simulations were found to be essentially linear over the interval from 0 to 3 ns, with the possible exception of the transverse mean-square displacement $Q_{i,xy}$ of the ions. Here, the curves showed sublinear (non-Einstein) behavior, especially visible at 500 K. This is probably a consequence of the confinement of the counter-ion atmosphere to a cylindrical region around the chain (Section 4.4.5).

Focusing on the eight main simulations, the diffusion coefficients D_z , D_{xy} and D_Θ associated with the chain rigid-body motion at 300 K are $0.85 \text{ nm}^2 \cdot \text{ns}^{-1}$, $0.38 \text{ nm}^2 \cdot \text{ns}^{-1}$ and $2.45 \text{ deg}^2 \cdot \text{ns}^{-1}$. Expectedly, the chain diffuses faster along its axis than transversally (by about a factor of two), which is related to the larger effective cross-section (exposed to solvent-induced friction) associated with the latter type of motion. The corresponding expected (root-mean-square) displacements over the 10 ns simulations are 4.1 nm, 3.9 nm (the corresponding expected displacements along x or y separately are 2.8 nm) and 7.0° . When comparing the sets of simulations initiated from two- and three-fold helices at 300 K, it appears that the former chains are somewhat more diffusive in all three degrees of freedom compared to the latter ones (by about 20-40%). This observation is probably related to the larger effective cross-section associated with three-fold helices compared to their two-fold analogs. Finally, as expected, the chain diffusion for pGulU (2_1 helix) is significantly faster at 500 K (by about a factor 3-7).

Again focusing on the main simulations, the average diffusion coefficients $D_{i,z}$ and $D_{i,xy}$ associated with the longitudinal and transverse motions of the counter-ions are 0.43 and 0.19

| $D_{\mathbf{q}}$ | 300 K (all) | 300 K (2-fold) | 300 K (3-fold) | 500K (2-fold) |
|--|-------------|----------------|----------------|---------------|
| D_z [$\text{nm}^2 \cdot \text{ns}^{-1}$] | 0.85 (0.29) | 0.92 (0.64) | 0.78 (0.24) | 3.26 (2.30) |
| D_{xy} [$\text{nm}^2 \cdot \text{ns}^{-1}$] | 0.38 (0.07) | 0.44 (0.13) | 0.31 (0.07) | 1.44 (0.51) |
| D_{θ} [$\text{deg}^2 \cdot \text{ns}^{-1}$] | 2.45 (0.55) | 2.68 (0.75) | 2.21 (1.02) | 18.12 (5.23) |
| $D_{i,z}$ [$\text{nm}^2 \cdot \text{ns}^{-1}$] | 0.41 (0.09) | 0.41 (0.12) | 0.40 (0.18) | 3.02 (1.99) |
| $D_{i,xy}$ [$\text{nm}^2 \cdot \text{ns}^{-1}$] | 0.19 (0.05) | 0.14 (0.02) | 0.24 (0.10) | 0.05 (0.01) |

Table 4.7: Diffusion constants evaluated for different sets of degrees of freedom of the chain and of the counter-ions. The results (together with error estimates; between parentheses) are reported in the form of averages over different sets of N_{sim} simulations (Table 4.2). The diffusion coefficients $D_{\mathbf{q}}$ are calculated from a linear least-square-fit of the corresponding mean-square displacements $Q_{\mathbf{q}}(t)$ over the time period from 0 to 3 ns. The chain diffusion is described in terms of the longitudinal (D_z), transverse (D_{xy}) and rotational (D_{θ}) diffusion constants. The ionic diffusion is described in terms the ionic longitudinal ($D_{i,z}$) and transverse ($D_{i,xy}$) diffusion constants, averaged over the N_{ions} positive counter-ions. The error estimate corresponds to the standard deviation of the calculated values over the N_{sim} simulations divided by $(N_{sim} - 1)^{1/2}$. The simulation sets correspond to all simulations at 300 K with Ca^{2+} counter-ions ($N_{sim}=8$), those started from two-fold ($N_{sim}=4$) or three-fold ($N_{sim}=4$) helical structure, or all simulations at 500 K with Ca^{2+} counter-ions ($N_{sim}=4$; pGulU; initiated from a two-fold helical structure). See Section 4.3.2 for the definition of $Q_{\mathbf{q}}(t)$ and $D_{\mathbf{q}}$. The calculation was performed considering the final 9 ns of the simulations and all possible time origins.

$\text{nm}^2 \cdot \text{ns}^{-1}$. The latter value is somewhat uncertain because the mean-square displacement is sub-linear in time (see above). This observation, together with the comparison of the two numbers suggests that the ions diffuse normally along the chain axis, while they are restricted in their transverse motion by confinement within the counter-ion atmosphere. The differences between the sets of simulations initiated from two- and three-fold helices at 300 K are probably not significant. Finally, although the longitudinal diffusion of the counter-ions around pGulU (2_1 helix) is significantly faster at 500 K (by about a factor 3), the opposite is true for the transverse diffusion (reduced by about a factor 3). The reasons for these two observations are probably the enhanced thermal motion (increase in $D_{i,z}$) and the contraction of the counter-ion atmosphere around the chain (decrease in $D_{i,xy}$) at elevated temperature. The latter effect is particularly visible in the corresponding mean-square displacement curve at 500 K, which reaches a plateau value of 1 nm^2 (comparable with the approximate square radii of the confinement cylinders; see above) after about 0.5 ns, implying that at this temperature, the ions are essentially unable to escape the close neighborhood of the chain (Section 4.4.5).

4.5 Conclusion

It is important to realize that the present simulations account for a somewhat "unphysical" situation (even at 300 K in the presence of Ca^{2+} or Na^+ counter-ions). Because artificial periodicity is applied along the chain axis based on repeat units (8 or 9 residues) that are significantly shorter than the persistence lengths of the polymers expressed on a per residue basis (about 20 for pGlcU and pManU; about 40 for pGalU and pGulU), it enforces an increased extent of regularity com-

pared to the experimental situation. Furthermore, the consideration of single chains excludes the possibility of dimer formation, which would occur experimentally in the presence of Ca^{2+} counter-ions at the concentration considered. In this sense, the simulations probe locally ordered states that may transiently exist (with limited equilibrium populations) within more disordered chains in the aqueous environment. However, the consideration of these more ordered states is justified by the fact that these are likely to represent the productive conformations for chain-chain association in junction zones and the formation of gels.

The main conclusions of the present study can be summarized as follows:

- A. Although regular helical structures, as inferred from fiber diffraction measurements in the solid state^{18,42–44,44–47,53} or built using molecular modeling procedures^{19,55} are useful conceptual models, they may represent a rather misleading picture of the complex conformational ensembles populated by single polyuronate chains in aqueous solution. The present simulations reveal an important conformational flexibility (in time) and variability (along the chain) for the glycosidic linkages and local helical parameters. Conformations in the appropriate range for being classified as regular helices only account for a limited fraction of the sampled configurations ($\leq 35\%$), and the helicity of the chain fluctuates locally (*i.e.* non-cooperatively) on a timescale ranging from 20 ps to 1 ns. The extent of regularity is likely to be even more limited in the absence of the (artificial) periodicity constraint employed in the present simulations, as suggested by other modeling studies using Monte Carlo sampling of the accessible glycosidic linkage conformations for long polymer chains^{19,55,57}.
- B. The binding of counter-ions to the polyuronate single chains appears to be largely non-specific irrespective of the type of polyuronate, choice of initial helical model, nature of the counter-ions and temperature. The counter-ion atmosphere is almost entirely confined within a cylinder of high ionic density around the chain axis. No well defined binding sites can be identified, and the differences observed in the modeling study of Braccini *et al.*⁵⁵ between diequatorially- and diaxially-linked polyuronates in terms of ion binding are not reproduced in the present simulations. At 300 K, the carboxylate groups frequently (about 25–30 % occurrences) bind counter-ions, but nearly always remain in their second solvation shells. Events of tight binding, as well as of multiple coordination, are scarce and of short durations. These findings are in line with recent theoretical models suggesting that counter-ion binding to polyuronate single chains is (predominantly) a (non-specific) counter-ion condensation process^{17,69–71}. In the present simulations, the counter-ion atmosphere leads to a nearly complete cancellation of the polymer charge within about 1.2 nm of the chain axis (charge reversal could possibly occur at higher counter-ion concentrations). This charge cancellation may significantly accelerate the kinetics of dimer formation by removal of the long-range Coulombic repulsion between the chains. Interestingly, an increase of the temperature to 500 K promotes a contraction of the counter-ion atmosphere and an increase in the occurrences of tight binding and double binding events, presumably through a decrease of the solvent permittivity (reduced dielectric screening).
- C. The simulations suggest that the helical propensities for single-chain homopoly-uronates in aqueous solution are 3_2 - (pGulU), 9_5 - (pManU), 2_1 - (pGalU) and 2_1 - (pGulU) helices, with the possibility of an alternative 3_1 -helix for pGalU. These forms are representative (in

an average sense; see point A) for the conformational ensembles probed by those of the present simulations at 300 K showing the highest extent of regularity. Although probably associated with limited equilibrium populations in long polyuronate chains, these helical forms may still represent the productive conformations for chain-chain association. Indeed, the suggested helical propensities for pGulU and pGalU are conform with the most popular structural models for junction zones in pectin and alginate gels^{13,28,30,36–38,45}.

- D. Intramolecular hydrogen bonding exclusively involves $H_2 \rightarrow O'_6$ and $O_5 \leftarrow H'_3$ hydrogen bonds across the glycosidic linkages, for all systems considered. Their occurrences in the simulations initiated from the different helical structures agree very well with the hydrogen bonding patterns suggested by the corresponding solid-state structures obtained from fiber diffraction measurements^{18,42–44,44–46,53}. These two hydrogen bonds can only be realized for specific local helical conformations at the level of the corresponding glycosidic linkage (which depend on the type of polyuronate). On the sole basis of the present simulations, it remains difficult to assess whether intramolecular hydrogen bonding represents a major driving force in determining the conformational preferences of the glycosidic linkages, or rather an "opportunistic" consequence of the favorable positioning of two groups given a specific linkage conformation (itself predominantly determined by steric and stereoelectronic effects), as has been suggested to be generally the case for carbohydrates in an aqueous environment⁸³.

Two main features of polyuronate single chains revealed by the present study, namely the important extent of irregularity and conformational flexibility (point A) and the absence of well defined counter-ion binding sites (point B), take a special significance if one compares them to the models proposed for chain-chain association (e.g. egg-box model^{29,36} and more recent alternatives^{38,39,45,46,50–52}). In these models, two chains associate in rigid and regular helical conformations providing specific sites for the chelation of calcium ions via contact (tight binding) interaction with two carboxylate groups and a number of sugar hydroxyl groups from the two chains (without the involvement of water molecules). If the main features of these models are correct, the observations made in the present study suggest that chain-chain association within junction zones must be accompanied by a substantial stiffening of the chains and a dramatic change in the ion binding mode.

A possible interpretation of this process is the following. The tight binding of a calcium ion (contact ion pair, *i.e.* including partial desolvation) is an energetically unfavorable process when a single carboxylate group is involved. It requires an interaction with two carboxylate groups that are optimally positioned and oriented (and probably also the assistance of sugar hydroxyl groups). Single-chain polyuronates are unable to provide such a geometry (or only at the expense of a too demanding conformational entropy decrease), and the counter-ions remain bound in a non-specific way within a dense counter-ion atmosphere. However, the counter-ion charges are not locally compensated and the ionic atmosphere is associated with an important (positive) self energy due to the Coulombic repulsion between the counter-ions. In other words, the calcium ions around the single chain can be viewed as being placed in a "reactive" state. The possibility of forming an appropriate chelating environment (involving a local compensation of the counter-ion charges and thus, the effective removal of inter-ionic repulsions) by association of a second polyuronate chain is likely to represent the leading driving force (in addition to presumably weaker direct chain-chain interactions) permitting chain stiffening, cation immobilization and chain-chain

association.

4.6 References

- [1] Ovodov, Y.S. *Pure. Appl. Chem.* **42** 351-369 (1974)
- [2] Rees, D.A. & Welsh, E.J. *Angew. Chem. Int. Ed.* **16** 214-224 (1977)
- [3] Morris, E.R. *Brit. Polym. J.* **18** 14-21 (1986)
- [4] Rees, D.A. *Pure Appl. Chem.* **53** 1-14 (1981)
- [5] Ramesh, H.P. & Tharanathan, R.N. *Crit. Rev. Biotech.* **23** 149-173 (2003)
- [6] Smidsrød, O. & Draget, K.I. *Carbohydr. Eur.* **14** 6-13 (1996)
- [7] Thakur, B.R., Singh, R.K. & Handa, A.K. *Crit. Rev. Food Sci. Nutrition* **37** 47-73 (1997)
- [8] Ridley, B.L., O'Neill, M.A. Mohnen, *Phytochemistry* **57** 929-967 (2001)
- [9] Willats, W.G.T., McCartney, L., Mackie W., Knox, *Plant Mol. Biol.* **47** 9-27 (2001)
- [10] Pérez, S., Mazeau, K. & Herve du Penhoat, C. *Plant Physiol. Biochem.* **38** 37-55 (2000)
- [11] Walkinshaw, M.D. & Arnott, S. *J. Mol. Biol.* **153** 1075-1085 (1981)
- [12] Powell, D.A., Morris, E.R., Gidley, M.J. & Rees, D.A. *J. Mol. Biol.* **155** 517-531 (1982)
- [13] Rees, D.A. *Carbohydr. Polym.* **2** 254-263 (1982)
- [14] Willats, W.G.T., Orfila, C., Limberg, G., Buchholt, H.C., van Alebeek, G-J.W.M., Voragen, A.G.J., Marcus, S.E., Christensen, T.M.I.E., Mikkelsen, J.D., Murray, B.S. & Knox, J.P. *J. Biol. Chem.* **276** 19404-19413 (2001)
- [15] Onsoyen, E. *Carbohydr. Eur.* **14** 26-31 (1996)
- [16] Rees, D.A. & Samuel, J.W.B. *J. Chem. Soc. C* **22** 2295-2298 (1967)
- [17] Siew, C.K. & Williams, P.A. *Biomacromolecules* **6** 963-969 (2005)
- [18] Heyraud, A., Dantas, L., Courtois, J., Courtois, B., Helbert, W. & Chanzy, H. *Carbohydr. Res.* **258** 275-279 (1994)
- [19] Braccini, I., Heyraud, A. & Pérez, S. *Biopolymers* **45** 165-175 (1998)
- [20] Dantas, L., Heyraud, A., Courtois, B., Courtois, J. & Milas, M. *Carbohydr. Polym.* **24** 185-191 (1994)
- [21] De Ruyter, G.A., Josso, S.L., Colquhoun, I.J., Voragen, A.G.J. & Rombouts, F.M. *Carbohydr. Polymers* **18** 1-7 (1992)

- [22] Heyraud, A., Courtois, J., Dantas, L., Colin-Morel, P. & Courtois, B. *Carbohydr. Res.* **240** 71-78 (1993)
- [23] Walter, R.H. Polysaccharide association structures in food. Marcel Dekker Inc., New York, U.S.A. (1998)
- [24] Shilpa, A., Agrawal, S.S. & Ray, A.R. *J. Macromol. Sci. C.* **43** 187-221 (2003)
- [25] Skjåk-Braek, G. & Espevik, T. *Carbohydr. Eur.* **14** 19-25 (1996)
- [26] Draget, K.I., Skjåk-Braek, G. & Smidsrød, O. *Int. J. Biol. Macromol.* **21** 47-55 (1997)
- [27] Anal, A.K., *Trends Food Sci. Tech.* **18** 240-251 (2007)
- [28] Davis, T.A., Volesky, B. & Mucci, A. *Water Res.* **37** 4311-4330 (2003)
- [29] Morris, E.R., Rees, D.A., Thom, D. & Boyd, J. *Carbohydr. Res.* **66** 145-154 (1978)
- [30] Thibault, J.-F. & Rinaudo, M. *Biopolymers* **25** 455-468 (1986)
- [31] Tibbits, C.W., MacDougall, A.J. & Ring, S.G. *Carbohydr. Res.* **310** 101-107 (1998)
- [32] Gardner, K.H. & Blackwell, J. *Biopolymers* **13** 1975-2001 (1974)
- [33] Sugiyama, J. Vuong & R., Chanzy, H. *Macromolecules* **24** 4168-4175 (1991)
- [34] Imberty, A. & Pérez, S. *Biopolymers* **25** 1205-1221 (1988)
- [35] Imberty, A., Chanzy, H., Pérez, S., Buléon, A. & Tran, V. *J. Mol. Biol.* **201** 365-378 (1988)
- [36] Grant, G.T., Morris, E.R., Rees, D.A., Smith, P.J.C. & Thom, D. *FEBS Lett.* **32** 195-198 (1973)
- [37] Jarvis, M.C. & Apperley, D.C. *Carbohydr. Res.* **275** 131-145 (1995)
- [38] Stokke, B.T., Draget, K.I., Smisrod, O., Yuguchi, Y., Urakawa, H. & Kajiwara, K. *Macromolecules* **33** 1853-1863 (2000)
- [39] Braccini, I. & Pérez, S. *Biomacromolecules* **2** 1089-1096 (2001)
- [40] Morris, D.F.C. *Electrochim. Acta* **27** 1481-1486 (1982)
- [41] Fang, Y., Al-Assaf, S., Philips, G.O., Nishinari, K., Funami, T., Williams, P.A. & Li, L. *J. Phys. Chem. B* **111** 2456-2462 (2007)
- [42] Atkins, E.D.T., Mackie, W. & Smolko, E.E. *Nature* **225** 626-628 (1970)
- [43] Atkins, E.D.T., Mackie, W., Parker, K.D. & Smolko, E.E. *Polymer letters* **9** 311-316 (1971)
- [44] Atkins, E.D.T., Nieduszynski, H.H., Mackie, W., Parker, K.D. & Smolko, E.E. *Biopolymers* **12** 1879-1887 (1973)

- [45] Sikorski, P., Mo, F., Skjåk-Braek, G. & Stokke, B.T. *Biomacromolecules* **8** 2098-2103 (2007)
- [46] Arnott, S., Bian, W., Chandrasekaran, R. & Mains, B.R. *Fibre Diffr. Rev.* **9** 44-51 (2000)
- [47] Mackie, W. *Biochem. J.* **125** 89P-89P (1971)
- [48] Morris, E.R., Powell, D.A., Gidley, M.J. & Rees, D.A. *J. Mol. Biol.* **155** 507-516 (1982)
- [49] Dheu-Andries, M.L & Pérez, S. *Carbohydr. Res.* **124** 324-332 (1983)
- [50] Donati, I., Holtan, S., Morch, Y.A., Borgogna, M., Dentini, M., *Biomacromolecules* **6** 1031-1040 (2005)
- [51] Yuguchi, Y., Urakawa, H., Kajiwara, K., Draget, K.I. & Stokke, B.T. *J. Mol. Struct.* **554** 21-34 (2000)
- [52] Li, L., Fang, Y., Vreeker, R. & Appelquist, I. *Biomacromolecules* **8** 464-468 (2007)
- [53] Walkinshaw, M.D. & Arnott, S. *J. Mol. Biol.* **153** 1055-1073 (1981)
- [54] Braccini, I., Rodriguez-Carvajal, M.A. & Pérez, S. *Biomacromolecules* **6** 1322-1328 (2005)
- [55] Braccini, I., Grasso, R.P. & Pérez, S. *Carbohydr. Res.* **317** 119-130 (1999)
- [56] Frei, E. & Preston, R.D. *Nature* **192** 939-934 (1961)
- [57] Boutherein, B., Mazeau, K. & Tvaroska, I. *Carbohydr. Polym.* **31** 1-12 (1997)
- [58] Larwood, V.L., Howlin, B.J. & Webb, G.A. *J. Mol. Model.* **2** 175-182 (1996)
- [59] Pérez, S., Kouwijzer, M., Mazeau, K. & Engelsens, S.B. *J. Mol. Graphics* **14** 307-321 (1996)
- [60] Manunza, B., Deiana, S. & Gessa, C. *J. Mol. Struct.* **368** 27-29 (1996)
- [61] Manunza, B., Delana, S., Pintore, B. & Gessa, C. *J. Mol. Struct.* **419** 169-172 (1997)
- [62] Manunza, B., Deiana, S., Pintore, M. & Gessa, C. *Carbohydr. Res.* **300** 85-88 (1997)
- [63] Marrone, T.J., Resat, H., Hodge, N., Chang, C.-H. & McCammon, J.A. *Protein Science* **7** 573-579 (1998)
- [64] Perry, T.D., Cygan, R.T. & Mitchell, R. *Geochim. Cosmochim. Acta* **70** 3508-3532 (2006)
- [65] Allinger, N.L. *J. Am. Chem. Soc.* **111** 8551-8566 (1989)
- [66] Allinger, N.L. *J. Am. Chem. Soc.* **111** 8566-8575 (1989)
- [67] Allinger, N.L. *J. Am. Chem. Soc.* **111** 8576-8582 (1989)

- [68] Goodford, P.J. *J. Med. Chem.* **28** 849-857 (1985)
- [69] Donati, I., Cesaro, A. & Paoletti, S. *Biomacromolecules* **7** 281-287 (2006)
- [70] Donati, I., Benegas, J.C., Cesaro, A. & Paoletti, S. *Biomacromolecules* **7** 1587-1596 (2006)
- [71] Donati, I., Benegas, J.C. & Paoletti, S. *Biomacromolecules* **7** 3439-3447 (2006)
- [72] Manunza, B. Deiana, S., Pintore, M., Gessa, *Glycoconjugate J.* **15** 297-300 (1998)
- [73] Noto, R., Martorana, V., Bulone, D., *Biomacromolecules* **6** 2555-2562 (2005)
- [74] van Gunsteren, W.F., Billeter, S.R., Eising, A.A., Hünenberger, P.H., Krüger, P., Mark, A.E., Scott, W.R.P. & Tironi, I.G. Biomolecular simulation: The GROMOS96 manual and user guide. Verlag der Fachvereine, Zürich, Switzerland. (1996)
- [75] Scott, W.R.P., Hünenberger, P.H., Tironi, I.G., Mark, A.E., Billeter, S.R., Fennen, J., Torda, A.E., Huber, T., Krüger, P. & van Gunsteren, W.F. *J. Phys. Chem. A* **103** 3596-3607 (1999)
- [76] Schuler, L.D. & van Gunsteren, W.F. *Mol. Simul.* **25** 301-319 (2000)
- [77] Schuler, L.D., Daura, X. & van Gunsteren, W.F. *J. Comput. Chem.* **22** 1205-1218 (2001)
- [78] Chandrasekhar, I., Kastenholz, M.A., Lins, R.D., Oostenbrink, C., Schüler, L.D., Tieleman, D.P. & van Gunsteren, W.F. *Eur. Biophys. J.* **32** 67-77 (2003)
- [79] Soares, T.A., Hünenberger, P.H., Kastenholz, M.A., Kräutler, V., Lenz, T., Lins, R.D., Oostenbrink, C. & van Gunsteren, W.F. *J. Comput. Chem.* **26** 725-737 (2005)
- [80] Börjesson, U. & Hünenberger, P.H. *J. Phys. Chem. B* **108** 13551-13559 (2004)
- [81] Lins, R.D. & Hünenberger, P.H. *J. Comput. Chem.* **26** 1400-1412 (2005)
- [82] Pereira, C.S., Kony, D., Baron, R., Müller, M., van Gunsteren, W.F. & Hünenberger, P.H. *Biophys. J.* **90** 4337-4344 (2006)
- [83] Kräutler, V., Müller, M. & Hünenberger, P.H. *Carbohydr. Res.* **342** 2097-2124 (2007)
- [84] Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F. & Hermans, J. Interaction models for water in relation to protein hydration. In: *Intermolecular Forces*. Pullman, B., Ed. Reidel, Dordrecht, The Netherlands. pp 331-342 (1981)
- [85] Hockney, R.W. *Methods Comput. Phys.* **9** 136-211 (1970)
- [86] Ryckaert, J.-P., Ciccotti, G. & Berendsen, H.J.C. *J. Comput. Phys.* **23** 327-341 (1977)
- [87] Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., Di Nola, A. & Haak, J.R. *J. Chem. Phys.* **81** 3684-3690 (1984)
- [88] van Gunsteren, W.F. & Berendsen, H.J.C. *Angew. Chem. Int. Ed.* **29** 992-1023 (1990)

- [89] Heinz, T.N., van Gunsteren, W.F. & Hünenberger, P.H. *J. Chem. Phys.* **115** 1125-1136 (2001)
- [90] Kony, D., Damm, W., Stoll, S., van Gunsteren, W.F. & Hünenberger, P.H. *Biophys. J.* **93** 442-455 (2007)
- [91] Stokke, B.T., Elgsaeter, A., Skjåk-Braek, G. & Smidsrød, O. *Carbohydr. Res.* **160** 13-28 (1987)
- [92] Stokke, B.T. & Brant, D.A. *Biopolymers* **30** 1161-1181 (1990)
- [93] Dentini, M., Rinaldi, G., Risica, D., Barbetta, A. & Skjåk-Braek, G. *Carbohydr. Polym.* **59** 489-499 (2005)
- [94] Cros, S., Garnier, C., Axelos, M.A.V., Imberty, A. & Pérez, S. *Biopolymers* **39** 339-352 (1996)
- [95] Plashchina, I.G., Semenova, M.G., E.E. Braudo, Tolstoguzov, *Carbohydr. Polym.* **5** 159-179 (1985)
- [96] Lee, K.Y., Bouhadur, K.H. & Mooney, D.J. *Biomacromolecules* **3** 1129-1134 (2002)
- [97] Maurstad, G., Danielsen, S. & Stokke, B.T. *J. Phys. Chem. B* **107** 8172-8180 (2003)
- [98] Insight II Molecular Modeling Package. Molecular Simulations Inc. (2000)
- [99] Karlson, P., Dixon, H.B. & Liebecq, C. *Eur. J. Biochem.* **131** 5-8 (1983)
- [100] Pérez, S. & Vergelati, C. *Biopolymers* **24** 1809-1822 (1985)
- [101] Lemieux, R.U. & Koto, S. *Tetrahedron* **30** 1933-1944 (1974)
- [102] Pérez, S. & Marchessault, R.H. *Carbohydr. Res.* **65** 114-120 (1978)
- [103] Graczyk, P.P. & Mikolajczyk, M. *Topics Stereochem.* **21** 159-349 (1994)
- [104] Thogersen, H., Lemieux, R.U., Bock, K. & Meyer, B. *Can. J. Chem.* **60** 44-57 (1982)
- [105] Tvaroška, I. & Carver, J.P. *Carbohydr. Res.* **309** 1-9 (1998)
- [106] Kastenholz, M.A. & Hünenberger, P.H. *J. Chem. Phys.* **124** 224501/1-224501/20 (2006)
- [107] Weast, R.C. Handbook of chemistry and physics. Volume . The Chemical Rubber Co, U.S.A. (1972)
- [108] de Bakker, P.I.W., Hünenberger, P.H. & McCammon, J.A. *J. Mol. Biol.* **285** 1811-1830 (1999)

4.7 Appendix

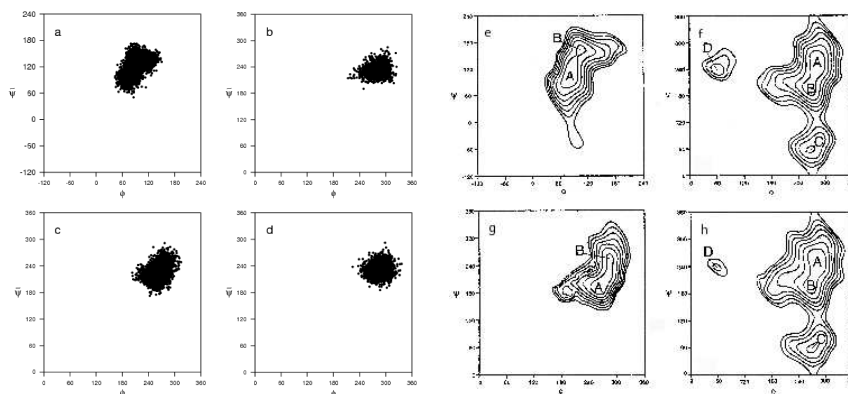


Figure 4.13: Ramachandran maps (ϕ , $\tilde{\psi}$) characterizing the conformation of the glycosidic linkage observed in explicit-solvent simulations of diuronates with the GROMOS 45A4 force field⁸¹ (a-d) or in implicit-solvent (permittivity 80) energy minimizations of the corresponding diuronic acids with the MM3 force field⁶⁵⁻⁶⁷ (e-f; reproduced from Braccini et al.;⁵⁵ isocontours drawn at 1 kcal·mol⁻¹ increments up to 8 kcal·mol⁻¹). The simulations were performed at constant temperature (298 K) and pressure (1 atm) within cubic periodic boxes of initial edge lengths 3 nm containing about 1000-1100 water molecules (no counterions) for a duration of 5 ns. Individual points represent conformations sampled at 1 ps intervals. Simulations were also performed with counter-ions (two Na⁺ or one Ca²⁺), leading to very similar results (data not shown). The graphs correspond to α -[D-GalU- α (1 \rightarrow 4)-D-GalU] (a,e), β -[D-GlcU- β (1 \rightarrow 4)-D-GlcU] (b,f), α -[L-GulU- α (1 \rightarrow 4)-L-GulU] (c,g) and β -[D-ManU- β (1 \rightarrow 4)-D-ManU] (d,h). See Section 4.3.2 for the definitions of ϕ and $\tilde{\psi}$, and the relationship between $\tilde{\psi}$ and ψ .

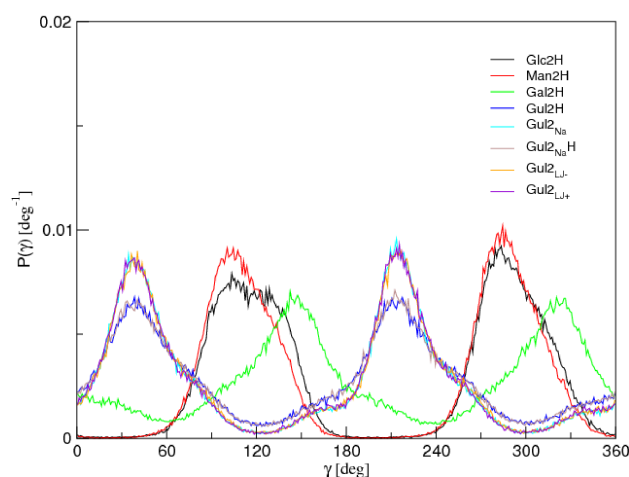


Figure 4.14: Normalized probability distributions of the γ dihedral angle characterizing the orientation of the carboxylate group, averaged over the N_{res} residues, for the eight simulation variants (500 K or different counter-ion environment; Table 4.2). The γ distribution accounts for the combined distributions of the two symmetry-related dihedral angles γ_1 and γ_2 . See Section 4.3.2 for the definitions of γ_1 , γ_2 and γ . Averaging was performed over the final 9 ns of the simulations.

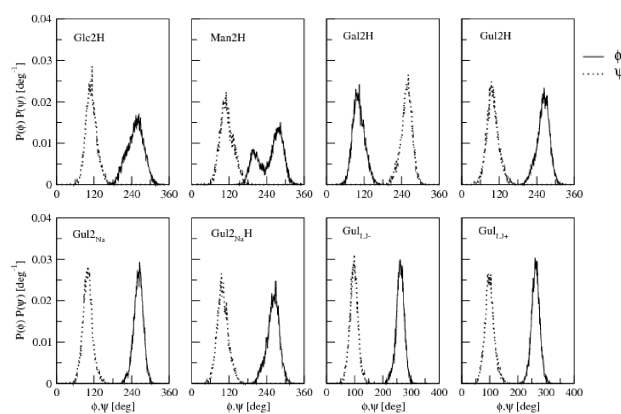


Figure 4.15: Normalized probability distributions of the glycosidic dihedral angles ϕ and ψ , averaged over the N_{res} linkages, for the eight simulation variants (500 K or different counter-ion environment; Table 4.2). See Section 4.3.2 for the definitions of ϕ and ψ . Averaging was performed over the final 9 ns of the simulations.

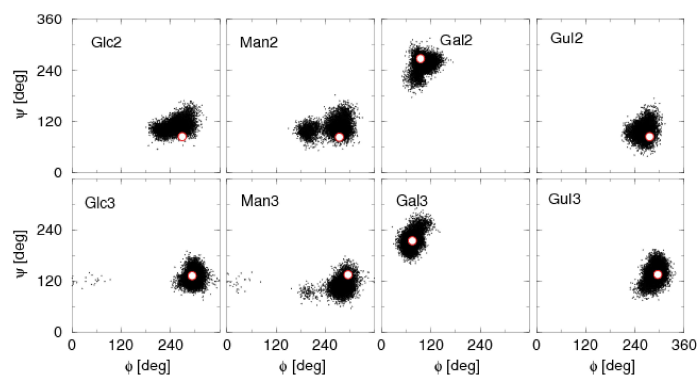


Figure 4.16: Ramachandran maps (ϕ , ψ) characterizing the conformation of the glycosidic linkages for the N_{res} linkages during the eight main simulations (300 K, Ca^{2+} counter-ions; Table 4.2). Individual points represent conformations sampled at 10 ps interval (over the entire 10 ns simulations). The reference values (ϕ_{ref} , ψ_{ref} ; Table 4.3) suggested for the different helix types by Braccini *et al.*,⁵⁵ based on MM3 relaxed-residue energy maps for the corresponding diuronic acids, are also indicated (ψ_{ref} was converted from the $\tilde{\psi}$ to the ψ definition). See Section 4.3.2 for the definitions of ϕ , ψ and $\tilde{\psi}$.

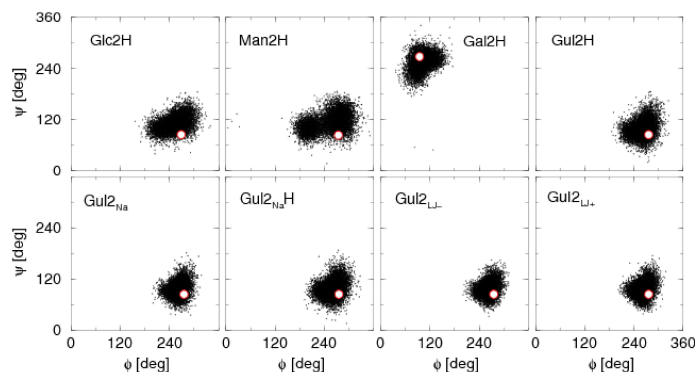


Figure 4.17: Ramachandran maps (ϕ , ψ) characterizing the conformation of the glycosidic linkages for the N_{res} linkages during the eight simulation variants (500 K or different counter-ion environment; Table 4.2). Individual points represent conformations sampled at 10 ps interval (over the entire 10 ns simulations). The reference values (ϕ_{ref} , ψ_{ref} ; Table 4.3) suggested for the different helix types by Braccini *et al.*,⁵⁵ based on MM3 relaxed-residue energy maps for the corresponding diuronic acids, are also indicated (ψ_{ref} was converted from the $\tilde{\psi}$ to the ψ definition). See Section 4.3.2 for the definitions of ϕ , ψ and $\tilde{\psi}$.

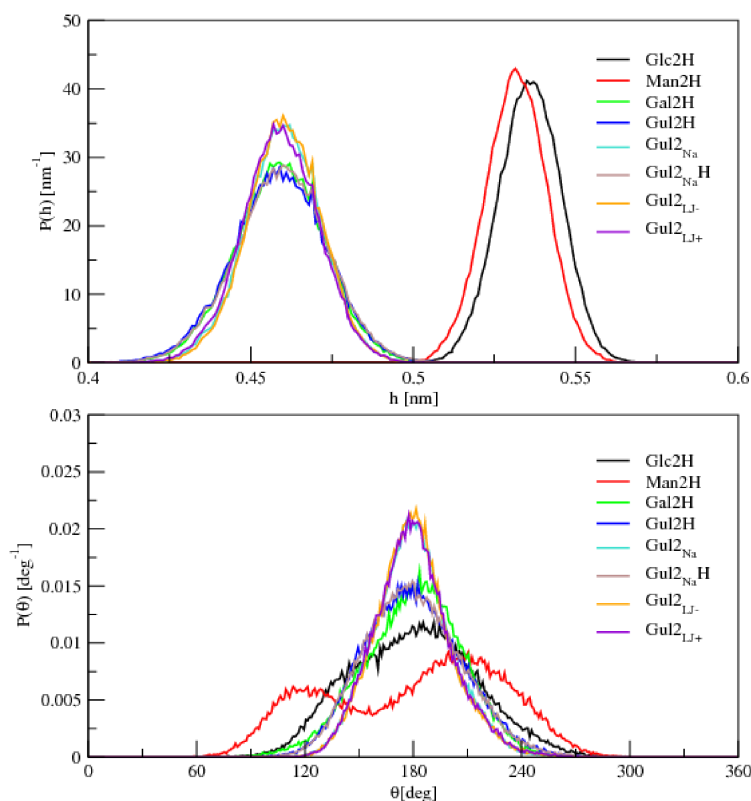


Figure 4.18: Normalized probability distributions of the helix raise h and turn angle θ , averaged over the N_{res} linkages, for the eight simulation variants (500 K or different counter-ion environment; Table 4.2). See Section 4.3.2 for the definition of h and θ . Averaging was performed over the final 9 ns of the simulations.

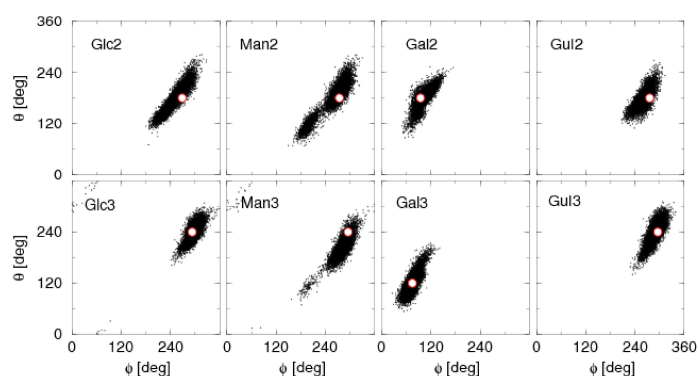


Figure 4.19: Maps characterizing the local conformation of the glycosidic linkages in the (ϕ, θ) space (glycosidic dihedral angle ϕ and turn angle θ) for the N_{res} linkages during the eight main simulations (300 K, Ca^{2+} counter-ions; Table 4.2). Individual points represent conformations sampled at 10 ps intervals (over the entire 10 ns simulations). The reference values (ϕ_{ref} ; Table 4.3; $\theta_{ref}=120, 180$ or 240° for simulations initiated from a 3_1 -, 2_1 - or 3_2 -helix, respectively) suggested for the different helix types by Braccini et al.,⁵⁵ based on MM3 relaxed-residue energy maps for the corresponding diuronic acids, are also indicated. See Section 4.3.2 for the definitions of ϕ and θ .

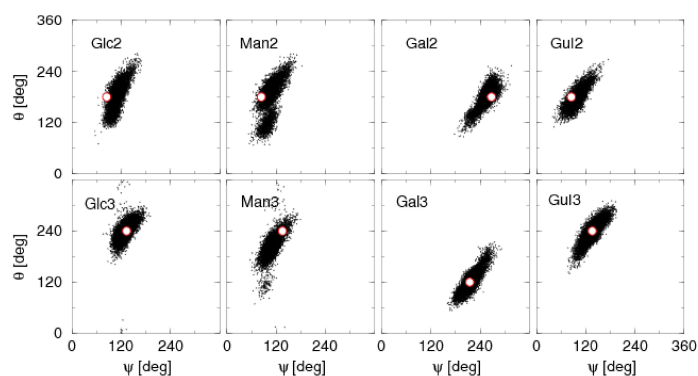


Figure 4.20: Maps characterizing the local conformation of the glycosidic linkages in the (ψ, θ) space (glycosidic dihedral angle ψ and turn angle θ) for the N_{res} linkages during the eight main simulations (300 K, Ca^{2+} counter-ions; Table 4.3). Individual points represent conformations sampled at 10 ps intervals (over the entire 10 ns simulations). The reference values (ψ_{ref} ; Table 4.2; $\theta_{ref}=120, 180$ or 240° for regular 3_1 -, 2_1 - or 3_2 -helices, respectively) suggested for the different helix types by Braccini et al.,⁵⁵ based on MM3 relaxed-residue energy maps for the corresponding diuronic acids, are also indicated (ψ_{ref} was converted from the $\tilde{\psi}$ to the ψ definition). See Section 4.3.2 for the definitions of ψ , $\tilde{\psi}$ and θ .

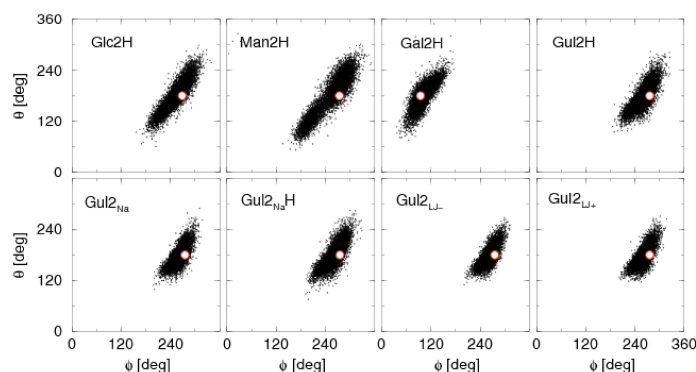


Figure 4.21: Maps characterizing the local conformation of the glycosidic linkages in the (ϕ, θ) space (glycosidic dihedral angle ϕ and turn angle θ) for the N_{res} linkages during the eight simulation variants (500 K or different counter-ion environment; Table 4.2). Individual points represent conformations sampled at 10 ps intervals (over the entire 10 ns simulations). The reference values (ϕ_{ref} ; Table 4.3; $\theta_{ref}=120, 180$ or 240° for regular 3_1 -, 2_1 - or 3_2 -helices, respectively) suggested for the different helix types by Braccini et al.,⁵⁵ based on MM3 relaxed-residue energy maps for the corresponding diuronic acids, are also indicated. See Section 4.3.2 for the definitions of ϕ and θ .

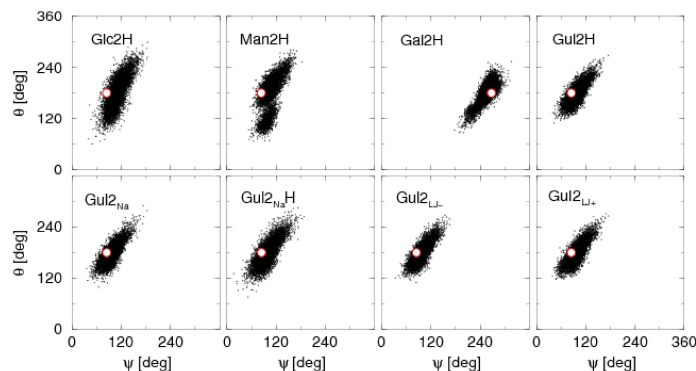


Figure 4.22: Maps characterizing the local conformation of the glycosidic linkages in the (ψ, θ) space (glycosidic dihedral angle ψ and turn angle θ) for the N_{res} linkages during the eight simulation variants (500 K or different counter-ion environment; Table 4.2). Individual points represent conformations sampled at 10 ps intervals (over the entire 10 ns simulations). The reference values (ψ_{ref} ; Table 4.3; $\theta_{ref}=120, 180$ or 240° for regular 3_1 -, 2_1 - or 3_2 - helices, respectively) suggested for the different helix types by Braccini et al.,⁵⁵ based on MM3 relaxed-residue energy maps for the corresponding diuronic acids, are also indicated (ψ_{ref} was converted from the $\tilde{\psi}$ to the ψ definition). See Section 4.3.2 for the definitions of ψ , $\tilde{\psi}$ and θ .

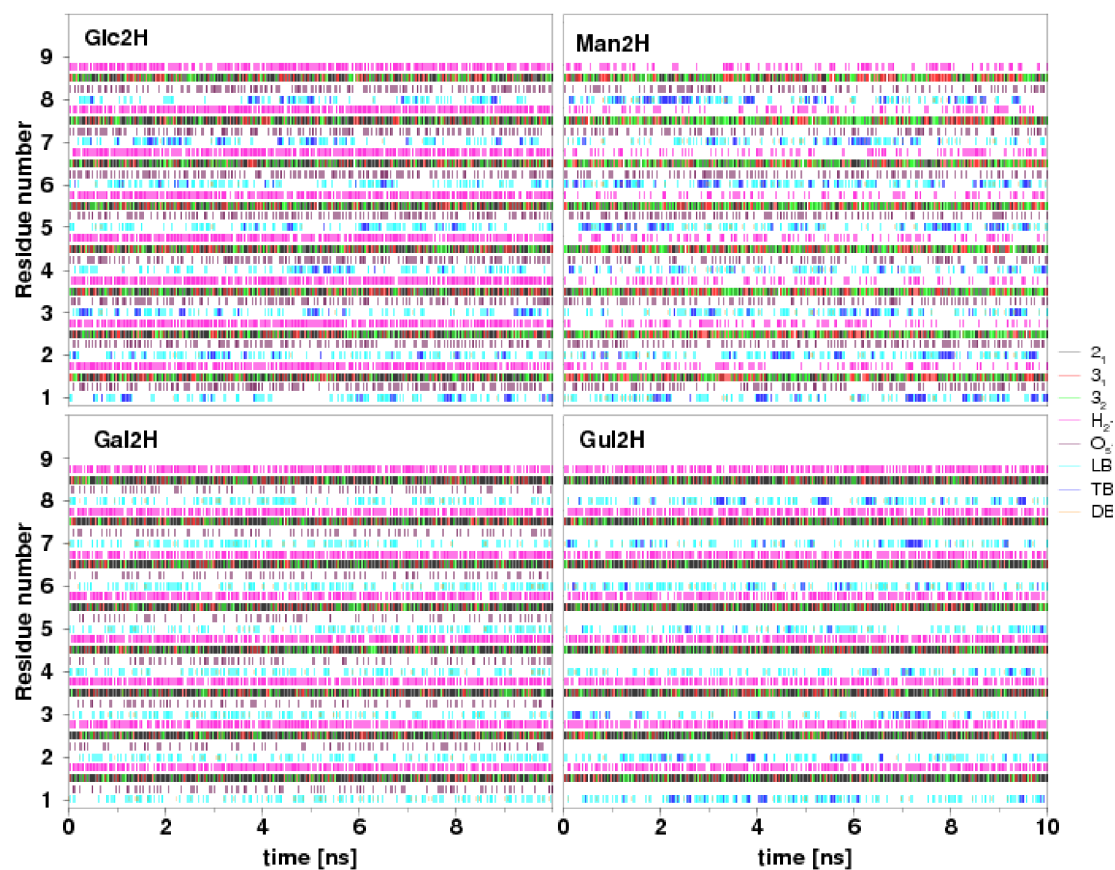


Figure 4.23: Time evolution of the local helical conformations and of the inter-residue hydrogen bonds for the N_{res} glycosidic linkages, as well as of ion binding events for the N_{res} residues, displayed for simulations Glc2H, Man2H, Gal2H and Gul2H (Table 4.2). The indicated events are the formation of a local 2_1 -, 3_1 - or 3_2 -helical conformation, of ($H_2 \rightarrow O'_6$ or $O_5 \leftarrow H'_3$) hydrogen bonds, of a solvent-separated (loose binding; LB) ion pair, of a contact (tight binding; TB) ion pair or of an ion pair (LB or TB) involving an ion also bound (LB or TB) to the carboxylate group of another residue (double binding; DB; smaller bars for the two involved carboxylate groups). See Section 4.3.2 for definitions of local helical conformations, hydrogen bonding and ion binding. The sampling period (individual bars) is 20 ps. Color code 2_1 (black), 3_1 (red), 3_2 (green), $H_2 \rightarrow O'_6$ (pink), $O_5 \leftarrow H'_3$ (brown), LB (cyan), TB (blue), DB (orange).

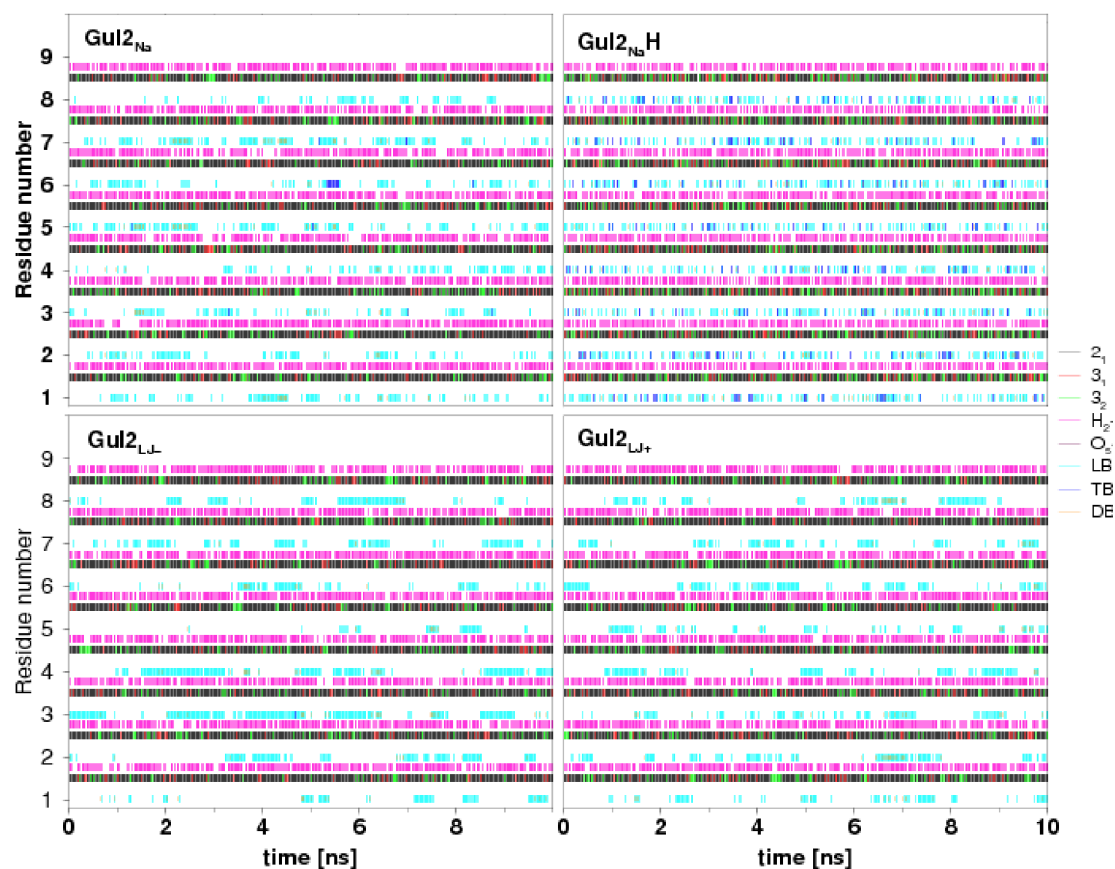


Figure 4.24: Time evolution of the local helical conformations and of the inter-residue hydrogen bonds for the N_{res} glycosidic linkages, as well as of ion binding events, for the N_{res} residues, displayed for simulations $Gul2_{Na}$, $Gul2_{NaH}$, $Gul2_{LJ-}$ and $Gul2_{LJ+}$ (Table 4.2). The indicated events are the formation of a local 2_1 -, 3_1 - or 3_2 -helical conformation, of ($H_2 \rightarrow O'_6$ or $O_5 \leftarrow H'_3$) hydrogen bonds, of a solvent-separated (loose binding; LB) ion pair, of a contact (tight binding; TB) ion pair or of an ion pair (LB or TB) involving an ion also bound (LB or TB) to the carboxylate group of another residue (double binding; DB; smaller bars for the two involved carboxylate groups). See Section 4.3.2 for definitions of local helical conformations, hydrogen bonding and ion binding. The sampling period (individual bars) is 20 ps. Color code 2_1 (black), 3_1 (red), 3_2 (green), $H_2 \rightarrow O'_6$ (pink), $O_5 \leftarrow H'_3$ (brown), LB (cyan), TB (blue), DB (orange).

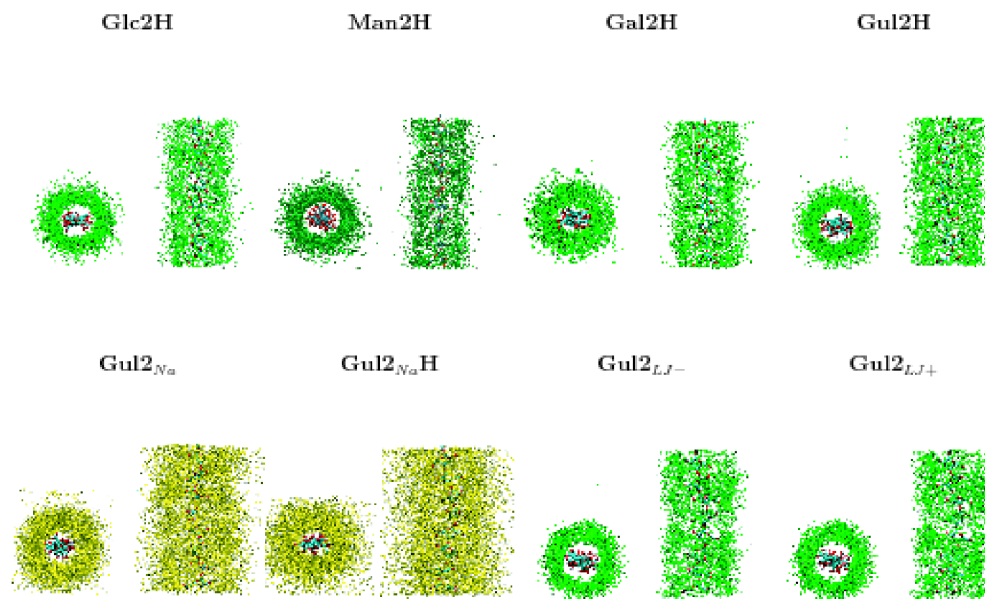


Figure 4.25: Positions of the counter-ions along the chain (top and side views), as sampled along of the eight simulation variants (500 K or different counter-ions environment; Table 4.2). The ion coordinates correspond to successive configurations superimposed onto the initial (after 0.2 ns equilibration) one (rotational and translational fit based on the C₂, C₃, C₅ and O₅ atoms). The latter reference configurations is also displayed. The sampling period (between displayed ion coordinates) is 10 ps. The figures correspond to the final 9 ns of the simulations.

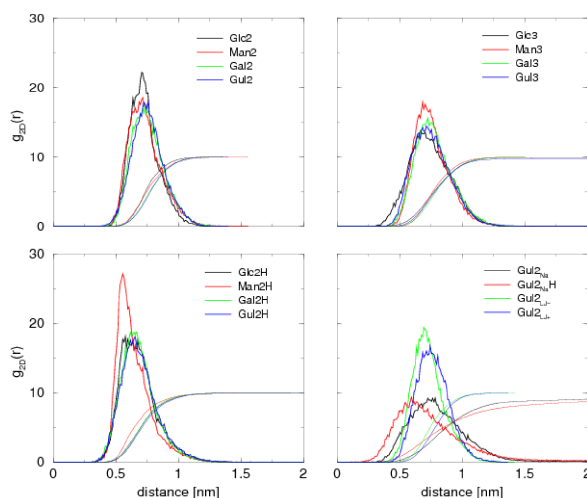


Figure 4.26: Two-dimensional radial distribution function $g_{2D}(r)$ for the coordinates of the counter-ions in the xy -plane relative to the helix axis, averaged over the N_{ions} (positive) counter-ions for the different simulations (Table 4.2). The cumulative integral of $(L_x \cdot L_y)^{-1} 2\pi r g_{2D}(r)$ (amplified by 10 for readability) is also displayed. See Section 4.3.2 for the definitions of $g_{2D}(r)$, L_x and L_y . Averaging was performed over the final 9 ns of the simulations.

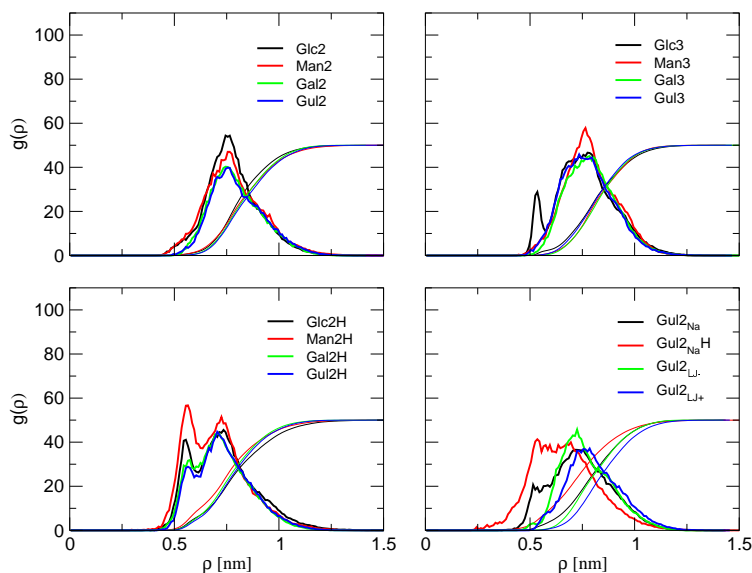


Figure 4.27: Radial distribution function $g(\rho)$ for the (minimum-image) distance between the sugar oxygen atom O_1 and the closest counter-ion, averaged over the N_{res} residues, for the different simulations (Table 4.2). The cumulative integral of $V^{-1} 4\pi \rho^2 g(\rho)$ (amplified by 50 for readability) is also displayed. See Section 4.3.2 for the definitions of $g(r)$ and V . Averaging was performed over the final 9 ns of the simulations.

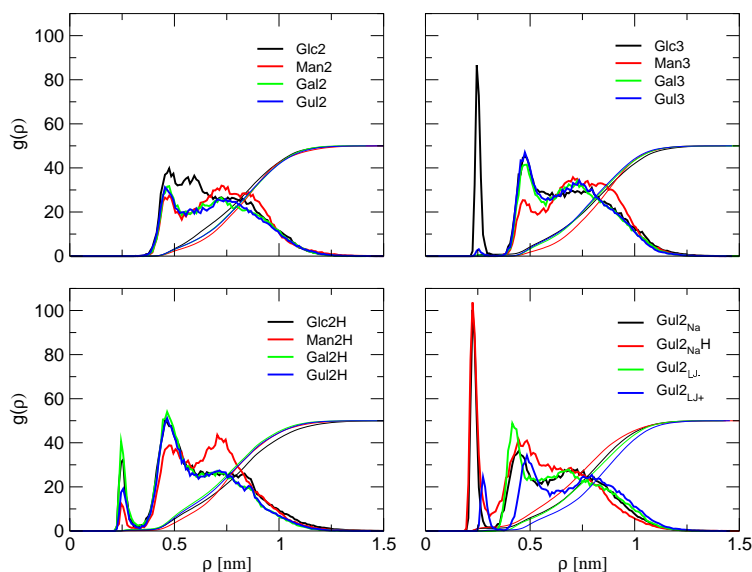


Figure 4.28: Radial distribution function $g(\rho)$ for the (minimum-image) distance between the sugar oxygen atom O_2 and the closest counter-ion, averaged over the N_{res} residues, for the different simulations (Table 4.2). The cumulative integral of $V^{-1}4\pi\rho^2g(\rho)$ (amplified by 50 for readability) is also displayed. See Section 4.3.2 for the definitions of $g(\rho)$ and V . Averaging was performed over the final 9 ns of the simulations.

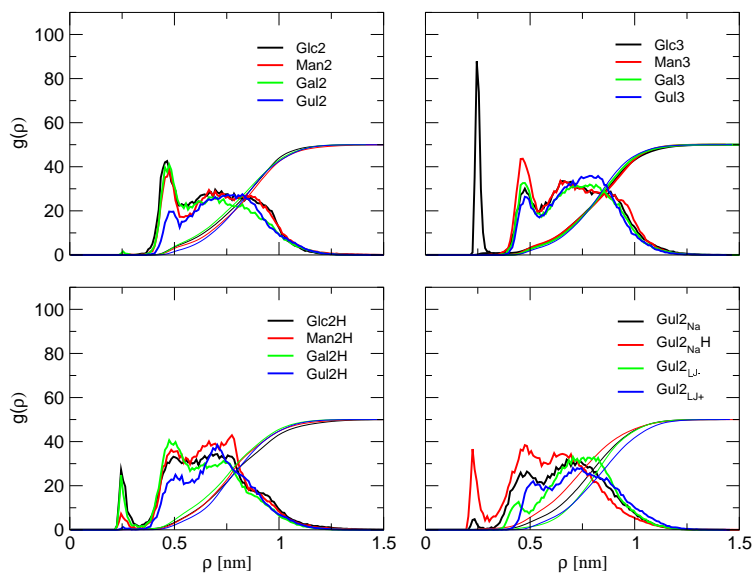


Figure 4.29: Radial distribution function $g(\rho)$ for the (minimum-image) distance between the sugar oxygen atom O_3 and the closest counter-ion, averaged over the N_{res} residues, for the different simulations (Table 4.2). The cumulative integral of $V^{-1}4\pi\rho^2g(\rho)$ (amplified by 50 for readability) is also displayed. See Section 4.3.2 for the definitions of $g(\rho)$ and V . Averaging was performed over the final 9 ns of the simulations.

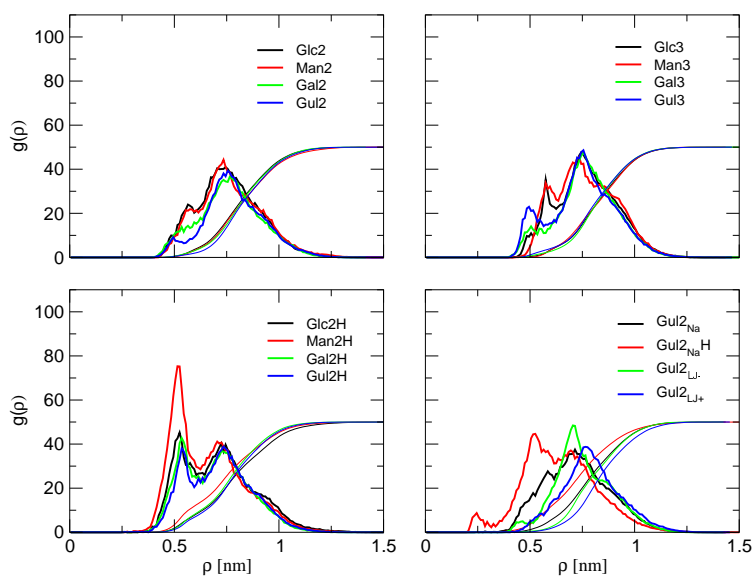


Figure 4.30: Radial distribution function $g(\rho)$ for the (minimum-image) distance between the sugar oxygen atom O_5 and the closest counter-ion, averaged over the N_{res} residues, for the different simulations (Table 4.2). The cumulative integral of $V^{-1}4\pi\rho^2g(\rho)$ (amplified by 50 for readability) is also displayed. See Section 4.3.2 for the definitions of $g(\rho)$ and V . Averaging was performed over the final 9 ns of the simulations.

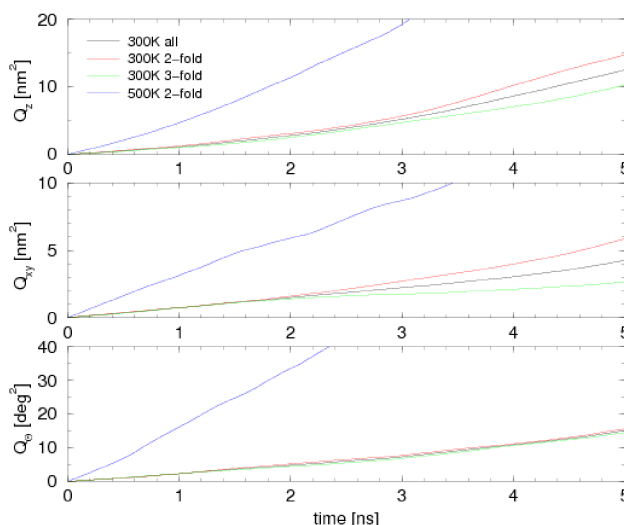


Figure 4.31: Mean-square displacements associated with the chain rigid-body motion, averaged over different sets of N_{sim} simulations (Table 4.2). The coordinates considered are the z -coordinate of the chain center, the coordinate vector of the chain center in the xy -plane and the chain orientational angle θ (mean-square displacement Q_z , Q_{xy} and Q_θ , respectively). The simulation sets correspond to all simulations at 300 K with Ca^{2+} counter-ions ($N_{sim}=8$), and those started from a two-fold ($N_{sim}=4$) or three-fold ($N_{sim}=4$) helical structure, or all simulations of pGulU at 500 K with Ca^{2+} counter-ions ($N_{sim}=4$) started from a two-fold helical structure. See Section 4.3.2 for the definitions of $Q_{\mathbf{q}}(t)$ and $D_{\mathbf{q}}$. Averaging was performed over the final 9 ns of the simulations considering all possible time origins.

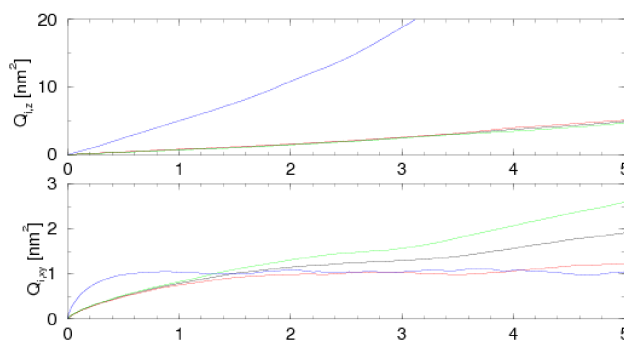


Figure 4.32: Mean-square displacements associated with the ion coordinates, averaged over the N_{ion} positive counter-ions and over different sets of N_{sim} simulations (Table 4.2). The ion coordinates are separated into the z -coordinate and the coordinate vector in the xy -plane (mean-square displacement $Q_{i,z}$, $Q_{i,xy}$, respectively). The simulation sets correspond to all simulations at 300 K with Ca^{2+} counter-ions ($N_{sim}=8$), and those started from a two-fold ($N_{sim}=4$) or three-fold ($N_{sim}=4$) helical structure, or all simulations of pGulU at 500 K with Ca^{2+} counter-ions ($N_{sim}=4$) started from a two-fold helical structure. See Section 4.3.2 for the definitions of $Q_{\mathbf{q}}(t)$ and $D_{\mathbf{q}}$. Averaging was performed over the final 9 ns of the simulations considering all possible time origins.

| Code | α_1 [deg] | α_2 [deg] | α_3 [deg] | α_4 [deg] | α_5 [deg] | α_6 [deg] | Chair [%] |
|----------------------|------------------|------------------|------------------|------------------|------------------|------------------|-----------|
| Glc2 | 306.5 (6.6) | 51.8 (6.0) | 306.1 (5.8) | 60.7 (7.0) | 301.8 (6.3) | 53.3 (5.7) | 99.89 |
| Man2 | 299.3 (5.4) | 56.3 (5.5) | 307.5 (5.9) | 56.3 (7.0) | 301.9 (6.1) | 58.4 (4.9) | 99.94 |
| Gal2 | 299.5 (5.1) | 62.4 (4.7) | 299.1 (4.9) | 59.2 (5.9) | 305.6 (5.8) | 54.1 (5.3) | 99.99 |
| Gul2 | 56.4 (4.9) | 302.8 (4.8) | 58.6 (5.0) | 300.7 (5.7) | 55.4 (5.7) | 306.2 (5.1) | 99.99 |
| Glc3 | 306.9 (6.9) | 50.8 (6.7) | 306.9 (6.5) | 61.4 (6.9) | 297.7 (6.2) | 56.3 (5.9) | 99.68 |
| Man3 | 299.4 (5.4) | 56.7 (5.6) | 307.5 (6.0) | 55.8 (6.9) | 301.3 (6.0) | 59.2 (4.9) | 99.90 |
| Gal3 | 299.2 (5.3) | 63.4 (4.8) | 297.9 (4.9) | 59.8 (5.9) | 306.9 (5.9) | 52.5 (5.4) | 99.97 |
| Gul3 | 57.9 (5.0) | 301.3 (4.7) | 59.2 (5.1) | 301.0 (6.0) | 53.5 (5.8) | 307.3 (5.1) | 99.98 |
| Gul2 _{Na} | 56.4 (4.9) | 302.7 (4.8) | 58.6 (5.0) | 300.7 (5.7) | 55.4 (5.7) | 306.2 (5.1) | 99.99 |
| Gul2 _{LJ-} | 56.4 (4.9) | 302.7 (4.8) | 58.7 (5.0) | 300.7 (5.7) | 55.4 (5.7) | 306.2 (5.1) | 100.00 |
| Gul2 _{LJ+} | 56.4 (4.9) | 302.7 (4.8) | 58.7 (5.0) | 300.7 (5.7) | 55.4 (5.7) | 306.1 (5.1) | 100.00 |
| Glc2H | 306.2 (12.1) | 51.6 (11.6) | 306.7 (9.3) | 60.6 (10.5) | 301.7 (9.6) | 53.6 (10.6) | 98.11 |
| Man2H | 299.1 (7.1) | 55.9 (7.0) | 307.9 (7.7) | 65.7 (9.4) | 302.8 (7.8) | 57.3 (6.3) | 99.17 |
| Gal2H | 299.5 (6.6) | 62.5 (6.2) | 298.9 (6.3) | 59.6 (7.6) | 305.7 (7.6) | 53.7 (6.9) | 99.69 |
| Gul2H | 56.3 (6.4) | 302.9 (6.2) | 58.6 (6.5) | 300.5 (7.5) | 55.4 (7.5) | 306.4 (6.7) | 99.77 |
| Gul2 _{Na} H | 56.3 (6.4) | 302.9 (6.2) | 58.5 (6.5) | 300.5 (7.5) | 55.3 (7.5) | 306.5 (6.6) | 99.79 |

Table 4.8: Ring conformations. Mean values and standard deviations (over the time series; between parentheses) of the ring dihedral angles α_k , $k=1..6$, together with the occurrences of the leading chair conformation (4C_1 for pGlcU, pManU and pGalU; 1C_4 for pGulU; Figure 4.1), averaged over the N_{res} residues for the different simulations (Table 4.2). See Section 4.3.2 for the definitions of the α_k dihedral angles and chair conformations. Averaging was performed over the final 9 ns of the simulations.

| Code | O ₁ [%] | O ₂ [%] | O ₃ [%] | O ₅ [%] |
|----------------------|--------------------|--------------------|--------------------|--------------------|
| Glc2 | 1.5 | 13.6 | 12.3 | 4.5 |
| Man2 | 1.7 | 8.6 | 10.2 | 3.8 |
| Gal2 | 0.3 | 11.2 | 13.6 | 4.1 |
| Gul2 | 0.5 | 11.2 | 6.9 | 2.7 |
| Glc3 | 3.9 | 14.6 (1.7) | 10.4 (1.7) | 3.3 |
| Man3 | 1.0 | 7.4 | 11.0 | 3.0 |
| Gal3 | 0.4 | 12.6 | 9.4 | 4.0 |
| Gul3 | 1.0 | 13.8 (0.1) | 7.8 | 6.2 |
| Gul2 _{Na} | 5.3 | 17.4 (2.6) | 12.5 (0.2) | 7.0 |
| Gul2 _{LJ-} | 0.9 | 16.4 | 5.5 | 2.5 |
| Gul2 _{LJ+} | 0.6 | 12.2 (0.9) | 7.4 | 2.0 |
| Glc2H | 6.6 | 18.4 (1.1) | 13.3 (0.9) | 11.9 |
| Man2H | 9.6 | 14.1 (0.4) | 12.8 (0.2) | 19.8 |
| Gal2H | 5.6 | 22.3 (1.5) | 17.5 (0.8) | 12.4 |
| Gul2H | 4.7 | 20.6 (0.8) | 9.9 | 10.3 |
| Gul2 _{Na} H | 14.0 (0.2) | 23.7 (3.8) | 19.3 (1.5) | 17.5 (0.6) |

Table 4.9: Ion binding to the different sugar oxygen atoms. The occurrences of ion binding to the oxygen atoms O₁ (glycosidic oxygen), O₂ and O₃ (hydroxyl groups) or O₅ (ring oxygen), averaged over the N_{res} residues, are reported for the different simulations (Table 4.2). The values correspond to the formation of a solvent-separated (loose binding; LB) or a contact (tight binding; TB; between parentheses if non-zero occurrences) interaction. See Section 4.3.2 for the definitions of the binding events. Averaging was performed over the final 9 ns of the simulations considering all possible time origins.

| Code | $D_z[\text{nm}^2 \cdot \text{ns}^{-1}]$ | $D_{xy}[\text{nm}^2 \cdot \text{ns}^{-1}]$ | $D_\Theta[\text{deg}^2 \cdot \text{ns}^{-1}]$ | $D_{i,z}[\text{nm}^2 \cdot \text{ns}^{-1}]$ | $D_{i,xy}[\text{nm}^2 \cdot \text{ns}^{-1}]$ |
|----------------------|---|--|---|---|--|
| Glc2 | 0.51 | 0.12 | 0.80 | 0.19 (0.08) | 0.05 (0.02) |
| Man2 | 0.54 | 0.25 | 1.69 | 0.52 (0.23) | 0.08 (0.01) |
| Gal2 | 2.55 | 0.37 | 0.78 | 0.65 (0.41) | 0.06 (0.01) |
| Gul2 | 0.08 | 0.15 | 2.09 | 0.29 (0.16) | 0.09 (0.06) |
| Glc3 | 1.21 | 0.08 | 1.37 | 0.43 (0.21) | 0.25 (0.35) |
| Man3 | 0.45 | 0.14 | 2.21 | 0.14 (0.07) | 0.06 (0.03) |
| Gal3 | 1.04 | 0.19 | 0.642 | 0.83 (0.58) | 0.09 (0.04) |
| Gul3 | 0.40 | 0.22 | 0.21 | 0.19 (0.10) | 0.08 (0.05) |
| Gul2 _{Na} | 0.42 | 0.26 | 10.57 | 0.94 (0.56) | 0.52 (0.95) |
| Gul2 _{LJ-} | 0.10 | 0.22 | 1.91 | 0.36 (0.19) | 0.09 (0.04) |
| Gul2 _{LJ+} | 0.68 | 0.11 | 0.99 | 0.19 (0.16) | 0.10 (0.03) |
| Glc2H | 2.67 | 0.67 | 12.33 | 1.15 (0.33) | 0.03 (0.01) |
| Man2H | 0.47 | 1.22 | 2.37 | 1.22 (1.18) | 0.02 (0.01) |
| Gal2H | 9.07 | 0.83 | 11.04 | 8.19 (2.08) | 0.03 (0.01) |
| Gul2H | 0.840 | 0.16 | 10.50 | 1.50 (0.76) | 0.04 (0.03) |
| Gul2 _{Na} H | 2.67 | 0.41 | 308.51 | 8.70 (4.58) | 3.30 (2.33) |

Table 4.10: Diffusion constants evaluated for different sets of degrees of freedom of the chain and of the counter-ions. The results are reported for all simulations (Table 4.2). The diffusion coefficients $D_{\mathbf{q}}$ are calculated from a linear least-square-fit of the corresponding mean-square displacements over the time period 0-3 ns. The chain diffusion is described in terms of the longitudinal (D_z), transverse (D_{xy}) and orientational (D_Θ) diffusion constants. The ionic diffusion is described in terms of the ionic longitudinal ($D_{i,z}$) and transverse ($D_{i,xy}$) diffusion constants, averaged over all N_{ion} positive counter-ions in the system (the corresponding standard deviation is reported between parentheses). See Section 4.3.2 for the definitions of $Q_{\mathbf{q}}(t)$ and $D_{\mathbf{q}}$. The calculation was performed considering the final 9 ns of the simulations.

Chapter 5

Interaction of alginate single-chain polyguluronate segments with mono- and divalent metal cations: a comparative molecular dynamics study

5.1 Summary

The interaction of a set of monovalent (Na^+ , K^+) and divalent (Mg^{2+} , Ca^{2+}) metal cations with single-chain homopolyguluronate segments of alginates (periodic chain based on a dodecameric repeat, 2_1 -helical conformation) is investigated using explicit-solvent molecular dynamics (MD) simulations (at 300 K and 1 bar). A total of 14 (neutralizing) combinations of the different ions are considered (single type of cation or simultaneous presence of two types of cation, either in the presence or absence of chloride anions). The main observations are that : (i) the chain conformation and intramolecular hydrogen-bonding is insensitive to the counter-ion environment; (ii) the binding of the cations is essentially non-specific for all ions considered (counter-ion atmosphere confined within a cylinder of high ionic density, but no well-defined binding sites); (iii) the density and tightness of the distributions of the different cations within of the counter-ion atmosphere follow the approximate order $\text{Ca}^{2+} > \text{Mg}^{2+} > \text{K}^+ > \text{Na}^+$ (preferential affinity for divalent over monovalent cations, and for Ca^{2+} over Mg^{2+}); (iv) the solvent-separated binding of the cations to the carboxylate groups of the chain is frequent, and its occurrence follows the approximate order $\text{K}^+ > \text{Na}^+ > \text{Ca}^{2+} > \text{Mg}^{2+}$ (contact binding events as well as the binding of a cation to multiple carboxylate groups are very infrequent); (v) the counter-ion atmosphere typically leads to a complete screening of the chain charge within 1.0-1.2 nm of the chain axis, and charge to a reversal at about 1.5 nm (*i.e.* the effective chain charge becomes positive and as high in magnitude as one quarter of the bare chain charge at this distance, before slowly decreasing to zero). These findings agree very well (in a qualitative sense) with available experimental data and predictions from simple analytical models, and provide further insight concerning the nature of alginate-cation interactions in aqueous solution.

5.2 Introduction

Polyuronates are (predominantly) polymers of uronic acids in (1→4)-linkage¹⁻⁵. These polymers have diverse biological functions in plants, such as the preservation of the structure, texture and flexibility, as well as the prevention of desiccation⁶⁻¹². These functions are related to their ability to form gels^{2-4,13-16}, typically in the presence of divalent metal cations. The three most important natural derivatives of polyuronates are pectins^{2-4,8-10,12,15,17}, alginates^{2-4,7,18} and glucuronans¹⁹⁻²¹. The properties of natural polyuronate-based materials are largely used in the food industry, where these polysaccharides serve as stabilizing, thickening, gelling or emulsifying agents^{8,18,22}. In addition, they have found many applications^{6,18,23-28} in the fields of medicine (encapsulation, prints), biochemistry (chromatography, culture media), biotechnology (immobilization of enzymes) and environmental protection (detoxification of industrial wastes). Most of these industrial and technological applications also rest on the ability of polyuronates to form gels.

Alginates^{2-4,7,18} are linear copolymers of β-D-mannuronate (β-D-ManU) and α-L-guluronate (α-L-GulU) residues in (1→4)-linkage, predominantly alternating homogeneous segments of either types of residues and regular repeats of their dimer²⁹. The relative amounts of the two acids and their distribution along the polymer chain vary widely depending on the natural source^{7,18,30,31}, and strongly influence the physico-chemical and biological properties of alginates^{7,30-34}. Alginates are present in large amounts in the extracellular medium of brown seaweeds (20-40% of the dry weight) and, for this reason, constitute the most abundant marine biopolymer.

An aqueous gel is formed when the polysaccharide chains associate through non-covalent interactions at the level of junction zones. These zones are separated by disordered segments consisting in general of only one chain. The framework of associated and disordered segments leaves large-sized cavities filled by water molecules. In the context of natural alginates, the size of the junction zones is typically limited by inhomogeneities in the polymer sequence. Homogeneous polyguluronate (pGulU) segments are believed to be the main contributors to junction zones³⁵ (although regular repeats of the α-[D-ManU-β(1→4)-L-GulU] dimer in (1→4)-linkage may also be involved^{32,36}) while homogeneous polymannuronate (pManU) segments (as well as segments with more heterogeneous sequences) show reduced propensities to chain-chain association. This size limitation of the junction zones prevents the formation of crystalline structures (precipitation) and, consequently, stabilizes the polymer-water system in the gel state.

The gelation of polyuronates is generally induced by the addition of divalent metal cations (typically Ca²⁺) at a pH close to neutrality. In this case, the leading intermolecular interactions responsible for chain-chain association in the junction zones are ionic interactions between the negatively-charged carboxylate groups and bridging divalent cations^{2-4,35,37-39}. The most popular structural model for calcium-induced chain-chain association between pGulU segments in alginates is the "egg-box" model^{37,40}, suggesting that pairs of segments in a 2₁-helical conformation are associated in an antiparallel fashion, leaving "niches" for the calcium ions that are coordinated by (two) carboxylate and (eight) hydroxyl groups of the two chains^{37,40-43}. This model is compatible with the selectivity for calcium over monovalent (*e.g.* sodium) ions^{37,44} (*i.e.* once bound within the gel, the Ca²⁺ ions cannot be displaced by Na⁺ ions) and the stoichiometry of calcium binding^{37,44-46} (one Ca²⁺ ion for four uronate residues), as well with the 2₁-helical conformation observed in the solid state for pGulU (acidic form⁴⁷⁻⁵¹, Na⁺ salt⁵² and, presumably, Ca²⁺ form⁵⁰). However, the egg-box model may not be the only model compatible with the avail-

able experimental data and has been questioned in a number of recent studies^{36,42,43,50,51,53–55}. In summary⁵, it appears that the original egg-box model^{37,40} is excessively idealized in view of all recent data. However, irrespective of the details of the chain conformation and cation coordination mode, a primary dimerization mechanism involving chain rigidification (regular helical structure) and tight ion binding (well defined coordination sites) at junction zones associating (principally) homogeneous pGulU segments is generally accepted for the cation-induced gelation of alginates. Note, however, that further (presumably weaker; see however⁴⁶) aggregation of the dimeric junction zones may occur when Ca^{2+} is the predominant counter-ion in solution^{42,46,50} (but not in the presence of a large excess of univalent cations^{44,56,57}) or upon dehydration⁵⁵.

One important factor influencing the gelation of polyuronates in general (and of alginates in particular), as well as the properties of the resulting gels, is the nature and concentration of the cations present in solution. The experimental investigation of the influence of different counter-ions is not only interesting in terms of the elucidation of the gel formation mechanism, but also in terms of environmental issues^{58–61} (*e.g.* concentration of heavy metals by plant roots or bacterial exopolysaccharides) as well as practical applications (*e.g.* detoxification of heavy metal containing wastes^{62–67}, inhibition of the uptake of radioactive isotopes by the human body⁶⁸, fouling of ultrafiltration or reverse-osmosis membranes^{69,70}, stabilization of *in situ* drug release systems⁷¹).

The results of independent experimental studies concerning the cation binding selectivity of alginates are sometimes difficult to compare in details, due to differences in the considered polymers (*e.g.* degree of polymerization, GulU vs ManU content and distribution), experimental conditions (*e.g.* soluble polymer vs gel, detailed composition of the aqueous environment) and monitored properties (*e.g.* cation-exchange coefficients, gel strength, spectroscopic properties). However, considering simple (alkali and alkali-earth) ions the following main trends appear to be largely systematic:

1. All (alkali-earth) divalent metal cations (except for the Mg^{2+} cation^{15,72}) are capable of promoting alginate gel formation. The selectivity towards these cations (as probed *e.g.* by cation-exchange coefficients or by the concentration of cations required to promote gel formation and the strength and permeability of the resulting gel) generally follows the sequence^{14,63,68,73–81} $\text{Ba}^{2+} > \text{Sr}^{2+} > \text{Ca}^{2+} \gg \text{Mg}^{2+}$. This sequence appears to be dominated by the selectivity towards pGulU segments^{63,68,75,76,81} (and inversions have been reported for ManU-rich alginates⁸¹). The same sequence appears to hold for both soluble (single chain) as well as gel (junction zones) alginates, although the selectivity is weaker in the former case⁷⁸.
2. The salts of alginates with monovalent (alkali) metal cations are typically water soluble (no gel formation in the absence of divalent cations). The selectivity towards these cations (as probed by ellipticity changes in circular dichroism measurements) appears to follow the sequence⁸² $\text{Li}^+ < \text{Na}^+ < \text{K}^+ > \text{Rb}^+ > \text{Cs}^+$. Here also, the sequence appears to be dominated by the selectivity towards pGulU segments⁸². Note also that the experimental results suggest a peculiar binding mode for the Na^+ cations (which was tentatively interpreted as resulting from egg-box like interchain interactions⁸²). Independent measurements of cation-exchange coefficients also suggested a slightly enhanced binding of Na^+ compared to K^+ in soluble (single chain) alginates⁸³ (*i.e.* an inversion between Na^+ and K^+ in the above sequence).
3. Alginate chains evidence preferential affinity for divalent (alkali-earth) over monovalent

(alkali) metal cations. This selectivity is well known for the gels^{37,44} (the divalent cations bound in junction zones cannot be displaced by monovalent cations) and also appears to hold for soluble (single-chain) alginates^{63,84,85}. Here also, the preferential affinity appears to be dominated by interactions with the pGulU segments^{63,85}.

For completeness, note that numerous studies have also addressed the (practically highly relevant) problem of the cation specificity of polyuronates towards heavy metal (as well as lanthanides and including some trivalent) cations^{35,58–61,63,64,66,67,73,76,77,79,80,86–91}. Note finally that the selectivity of alginates towards various mono- or divalent cations (as summarized above) may differ from that observed for other polyanionic biopolymers^{14,72,77,86,87,91–99} (*e.g.* other polyuronates, C₆-oxidized cellulose, carrageenans or sulfated polysaccharides).

The effect of different counter-ions (charge, size and concentration) on the gelation and resulting gel properties of polyuronates in general (and alginates in particular) can be thought of as resulting from differences in polysaccharide-ion interactions in two distinct states: (*i*) free single-chains in aqueous solution; (*ii*) associated chains in junction zones. The present study exclusively focuses on the first state (the investigation of the second state will be the scope of a subsequent study).

In the free (unassociated) state, a polyuronate chain can be idealized as an infinite rod with regularly positioned negative charges, surrounded by (bound or aqueous) cations and neutralizing (aqueous) anions. There exist a number of theoretical approaches to model such a rod-like polyanion surrounded by a aqueous counter-ion atmosphere. These include (in order of increasing complexity and expected accuracy): (*i*) analysis of the potential generated by the polyanion in vacuum (chain: atomistic, rigid; solvent: neglected; ion-ion interactions: neglected) and deduction of possible binding sites^{43,100}; (*ii*) analysis of the counter-ion distribution around the polyanion (chain: atomistic or simplified charge distribution, rigid) using simple solvation models based on continuum electrostatics (solvent: implicit) along with different approximate treatments of interionic interactions (ion-ion interaction: mean-field plus possible corrections), *e.g.*¹⁰¹ linearized Poisson-Boltzmann^{102,103}, Poisson-Boltzmann^{104–107} or Manning condensation^{105,106,108–112} and extended^{113–115} approaches; (*iii*) analysis of counter-ion trajectories (and corresponding calculated distributions) around the polyanion (chain: atomistic or simplified charge distribution; rigid or flexible) using Monte Carlo (MC), molecular dynamics (MD), stochastic dynamics (SD) or Brownian dynamics (BD) simulation techniques with an explicit representation of the counterions along with an implicit (MC, SD, BD) or explicit (MC, MD) representation of the solvent molecules¹¹⁶. Although these approaches have been most commonly applied to the study of the deoxyribonucleic acid double-helix, some have also been applied to polyuronates^{5,43,100,117,118}. One of the main observations resulting from the comparison of various approaches is that the neglect (or, to a lesser extent, the mean-field treatment) of ion-ion interactions does not permit even a qualitatively accurate description of these systems in the functionally relevant concentration regime.

In a previous study⁵ (see Chapter 4 of the present thesis), we applied an approach of the third type (chain: atomistic, partially flexible; ions and solvent: explicit; MD) to investigate the binding of calcium ions (neutralizing amount) to four types of homopolyuronate single-chains at 300 K, starting from alternative regular two- and three-fold helical structures. In this work, the chains were made formally infinite by application of artificial periodicity along the chain axis (with octameric or nonameric repeat units). Two main observations made in these simulations

were that : (i) the glycosidic linkages (and local helical parameters) showed an important flexibility (in time) and variability (along the chains), and regular helical structures only accounted for a limited fraction of the conformational ensembles populated at 300 K; (ii) for all systems considered, the binding of the calcium counter-ions was essentially non-specific, with the formation of a dense counter-ion atmosphere around the chains (condensation), but no specific (tight-binding) interactions at well-defined coordination sites. Taken together, these observations suggested that if chain dimers are appropriately described by the egg-box model (or any other structural model with similar qualitative features), chain-chain association within junction zones in gels must be accompanied by a substantial chain stiffening and a dramatic change in the ion-binding mode. A possible interpretation of this process is the following. The tight binding of a calcium ion (contact ion pair, *i.e.* including partial desolvation) is an energetically unfavorable process when a single carboxylate group is involved. It requires an interaction with two carboxylate groups that are optimally positioned and oriented (and probably also the assistance of sugar hydroxyl groups). Single-chain polyuronates are unable to provide such a geometry (or only at the expense of a too demanding conformational entropy decrease), and the counter-ions remain bound in a non-specific way within a dense counter-ion atmosphere. However, the carboxylate and counter-ion charges are not locally compensated and both the chain and the ionic atmosphere are associated with important (positive) self energies due to the Coulombic repulsion between like charges. In other words, the calcium ions around the single chain can be viewed as being placed in a "reactive" state. The possibility of forming an appropriate chelating environment (involving a local compensation of the counter-ion charges and thus, the effective removal of inter-ionic repulsions) by association of a second polyuronate chain is likely to represent the leading driving force (in addition to presumably weaker direct chain-chain interactions) permitting chain stiffening, cation immobilization and chain-chain association.

The goal of the present study is to investigate and compare the interaction of a set of monovalent (Na^+ , K^+) and divalent (Mg^{2+} , Ca^{2+}) cations with single-chain pGulU (at 300 K and 1 bar) using explicit-solvent MD simulation. The approach employed is similar to that used in previous work⁵ (see Chapter 4 of the present thesis), based on a (periodic) dodecameric pGulU chain in 2_1 -helical conformation, and considers 14 possible combinations in terms of possible (neutralizing) solution environments (single type of cation or simultaneous presence of two types of cation, either in the absence or presence of chloride anions).

5.3 Computational methods

5.3.1 Molecular dynamics simulation

All MD simulations in this study were performed using the GROMOS96 program^{119,120}, together with GROMOS the 45A4 force field¹²¹⁻¹²⁶ (encompassing a recently reoptimized parameter set for hexopyranose-based carbohydrates¹²⁶⁻¹³⁰, including uronates⁵) and the SPC water model¹³¹. The non-bonded interaction parameters for all the ions were taken directly from the 45A4 parameter set except for K^+ ion (parameters optimized based on experimental single-ion solvation free energies assuming a value of $-1100 \text{ kJ mol}^{-1}$ for the proton; M. Reif; personal communication) and are reported in Table 5.1.

A total of 14 explicit-solvent simulations were carried out involving a single (periodic) pGulU

| ion | $C_6(I,I)$ | $C_6(I,W)$ | $C_6(I,O)$ | $C_{12}(I,I)$ | $C_{12}(I,W)$ | $C_{12}(I,O)$ | $R(I,I)$ | $\epsilon(I,I)$ | $R(I,W)$ | $\epsilon(I,W)$ | $R(I,O)$ | $\epsilon(I,O)$ |
|------------|--------------------------------------|------------|------------|---|---------------|---------------|----------|-------------------|----------|-------------------|----------|-------------------|
| | [10^{-4} kJ mol $^{-1}$ nm 6] | | | [10^{-8} kJ mol $^{-1}$ nm 12] | | | [nm] | [kJ mol $^{-1}$] | [nm] | [kJ mol $^{-1}$] | [nm] | [kJ mol $^{-1}$] |
| Na $^+$ | 0.7206 | 4.3429 | 4.0374 | 2.1025 | 23.5335 | 44.4860 | 0.2891 | 0.0617 | 0.3204 | 0.2000 | 0.3607 | 0.0916 |
| K $^+$ | 12.9600 | 18.4176 | 17.1216 | 186.5600 | 221.6856 | 419.0580 | 0.3771 | 0.2251 | 0.3661 | 0.3825 | 0.4121 | 0.1748 |
| Mg $^{2+}$ | 0.6528 | 4.1337 | 3.8418 | 0.3408 | 9.4750 | 17.9109 | 0.2169 | 0.3126 | 0.2777 | 0.4508 | 0.3126 | 0.2060 |
| Ca $^{2+}$ | 10.0489 | 16.2177 | 15.0765 | 49.8013 | 114.5351 | 216.5088 | 0.3157 | 0.5069 | 0.3349 | 0.5741 | 0.3770 | 0.2624 |
| Cl $^-$ | 138.0625 | 60.1130 | 55.8830 | 10691.5600 | 1678.1820 | 890.3774 | 0.4993 | 0.4457 | 0.4218 | 0.5387 | 0.3836 | 0.8768 |

Table 5.1: Non-bonded interaction parameters for the considered ions. The reported entries for the 5 ions are the Lennard-Jones (London) dispersion and (Pauli) repulsion coefficients C_6 and C_{12} , as well as the distance R and energy ϵ at the minimum of the Lennard-Jones interaction curve, for ion-ion (I,I), ion-water (I,W) and ion-carboxylate oxygen atom (I,O) interactions. The Lennard-Jones interaction energy is $E_{LJ} = C_{12}r^{-12} - C_6r^{-6} = \epsilon[(R^{-1}r)^{-12} - 2(R^{-1}r)^{-6}]$ where $R = (2C_{12})^{1/6}C_6^{-1/6}$ and $\epsilon = (4C_{12})^{-1}C_6^2$. The GROMOS force-field relies on a geometric combination rule for C_{12} and C_6 (or, equivalently, for ϵ and R). However, the C_{12} parameters used in the combination rule for interactions between charged species (I,I and I,O) differ from those used for interactions between charged and polar species (I,W). The C_6 and C_{12} parameters for the oxygen of the SPC water model¹³¹ are $26.1735 \cdot 10^{-4}$ kJ mol $^{-1}$ nm 6 and $263.4129 \cdot 10^{-8}$ kJ mol $^{-1}$ nm 12 , respectively.

chain in different counter-ion solutions (see below) at a temperature of 300 K and a pressure of 1 bar. The simulations were carried out under periodic boundary conditions based on cubic computational boxes (box edge L). The equations of motion were integrated using the leapfrog scheme¹³² with a timestep of 2 fs. The SHAKE algorithm¹³³ was applied to constrain all bond lengths with a relative geometric tolerance of 10^{-4} . The temperature was maintained close to its reference value (300 K) by weakly coupling the solute and solvent degrees of freedom separately to a heat bath¹³⁴, with a relaxation time of 0.1 ps. The pressure was also maintained close to its reference value (1 bar) by weakly coupling the particle coordinates and box dimensions (isotropic coordinate scaling) to a pressure bath¹³⁴, with a relaxation time of 0.5 ps and an isothermal compressibility of $0.4575 \cdot 10^{-3}$ kJ $^{-1}$ mol nm 3 as appropriate for water¹¹⁹. The center of mass motion was removed every 200 ps. Non-bonded interactions were computed using a twin-range scheme^{116,119}, with short- and long-range cutoff distances set to 0.8 and 1.4 nm respectively, and a frequency of 5 timesteps for the update of the short-range pairlist and intermediate-range interactions. A reaction-field correction^{135,136} was applied to account for the mean effect of omitted electrostatic interactions beyond the long-range cutoff distance, using a relative dielectric permittivity of 61 as appropriate for the SPC water model¹³⁷. All simulations were carried out for a duration of 10 ns after an equilibration period of 1 ns, and coordinates were saved to file every 1 ps for subsequent analysis.

To avoid the presence of chain-end effects and the reduction of cooperativity associated with the simulation of finite (short) oligomers^{117,138–143}, the choice was made to simulate infinite chains by taking advantage of the periodic boundary conditions⁵ (note that an alternative approach has been recently suggested¹⁴³). To this purpose, a polymer segment consisting of 12 residues within the computational box (with the chain axis aligned along the z -direction) was made infinite by linking the two terminal residues *via* a glycosidic bond across the periodic boundary (resulting in a formally cyclic topology¹⁴⁴). In this setup, the 12 residues are thus associated through 12 glycosidic linkages connecting residues i to $i+1$ with $i=1\dots 11$, the last linkage connecting residue

12 to residue 1. This approach is expected to account more appropriately for the cooperativity effects present within longer polymer chains and to entirely eliminate chain-end effects. However, one should be aware that the inclusion of an artificial periodic constraint with a repeat unit of 12 residues (along with an essentially fixed end-to-end vector for this 12 residue repeat, as imposed by the isotropic pressure coupling employed in the simulations and the limited compressibility of water) results in a restriction of the chain longitudinal and torsional flexibility⁵. The effect of the artificial longitudinal constraint is expected to be limited in view of long persistence length (about 40 residues^{5, 145, 146}) of pGulU. The effect of the torsional constraint is also expected to be limited by the consideration of a dodecameric periodic repeat (longer than in previous work⁵ and compatible with both regular 2- and 3-fold helices). The initial structure for the simulations was

| System | Na ⁺ | K ⁺ | Mg ²⁺ | Ca ²⁺ | Cl ⁻ | H ₂ O | <i>L</i> | <i>I</i> | <i>I</i> _{tot} | κ^{-1} | κ_{tot}^{-1} |
|--------|-----------------|----------------|------------------|------------------|-----------------|------------------|----------|-------------------------|-------------------------|---------------|---------------------|
| | | | | | | | [nm] | [mol kg ⁻¹] | [mol kg ⁻¹] | [nm] | [nm] |
| Na | 12 | - | - | - | - | 4147 | 5.05 | 0.08 | 0.16 | 1.08 | 0.76 |
| K | - | 12 | - | - | - | 4147 | 5.06 | 0.08 | 0.16 | 1.08 | 0.76 |
| NaCl | 24 | - | - | - | 12 | 4123 | 5.04 | 0.24 | 0.32 | 0.62 | 0.54 |
| KCl | - | 24 | - | - | 12 | 4123 | 5.05 | 0.24 | 0.32 | 0.62 | 0.54 |
| NaKCl | 12 | 12 | - | - | 12 | 4123 | 5.05 | 0.24 | 0.32 | 0.62 | 0.54 |
| Mg | - | - | 6 | - | - | 4153 | 5.05 | 0.16 | 0.24 | 0.76 | 0.62 |
| Ca | - | - | - | 6 | - | 4153 | 5.05 | 0.16 | 0.24 | 0.76 | 0.62 |
| MgCl | - | - | 12 | - | 12 | 4135 | 5.04 | 0.40 | 0.48 | 0.48 | 0.44 |
| CaCl | - | - | - | 12 | 12 | 4135 | 5.04 | 0.40 | 0.48 | 0.48 | 0.44 |
| MgCaCl | - | - | 6 | 6 | 12 | 4135 | 5.05 | 0.40 | 0.48 | 0.48 | 0.44 |
| MgNaCl | 12 | - | 6 | - | 12 | 4129 | 5.04 | 0.32 | 0.40 | 0.54 | 0.48 |
| MgKCl | - | 12 | 6 | - | 12 | 4129 | 5.05 | 0.32 | 0.40 | 0.54 | 0.48 |
| CaNaCl | 12 | - | - | 6 | 12 | 4129 | 5.05 | 0.32 | 0.40 | 0.54 | 0.48 |
| CaKCl | - | 12 | - | 6 | 12 | 4129 | 5.05 | 0.32 | 0.40 | 0.54 | 0.48 |

Table 5.2: Summary of the simulated systems. The reported entries for the 14 simulations are: the simulation code, the types and numbers of counter-ions in the simulation box, the number of water molecules in the computational box, the average (cubic) box edge length (*L*) during the 10 ns simulations, the ionic strength (*I*: counter-ions only; *I*_{tot}: including the saccharide chain) and the Debye screening length (κ^{-1} : counter-ions only; κ_{tot}^{-1} : including the saccharide chain). The (root-mean-square) fluctuations of *L* are between 0.004-0.006 nm for all simulations. The ionic strength is calculated as $I = 1/2 \sum b_i z_i^2$ where *b_i* is the solution molality of species *i* (mol solute per kg solvent) and *z_i* the integer charge of the species (values including the saccharide consider individual monomeric units with an integer charge of -1). The Debye screening length is calculated as $\kappa^{-1} = CI^{-1/2}$ with $C = [(2N_A e^2 \rho_s)^{-1} \epsilon_0 \epsilon_s k_B T]^{1/2}$, where *N_A* is Avogadro's number, *e* the electron charge, ρ_s the solvent density, ϵ_0 the permittivity of vacuum, ϵ_s the relative dielectric permittivity of the solvent, *k_B* Boltzmann's constant and *T* the absolute temperature. Using the experimental values¹⁴⁷ $\rho_s = 997.048 \text{ kg m}^{-3}$ and $\epsilon_s = 78.36$ at *T* = 300 K, one has $C = 0.305 \text{ nm mol}^{1/2} \text{ kg}^{-1/2}$ (used for this table). Note that the use of the appropriate values for the SPC water model, namely^{137, 148} $\rho_s = 994 \text{ kg m}^{-3}$ and $\epsilon_s = 61$ would lead to a slightly different value $C = 0.270 \text{ nm mol}^{1/2} \text{ kg}^{-1/2}$. All simulations were carried out for 10 ns duration (after 1 ns equilibration) at 300 K and 1 bar.

chosen to be a regular 2₁-helix (6 full turns per dodecameric repeat) with a turn angle per residue of 180° and a rise per residue $h_{ref} = 0.427 \text{ nm}$. The latter value was taken from the modeling

studies by Braccini *et al.*¹⁰⁰ (based on MM3 relaxed-residues energy maps for the corresponding diuronic acids) and is also close to the experimental value of 0.436 nm based on the fiber diffraction experiments by Atkins *et al.*⁴⁷⁻⁴⁹ and Sikorski *et al.*⁵⁰. This value is, however, lower than the value of 0.459 nm used in previous work⁵. Initial coordinates for this chain configuration were generated using the program InsightII¹⁴⁹, assuming ideal ⁴C₁ ring geometries and using the reference ϕ and ψ glycosidic dihedral angle values ($\phi_{ref} = 275.8^\circ$ and $\psi_{ref} = 84.7^\circ$) reported by Braccini *et al.*¹⁰⁰. The initial cubic box dimension L was set to $12 h_{ref}$. The computational boxes corresponding to the different systems were brought to overall neutrality by inclusion of counterions at random positions (see below) and filled to a water density of about $1 \text{ g}\cdot\text{cm}^{-3}$. After energy minimization, equilibration was carried out by performing 0.2 ns MD simulation, increasing the temperature from 50 to 300 K in six successive steps followed by 0.8 ns at 300K. The resulting configurations were then used for the production simulations. The 14 simulations involved different counter-ion solutions in the form of overall neutralizing sets of the different ions considered (Na^+ , K^+ , Mg^{2+} , Ca^{2+} and Cl^-) dissolved in about 4000 water molecules. The different combinations investigated (along with the resulting average box sizes, solution ionic strengths I and associated Debye screening lengths κ^{-1}) are summarized in Table 5.2. For the ease of reference, all simulations are labeled with a unique code referring to the counter-ion solution involved.

5.3.2 Analysis of the trajectories

The analysis of the trajectories was performed in terms of : (i) local helical conformation and intramolecular hydrogen-bonding; (ii) ion distribution and binding; (iii) effective chain charge (charge of a cylindrical volume of given radius centered on the helix axis); (iv) chain (rotational, longitudinal and transverse) and ionic (three-dimensional, longitudinal and transverse) diffusion. The detailed procedures for these analyzes are described elsewhere⁵ (see Chapter 4 of the present thesis) and only the most important points are summarized below.

The local helical conformation at a glycosidic linkage was assessed by monitoring the corresponding turn angle. This angle is calculated as the dihedral angle between the "reference vectors" of two successive residues, this vector being defined here as the projection of the vector connecting the midpoint of the C₂-C₃ bond to the midpoint of the C₅-O₅ bond onto the xy -plane. Values within ranges of $\pm 30^\circ$ around 120° , 180° or 240° were associated to local 3₁-, 2₁- or 3₂-helical conformations, respectively.

The intramolecular hydrogen-bonding was only analyzed in terms of the corresponding occurrences of the interresidue H₂→O₆' and O₅←H₃' hydrogen-bonds at the successive glycosidic linkages (where the prime residue is next to the unprimed residue towards the reducing end of the chain), since these two types of hydrogen-bonds were the only ones found to be populated in regular polyuronate chains⁵. The presence of a hydrogen-bond was defined by a maximal hydrogen-oxygen distance of 0.25 nm and a minimal oxygen-hydrogen-oxygen angle of 135° . A hydrogen-bond between a hydroxyl and a carboxylate group was assumed to be present if the hydroxyl group formed a hydrogen-bond with either (or both) of the carboxylate oxygen atoms.

The counter-ion distribution was analyzed (for each ion type separately) in terms of the two-dimensional radial distribution function $g_{2D}(r)$ corresponding to the (minimum-image) distance r (in the xy -plane) between the helix axis and the ions of the given type. This function is defined in such a way that the integral of $L^{-2}2\pi r g_{2D}(r)$ over the entire distance range is equal to one. The counter-ion distribution was also analyzed in terms of the (three-dimensional) radial distribution

function $g_{3D}(\rho)$ corresponding to the smallest (minimum-image) distance ρ between any of the two carboxylate oxygen atoms and the closest counter-ion. This function is defined in such a way that the integral of $L^{-3}4\pi\rho^2g_{3D}(\rho)$ over the entire simulation box (or any larger volume) is equal to one. All radial distribution functions were computed with a bin width of 0.01 nm. In addition, cation-binding events were monitored for the 12 residues, by assigning two cutoff values ρ_t and ρ_l (with $\rho_t < \rho_l$) to the distance ρ . The condition $\rho \leq \rho_t$ was associated to a tight binding event (TB; contact ion pair), the condition $\rho_t < \rho \leq \rho_l$ to a loose binding event (LB; solvent-separated ion pair), while $\rho > \rho_l$ corresponded to the absence of ion binding. The values of ρ_t and ρ_l were chosen as the locations of the first and the second minima in $g_{3D}(\rho)$ for a given type of ion, *i.e.* the values of ρ_t were set to 0.30 (Na⁺), 0.36 (K⁺), 0.32 (Mg²⁺) or 0.33 (Ca²⁺) nm, while the corresponding values of ρ_l were set to 0.54 (Na⁺), 0.61 (K⁺), 0.52 (Mg²⁺) or 0.57 (Ca²⁺) nm. The binding (LB or TB) of the carboxylate group of a given residue to a cation that is simultaneously bound (LB or TB) to the carboxylate group of a different residue was referred to as a double binding (DB; ion bridging) event. Two different types of DB events were further distinguished depending on whether they involved the carboxylate groups of two neighbor residues (DB_n) or of two second-neighbor residues (DB_s; *i.e.* separated by single residue along the chain). Finally, the counter-ion distributions were also visualized by superimposing successive trajectory frames (chain and counter-ions) on the initial (equilibrated) chain configuration (rotational and translational fit based on the C₂, C₃, C₅ and O₅ atoms), and displaying the positions of all ions at successive 20 ps intervals.

The effective chain charge $Q(r)$ is defined as the net charge of a cylindrical volume of radius r centered on the helix axis (and of length corresponding to the simulated dodecameric unit). This quantity was evaluated as a function of r by summation of the charges contributed by all counter-ions (calculated *via* integration of the corresponding $g_{2D}(r)$ functions) with the boundary condition $Q(0) = -12e$ (chain charge).

Diffusion coefficients were calculated by monitoring the mean-square displacements $Q_{\mathbf{q}}(t)$ of various time dependent quantities $\mathbf{q}(t)$, with averaging over all possible time origins. The corresponding diffusion coefficients $D_{\mathbf{q}}$ were calculated based on the Einstein relation, taking into account the dimensionality n (1, 2 or 3) of the quantity \mathbf{q} . In practice, $D_{\mathbf{q}}$ was estimated from the slope of a regression line fitting $Q_{\mathbf{q}}(t)$ over the interval from 0 to 3 ns where these functions were found to be approximately linear. The quantities \mathbf{q} considered and the corresponding diffusion coefficients $D_{\mathbf{q}}$ were: (i) the chain orientation angle Θ (mean of the "reference angles" of all residues, this angle being defined by the deviation of the above defined reference vector relative to its orientation in the initial configuration; with $n=1$), to evaluate the chain rotational diffusion constant D_{Θ} ; (ii) the z -coordinate of the chain center (center of geometry of the "reference points" of all residues, this point being defined here as the center of geometry of the ring atoms C₂, C₃, C₅ and O₅; with $n=1$), to evaluate the chain longitudinal diffusion constant D_z ; (iii) the coordinate vector of the chain center in the xy -plane (with $n=2$), to evaluate the chain transverse diffusion constant D_{xy} ; (iv) the coordinate vector of a given counter-ion (with $n = 3$) to evaluate the three-dimensional diffusion constant $D_{i,r}$ for all counter-ions; (v) the z -coordinate of a given counter-ion (with $n=1$), to evaluate longitudinal diffusion constants $D_{i,z}$ for all counter-ions; (vi) the coordinate vector of a given counter-ion in the xy -plane (with $n=2$), to evaluate the transverse diffusion constants $D_{i,xy}$ for all counter-ions. While evaluating the corresponding time series $\mathbf{q}(t)$, care was taken to follow periodic coordinates as they diffused across different periods, without "refolding" to the reference interval.

5.4 Results

5.4.1 Chain conformation and hydrogen-bonding

The time evolution of the local helical conformation and interresidue hydrogen-bonding at the level of the 12 glycosidic linkages was found to be qualitatively very similar for all linkages and in all simulations. The results for simulation Na are displayed in Figure 5.1 as an illustrative example. The corresponding average occurrences are reported in Table 5.3 for the 14 simulations. In all cases, the number of full turns per dodecameric unit (sum of the turn angles of the 12 glycosidic linkages divided by 360°) remained exactly 6 throughout the entire simulation time (as appropriate for a chain conformation close to a 2_1 -helix). The average rise per residue (average box edge length L divided by 12; Table 5.2) of about 0.420-0.422 nm is also the one appropriate for an ideal pGulU 2_1 -helix (Section 5.3.1). Locally, the 2_1 -helical conformation is also found to be predominant (74.0-77.3 %), although alternative 3_1 - (11.3-13.6 %) and 3_2 - (10.5-12.4 %) conformations are also transiently observed. The $H_2 \rightarrow O'_6$ hydrogen-bond is the only one found in the simulations (66.7-68.9 %), while the $O_5 \leftarrow H_3'$ hydrogen-bond is never observed. The occurrences of local helical conformations and intramolecular hydrogen-bonds found in the present simulations differ slightly (about 10% at most) from those reported in previous work⁵ for pGulU. This is probably due to the consideration of dodecamers (rather than octamers) and to the slightly smaller value chosen here for the rise per residue.

5.4.2 Distribution of the ions around the chain

The two-dimensional radial distribution functions $g_{2D}(r)$ of the different ions around the helix axis are displayed in Figures 5.2-5.4 for the 14 simulations. For the ease of comparison, a summary of the main features of these curves is also provided in Table 5.4.

The following general observations apply to all systems considered: (i) in the absence of Cl^- anions (simulations Na, K, Mg and Ca), the cation radial distributions are relatively narrow (narrow *vs* broad referring here to the distance range spanned by the distribution) and tight (tight *vs* loose referring here to the mean distance from the helix axis), presenting a single maximum at about 0.5-0.6 nm with a magnitude of about 6 (Na^+ , K^+) or 8 (Mg^{2+} , Ca^{2+}), and essentially reaching zero (<0.25) beyond about 1.7 (Na^+ , K^+) or 1.4 (Mg^{2+} , Ca^{2+}) nm; (ii) in the presence of Cl^- anions along with a single cation type (simulations NaCl, KCl, MgCl and CaCl) or isovalent cation types (simulations NaKCl and MgCaCl), the cation distributions tend to be somewhat broader and looser (peak location usually shifted to a slightly higher distance, peak height decreased by factor of about 1.8, appearance of a long-range tail in the distribution), compared to the corresponding simulations involving a single cation type in the absence of Cl^- (simulations Na, K, Mg and Ca); (iii) in the presence of Cl^- ions along with cation types of different valences (simulations MgNaCl, MgKCl, CaNaCl and CaKCl), the distribution of the monovalent cation is significantly broadened and loosened (peak location shifted to about 0.8-0.9 nm, peak height decreased by a factor of about 2.2 in the presence of Mg^{2+} or 2.9 in the presence of Ca^{2+} , pronounced long-range tail in the distribution) compared to the corresponding simulation involving the monovalent ion alone (simulations Na and K), while the corresponding change in the distribution of the divalent cation (relative to simulations Mg and Ca) is much more limited; (iv) the distributions of the Cl^- anion (whenever present) are very broad and loose, presenting a

| system | 2_1 [%] | 3_1 [%] | 3_2 [%] | $H_2 \rightarrow O_6'$ [%] | $O_5 \leftarrow H_3'$ [%] |
|--------|--------------|--------------|--------------|-------------------------------|------------------------------|
| Na | 76.2 | 12.4 | 11.3 | 67.2 | 0.0 |
| K | 76.7 | 12.3 | 11.0 | 67.6 | 0.0 |
| NaCl | 76.6 | 12.2 | 11.0 | 67.0 | 0.0 |
| KCl | 76.3 | 12.4 | 11.1 | 66.7 | 0.0 |
| NaKCl | 78.1 | 11.3 | 10.5 | 68.0 | 0.0 |
| Mg | 76.4 | 12.0 | 11.4 | 67.3 | 0.0 |
| Ca | 75.2 | 12.9 | 11.9 | 66.9 | 0.0 |
| MgCl | 75.3 | 12.8 | 11.8 | 67.8 | 0.0 |
| CaCl | 77.3 | 11.8 | 10.7 | 68.9 | 0.0 |
| MgCaCl | 75.8 | 12.5 | 11.6 | 67.0 | 0.0 |
| MgNaCl | 76.2 | 12.6 | 11.1 | 66.7 | 0.0 |
| MgKCl | 76.9 | 12.3 | 10.7 | 68.2 | 0.0 |
| CaNaCl | 74.9 | 13.2 | 11.9 | 68.5 | 0.0 |
| CaKCl | 74.0 | 13.6 | 12.4 | 67.0 | 0.0 |

Table 5.3: Occurrences of local helical conformations and interresidue hydrogen-bonds for all simulations (Table 5.2). The occurrences are averaged over the 12 linkages and over the entire simulation time (10 ns). The indicated properties are the formation of a local 2_1 -, 3_1 - or 3_2 -helical conformation and of $H_2 \rightarrow O_6'$ or $O_5 \leftarrow H_3'$ interresidue hydrogen-bonds.

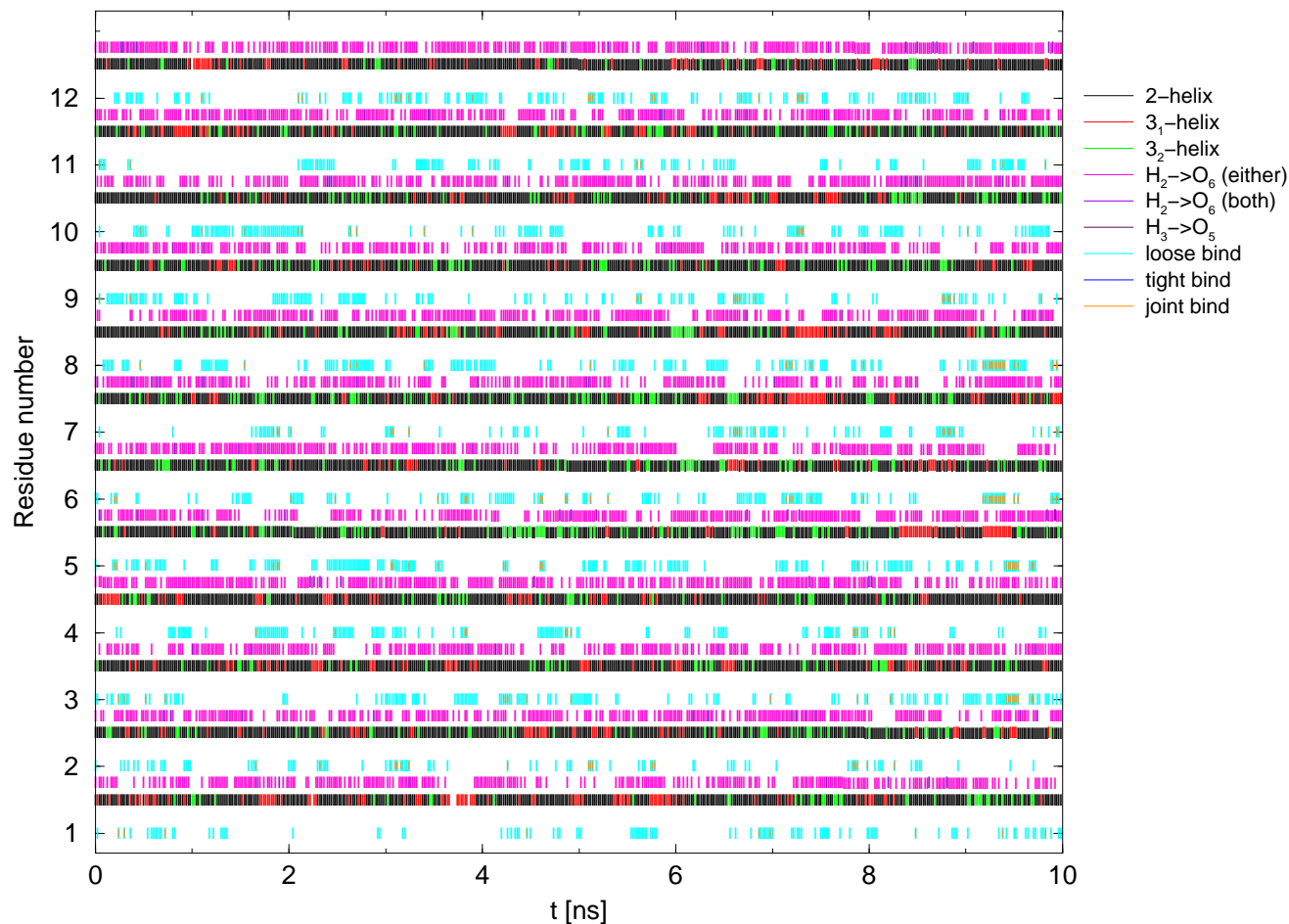


Figure 5.1: Time evolution of the local helical conformation, interresidue hydrogen-bonds and ion-binding events for the 12 glycosidic linkages in simulation Na (Table 5.2). The indicated events are the formation of a local 2₁-, 3₁- or 3₂-helical conformation, of H₂→O₆^l or O₅←H₃^l hydrogen-bonds (H₂→O₆^l either: involving only one carboxylate oxygen atom; H₂→O₆^l: involving both carboxylate oxygen atoms), of a solvent-separated (loose binding; LB) ion pair, of a contact (tight binding; TB) ion pair or of an ion pair (LB or TB) involving an ion also bound (LB or TB) to the carboxylate group of another residue (double binding; DB; smaller bars for the two involved carboxylate groups). See Section 5.3.2 for definitions of local helical conformations, hydrogen-bonding and ion binding. The sampling period (individual bars) is 20 ps. Color code 2₁ (black), 3₁ (red), 3₂ (green), H₂→O₆^l (pink), O₅←H₃^l (brown), LB (cyan), TB (blue), DB (orange).

single maximum at about 1.4-1.7 nm with a magnitude of about 1.4-1.7, and essentially vanishing (<0.25) below about 0.2-0.8 nm.

Qualitatively, these observations can be summarized as follows. When the chain charge is exactly neutralized by a single type of cation, divalent cations form a denser and tighter counter-ion atmosphere around the chain compared to monovalent cations. Upon addition of one equivalent of the chloride salt of the same ion or of an isoivalent ion, the counter-ion atmosphere becomes slightly more diffuse and less tight, as a result of the increased ionic strength (electrostatic screening) and the competition between the chain and the chloride ions (predominantly distributed in the bulk region far away from the chain) for interactions with the cations. However, when one equivalent of the chloride salt of a monovalent cation is added to the chain neutralized by divalent cations (or, equivalently, the inverse process), the divalent cations preferentially interact with the chain (distribution similar to that observed for a chain neutralized by these sole cations) while the monovalent cations preferentially interact with the chloride ions (distribution significantly expanded compared to that observed for a chain neutralized by these sole cations). In other words, single pGulU chains in aqueous solution evidence a clear preferential affinity for divalent over monovalent cations when both types of ions are present in solution.

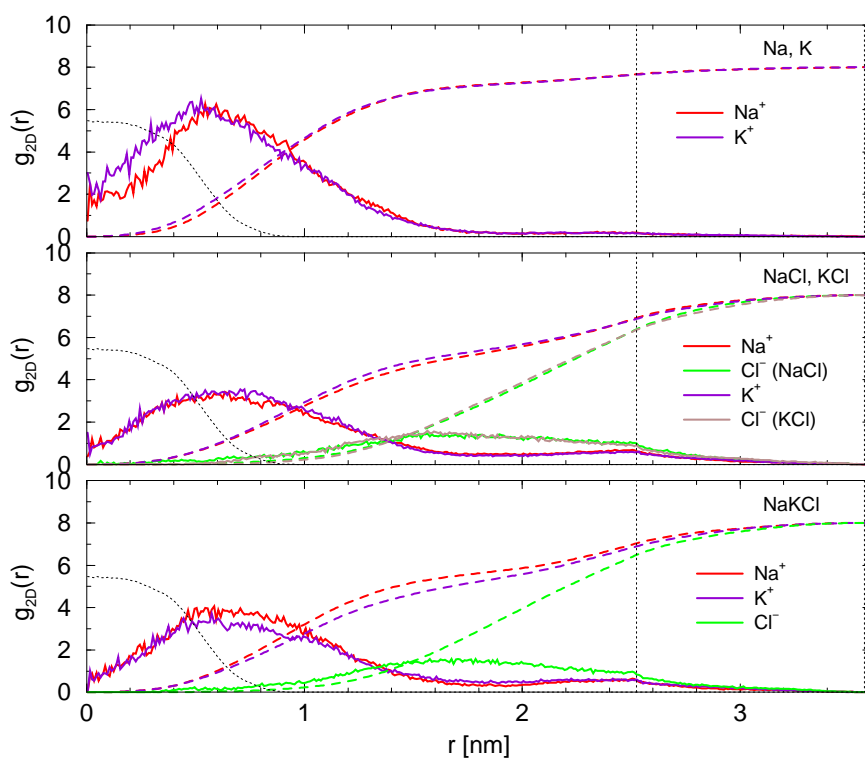


Figure 5.2: Two-dimensional radial distribution function $g_{2D}(r)$ for the coordinates of the counter-ions in the xy -plane relative to the helix axis for simulations Na, K, NaCl, KCl and NaKCl (Table 5.2). The functions are averaged over all ions of a given type in the system and calculated over the entire simulation time (10 ns) using a bin size of 0.01 nm. The cumulative integral of $L^{-2}2\pi r g_{2D}(r)$ (amplified by a factor 8 for readability) is also displayed, as well as the corresponding distribution of all chain atoms (black dotted line; average over all simulations; scaled by a factor 5 for readability).

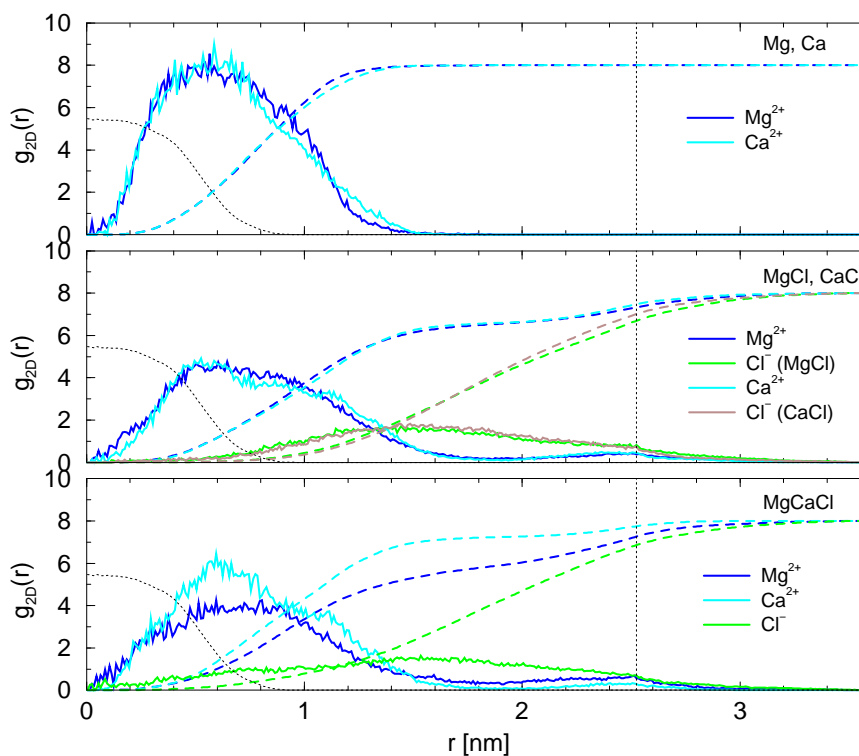


Figure 5.3: Two-dimensional radial distribution function $g_{2D}(r)$ for the coordinates of the counter-ions in the xy -plane relative to the helix axis for simulations *Mg*, *Ca*, *MgCl*, *CaCl* and *MgCaCl* (Table 5.2). The functions are averaged over all ions of a given type in the system and calculated over the entire simulation time (10 ns) using a bin size of 0.01 nm. The cumulative integral of $L^{-2}2\pi r g_{2D}(r)$ (amplified by a factor 8 for readability) is also displayed, as well as the corresponding distribution of all chain atoms (black dotted line; average over all simulations; scaled by a factor 5 for readability).

Besides these general considerations, the detailed comparison of the different systems reveals the following differences between corresponding simulations involving a different type of monovalent (Na^+ or K^+) or divalent (Mg^{2+} or Ca^{2+}) cation. In the absence of divalent cations, the systems involving Na^+ or K^+ (simulations *Na*, *K*, *NaCl*, *KCl* and *NaKCl*) present very similar distributions for all types of ions (except for a slightly tighter distribution for K^+ in simulation *K* compared to Na^+ in simulation *Na* in the immediate vicinity of the chain). The same applies to the systems involving Mg^{2+} or Ca^{2+} separately in the absence of monovalent cations (simulations *Mg*, *Ca*, *MgCl* and *CaCl*). However, when the two types of divalent cations are included simultaneously (simulation *MgCaCl*), the Ca^{2+} distribution is seen to be noticeably tighter compared to that of Mg^{2+} (simultaneously, the Cl^- distribution extends closer to the helix axis compared to all other systems). Finally, in the simulations involving cations of different valences (simulations *MgNaCl*, *MgKCl*, *CaNaCl* and *CaKCl*), a systematic difference is also seen between the systems involving the two different types of divalent cations. Irrespective of the monovalent cation considered, the Mg^{2+} distribution reaches closer to the helix axis (distance range 0.0-0.3 nm) compared to the Ca^{2+} distribution. However, in spite of this close-range difference, the Ca^{2+} distribution appears overall somewhat tighter (in contrast to Mg^{2+} , no long-distance tail is observed for Ca^{2+})

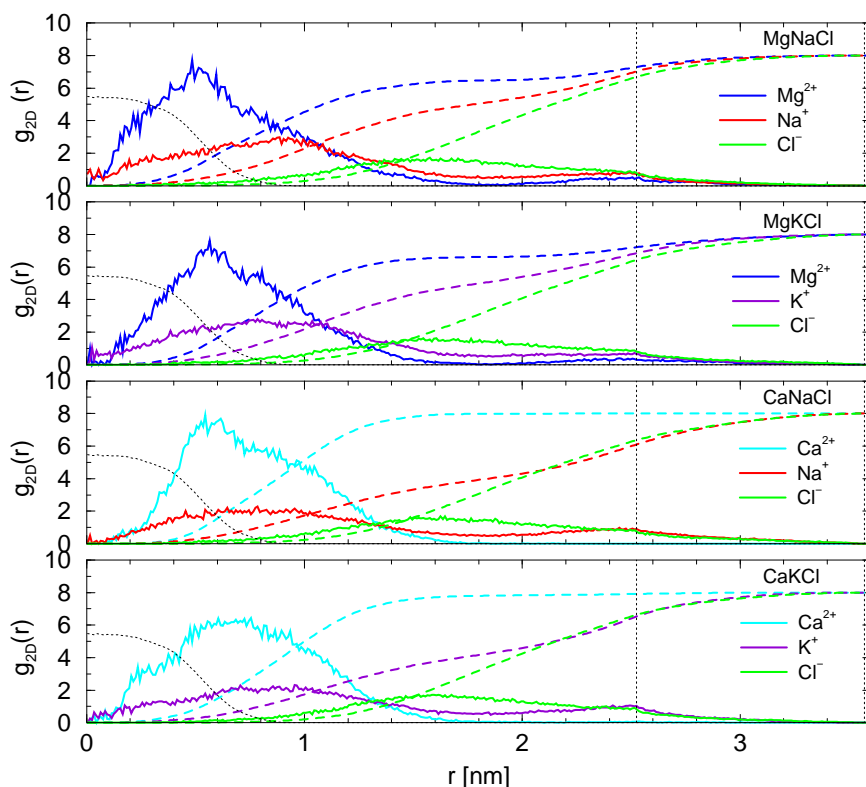


Figure 5.4: Two-dimensional radial distribution function $g_{2D}(r)$ for the coordinates of the counter-ions in the xy -plane relative to the helix axis for simulations MgNaCl, MgKCl, CaNaCl and CaKCl (Table 5.2). The functions are averaged over all ions of a given type in the system and calculated over the entire simulation time (10 ns) using a bin size of 0.01 nm. The cumulative integral of $L^{-2}2\pi r g_{2D}(r)$ (amplified by a factor 8 for readability) is also displayed, as well as the corresponding distribution of all chain atoms (black dotted line; average over all simulations; scaled by a factor 5 for readability).

and the preferential exclusion of the monovalent cations is noticeably stronger for this divalent cation.

In summary, the density and tightness of the counter-ion atmosphere (in terms of an ion of a given type) appears to be correlated with the ion charge and size, resulting the approximate ranking $\text{Ca}^{2+} > \text{Mg}^{2+} > \text{K}^+ \sim \text{Na}^+$.

Illustrative counter-ion distributions corresponding to simulations Na, NaCl, Ca, CaCl, NaKCl, MgCaCl, MgNaCl and CaNaCl are displayed graphically in Figure 5.5.

5.4.3 Ion binding to the chain

The statistics concerning the binding of the different cations to the carboxylate groups of the pGulU chain are reported in Table 5.5 for the 14 simulations.

The following general considerations apply to all systems considered: (i) the occurrence of ion-binding events (B), *i.e.* the average fraction of the simulation time where a carboxylate group binds a cation either tightly (TB; contact ion pair) or loosely (LB; solvent-separated ion pair), is significant (8.5-46.3 %); (ii) for the systems with one single cation type or two isovalent cation

| System | ion | zero [nm] (-) | low [nm] (-) | peak [nm] (-) | high [nm] (-) |
|--------|------------------|------------------|-----------------|------------------|------------------|
| Na | Na ⁺ | 0.00 (2.05) | - | 0.58 (6.12) | 1.73 (0.25) |
| K | K ⁺ | 0.00 (2.23) | - | 0.48 (6.16) | 1.75 (0.25) |
| NaCl | Na ⁺ | 0.00 (0.91) | - | 0.64 (3.31) | 2.83 (0.25) |
| | Cl ⁻ | 0.00 (0.00) | 0.60 (0.25) | 1.56 (1.41) | 3.00 (0.25) |
| KCl | K ⁺ | 0.00 (0.66) | - | 0.70 (3.43) | 2.85 (0.25) |
| | Cl ⁻ | 0.00 (0.00) | 0.72 (0.25) | 1.58 (1.52) | 3.04 (0.25) |
| NaKCl | Na ⁺ | 0.00 (0.83) | - | 0.56 (3.91) | 2.77 (0.25) |
| | K ⁺ | 0.00 (0.80) | - | 0.59 (3.45) | 2.87 (0.25) |
| | Cl ⁻ | 0.00 (0.00) | 0.76 (0.25) | 1.72 (1.52) | 2.96 (0.25) |
| Mg | Mg ²⁺ | 0.00 (0.30) | 0.07 (0.25) | 0.56 (7.97) | 1.40 (0.25) |
| Ca | Ca ²⁺ | 0.00 (0.00) | 0.10 (0.25) | 0.59 (8.20) | 1.47 (0.25) |
| MgCl | Mg ²⁺ | 0.00 (0.08) | 0.04 (0.25) | 0.52 (4.56) | 1.71 (0.25) |
| | Cl ⁻ | 0.00 (0.00) | 0.59 (0.25) | 1.40 (1.62) | 0.59 (0.25) |
| CaCl | Ca ²⁺ | 0.00 (0.07) | 0.12 (0.25) | 0.52 (4.73) | 1.66 (0.25) |
| | Cl ⁻ | 0.00 (0.00) | 0.68 (0.25) | 1.50 (1.76) | 2.75 (0.25) |
| MgCaCl | Mg ²⁺ | 0.00 (0.15) | 0.03 (0.25) | 0.81 (3.98) | 2.73 (0.25) |
| | Ca ²⁺ | 0.00 (0.00) | 0.12 (0.25) | 0.57 (6.09) | 1.64 (0.25) |
| | Cl ⁻ | 0.00 (0.20) | 0.17 (0.25) | 1.44 (1.51) | 2.87 (0.25) |
| MgNaCl | Mg ²⁺ | 0.00 (0.30) | - | 0.51 (7.12) | 2.73 (0.25) |
| | Na ⁺ | 0.00 (0.80) | - | 0.86 (2.89) | 2.77 (0.25) |
| | Cl ⁻ | 0.00 (0.00) | 0.70 (0.25) | 1.58 (1.61) | 2.92 (0.25) |
| MgKCl | Mg ²⁺ | 0.00 (0.50) | 0.10 (0.25) | 0.56 (7.14) | 1.53 (0.25) |
| | K ⁺ | 0.00 (0.65) | - | 0.77 (2.72) | 2.87 (0.25) |
| | Cl ⁻ | 0.00 (0.00) | 0.73 (0.25) | 1.54 (1.56) | 2.98 (0.25) |
| CaNaCl | Ca ²⁺ | 0.00 (0.00) | 0.12 (0.25) | 0.60 (7.50) | 1.60 (0.25) |
| | Na ⁺ | 0.00 (0.25) | 0.12 (0.25) | 0.80 (2.10) | 3.10 (0.25) |
| | Cl ⁻ | 0.00 (0.00) | 0.74 (0.25) | 1.52 (1.57) | 3.06 (0.25) |
| CaKCl | Ca ²⁺ | 0.00 (0.15) | 0.04 (0.25) | 0.70 (6.25) | 1.60 (0.25) |
| | K ⁺ | 0.00 (0.20) | 0.01 (0.25) | 0.86 (2.18) | 2.96 (0.25) |
| | Cl ⁻ | 0.00 (0.00) | 0.79 (0.25) | 1.60 (1.70) | 2.90 (0.25) |

Table 5.4: Main characteristics of the two-dimensional radial distribution functions $g_{2D}(r)$ displayed in Figures 5.2-5.4. The reported entries for the 14 simulations and involved ion types are couples of distances (r) and function ($g_{2D}(r)$) values (between parentheses) corresponding to: $r = 0$ nm (zero), first occurrence of $g_{2D}(r) \geq 0.25$ along the increasing portion of the curve (low), peak position (peak), first occurrence of $g_{2D}(r) \leq 0.25$ along the decreasing portion of the curve (high). Note that the curves were smoothed by five-point averaging (effective bin size 0.05 nm) prior to the determination of these characteristic features. See Table 5.2 for the simulation codes.

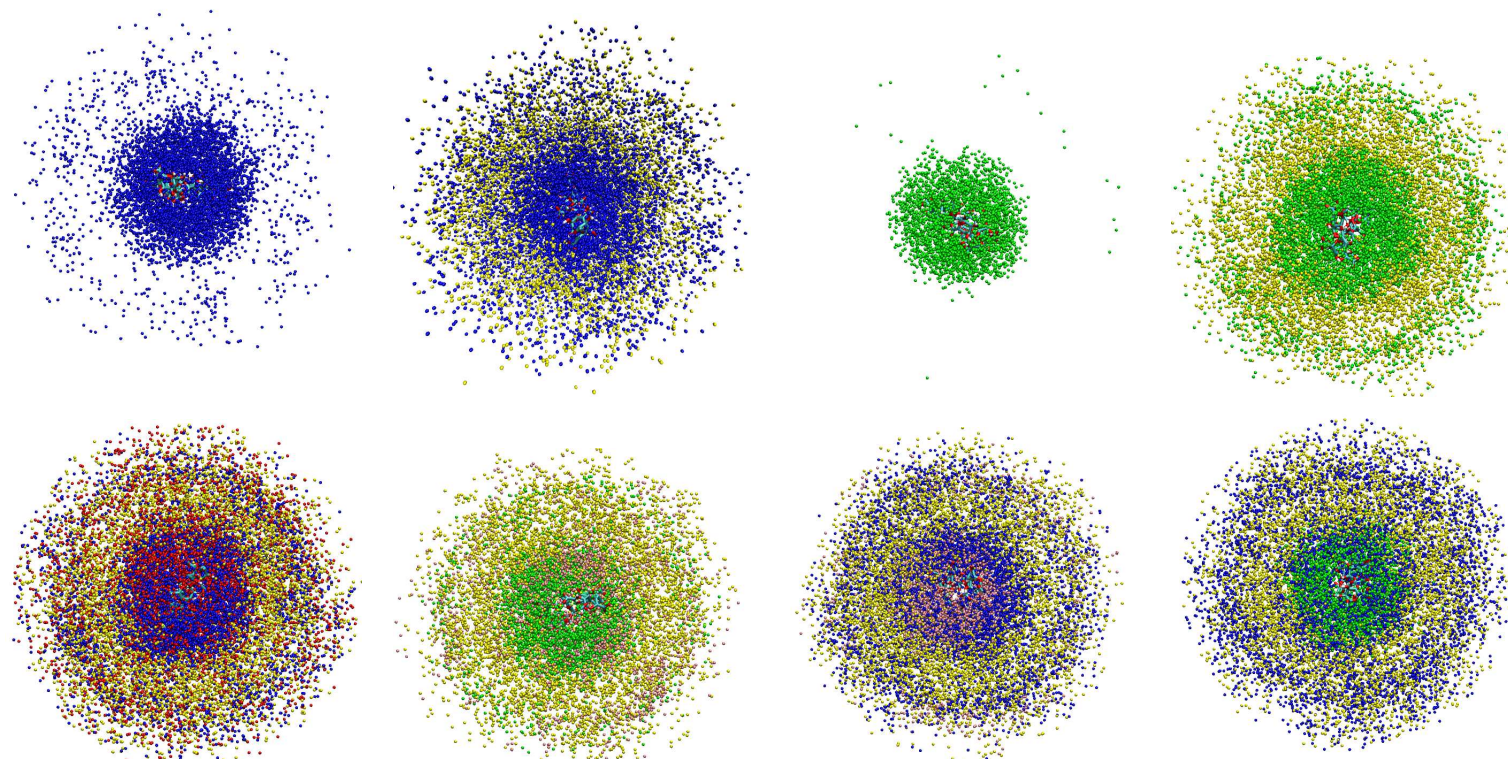


Figure 5.5: Positions of the counter-ions along the chain (top views), as sampled along the simulations (Table 5.2) Na, NaCl, Ca, CaCl (upper row) and NaKCl, MgCaCl, MgNaCl, CaNaCl (bottom row). The ion coordinates correspond to successive configurations superimposed onto the initial one (rotational and translational fit based on the C₂, C₃, C₅ and O₅ atoms) based on the entire (10 ns) simulations. The latter reference configuration is also displayed. Atom coloring: Na⁺ (blue), K⁺ (red), Mg²⁺ (pink), Ca²⁺ (green) and Cl⁻ (yellow). The sampling period (between displayed ion coordinates) is 20 ps.

types (simulations Na, K, NaCl, KCl, NaKCl, Mg, Ca, MgCl, CaCl and MgCaCl), the binding events involving a monovalent cation are more frequent (27.4-46.3%) than those involving a divalent cation (23.1-27.4%); (iii) for the systems involving cation types of different valences (MgNaCl, MgKCl, CaNaCl and CaKCl), binding events involving a monovalent cation may be more frequent (Mg^{2+}) or less frequent (Ca^{2+}) than those involving a divalent cation; (iv) the occurrence of tight-binding events (TB) is low (0.0-4.0 %) and these events are noticeably less frequent for divalent (0.0-0.3 %) compared to monovalent (0.3-4.0 %) ions; (v) the occurrences of double binding events (DB; ion bridging), *i.e.* the average fraction the simulation time where a carboxylate group binds a cation that is simultaneously bound to another carboxylate group (in both cases either tightly or loosely) is also low (0.6-11.9%), and these events are noticeably less frequent for divalent (0.6-3.9%) compared to monovalent (1.4-11.9%) ions; (vi) the dominant contribution to DB events involves the simultaneous binding of an ion to two second-neighbour residues (DB_s) rather than two nearest-neighbour (DB_n) residues; (vii) the inclusion of Cl^- ions does not significantly affect the statistics of binding events (compared to the corresponding simulations without Cl^- ions). Note that triple binding events were sometimes seen (their occurrence never exceeding 0.2 %), while chloride binding events were never observed.

Qualitatively, these observations can be summarized as follows (in the light of the corresponding general observations made in Section 5.4.2). The electrostatic attraction between the carboxylate groups and the cations is strong enough to promote the formation of a dense counter-ion atmosphere around the chain, along with a significant extent of non-specific solvent-separated counter-ion binding. However, this interaction is insufficient to induce the tight-binding (contact ion pair) of the cations at specific sites, which would involve the partial desolvation of these ions. Furthermore, as discussed in previous work⁵ (see Figure 11 therein; Chapter 4 of the present thesis and Figure 4.11), due to the geometry of the chain (in the 2_1 -helical conformation), the coordination of an ion by two carboxylate groups implies either a significant local distortion of the chain (DB_n event) or relatively long carboxylate-ion distances (DB_s event), so that such double-binding events are scarce (and only involve solvent-separated cation binding by the two carboxylate groups involved). These findings are in line with recent implicit-solvent modeling studies, suggesting that counter-ion binding to polyuronate single chains is (predominantly) a (non-specific) counter-ion condensation process^{34,113-115}. The somewhat higher propensity of monovalent cations to be involved in LB, TB or DB events in the absence of divalent cations can be explained by their weaker solvation. In the presence of divalent ions, the inversion of this trend observed for Ca^{2+} probably results from the stronger preferential exclusion of the monovalent cations in this case (Section 5.4.2).

Besides these general considerations, the detailed comparison of the different systems reveals the following differences between corresponding simulations involving a different monovalent (Na^+ or K^+) or divalent (Mg^{2+} or Ca^{2+}) cation. In the absence of divalent cations, the systems involving Na^+ or K^+ (simulations Na, K, NaCl, KCl and NaKCl) reveal a systematically higher occurrence of binding events of all types (B, LB, TB, DB_n , DB_s) for K^+ compared to Na^+ (*e.g.* increase by about 35-50% for B). Note that this systematic trend is not clearly correlated with a difference in the corresponding two-dimensional radial distribution functions (Section 5.4.2; with the possible exception of simulations Na vs K). In contrast, the systems involving Mg^{2+} or Ca^{2+} separately in the absence of monovalent cations (simulations Mg, Ca, MgCl and CaCl) do not reveal a striking difference in terms of ion-binding properties. No differences were observed either in the corresponding radial distribution functions. However, when the two types of divalent cations

| system | I | B | LB | TB | DB | DB _n | DB _s |
|--------|------------------|------|------|-----|------|-----------------|-----------------|
| | | [%] | [%] | [%] | [%] | [%] | [%] |
| Na | Na ⁺ | 27.4 | 26.8 | 0.6 | 3.9 | 0.5 | 3.4 |
| K | K ⁺ | 41.7 | 38.2 | 3.5 | 11.9 | 2.4 | 9.6 |
| NaCl | Na ⁺ | 33.2 | 32.7 | 0.5 | 6.3 | 0.4 | 5.9 |
| KCl | K ⁺ | 46.3 | 42.4 | 4.0 | 10.9 | 2.8 | 8.1 |
| NaKCl | Na ⁺ | 18.2 | 17.9 | 0.3 | 3.4 | 0.2 | 3.1 |
| | K ⁺ | 24.5 | 22.8 | 1.7 | 7.0 | 1.3 | 5.7 |
| Mg | Mg ²⁺ | 27.4 | 27.4 | 0.0 | 2.1 | 0.2 | 1.9 |
| Ca | Ca ²⁺ | 25.8 | 25.8 | 0.0 | 3.9 | 0.4 | 3.5 |
| MgCl | Mg ²⁺ | 23.1 | 23.1 | 0.0 | 1.1 | 0.2 | 1.0 |
| CaCl | Ca ²⁺ | 24.8 | 24.8 | 0.0 | 3.6 | 0.2 | 3.4 |
| CaMgCl | Mg ²⁺ | 8.5 | 8.4 | 0.1 | 0.6 | 0.0 | 0.6 |
| | Ca ²⁺ | 16.8 | 16.6 | 0.2 | 2.6 | 0.2 | 2.4 |
| MgNaCl | Mg ²⁺ | 13.1 | 13.1 | 0.0 | 2.7 | 0.3 | 2.4 |
| | Na ⁺ | 15.1 | 14.6 | 0.4 | 3.0 | 0.6 | 2.5 |
| MgKCl | Mg ²⁺ | 15.0 | 15.0 | 0.0 | 1.2 | 0.0 | 1.2 |
| | K ⁺ | 19.6 | 18.3 | 1.3 | 4.5 | 1.3 | 3.3 |
| CaNaCl | Ca ²⁺ | 19.4 | 19.1 | 0.3 | 2.1 | 0.4 | 1.7 |
| | Na ⁺ | 10.8 | 10.4 | 0.3 | 1.4 | 0.2 | 1.1 |
| CaKCl | Ca ²⁺ | 18.0 | 18.0 | 0.0 | 2.6 | 0.1 | 2.4 |
| | K ⁺ | 15.4 | 14.3 | 1.1 | 4.1 | 1.0 | 3.1 |

Table 5.5: Occurrence of cation-binding events of different types for all simulations (Table 5.2). The occurrences are averaged over the 12 carboxylate groups and over the entire simulation time (10 ns). The indicated properties are B (binding), solvent-separated (loose binding; LB) ion pair, contact (tight binding; TB) ion pair, double-binding (DB; carboxylate group bound to an ion that is simultaneously bound to another carboxylate group), nearest-neighbour double-binding (DB_n; DB event involving two nearest-neighbour residues), second-neighbour double-binding (DB_s; DB event involving two second-neighbour residues). The notation "cat" represents any cation in the system irrespective of its identity.

are included simultaneously (simulation MgCaCl), a higher occurrence of binding events of all types is observed for Ca²⁺ compared to Mg²⁺ (e.g. occurrence higher by about a factor two for B). In this case, the difference is clearly reflected in the corresponding radial distribution functions, where the Ca²⁺ distribution was found to be tighter than the Mg²⁺ distribution. Finally, in the simulations involving simultaneously ions with different valences (MgCaCl, MgKCl, CaNaCl and CaKCl), a systematic difference is also seen in terms of ion-binding between both the monovalent and the divalent cations. First, irrespective of the monovalent cation considered (Na⁺ or K⁺), the occurrences of binding events of all types are higher for Ca²⁺ compared to Mg²⁺. Second,

irrespective of the divalent cation considered, the corresponding occurrences are higher for K^+ compared to Na^+ . Third, the binding of Mg^{2+} is enhanced by the presence of K^+ compared to Na^+ , while the opposite is true for Ca^{2+} . Fourth, the binding of both Na^+ and K^+ are enhanced by the presence of Mg^{2+} compared to Ca^{2+} . Only the last trend could have been anticipated based on the observation of the corresponding radial distribution functions. In contrast, the first trend is somewhat unexpected considering that the close-range (0-3 nm) distribution of Mg^{2+} is higher than that of Ca^{2+} .

In summary, there appears to be no entirely systematic correlation between the trends observed in the distribution of the ions around the chain (Section 5.4.2) and the occurrence of ion-binding events (present section). Unless preferential affinity effects are involved (simulations involving ions of different valences simultaneously), the occurrence of binding events appears to be correlated with the ion charge and size, resulting in the approximate ranking $K^+ > Na^+ > Ca^{2+} > Mg^{2+}$. This ranking is also the one qualitatively expected in terms of increasing hydration strength of the cation. For example, the Na^+ cation is small, well hydrated, and usually classified as “kosmotrope”¹⁵⁰ (meaning that Na^+ tends to promote water structuring¹⁵¹ around itself). In contrast, potassium is larger, less well hydrated, and usually classified as “chaotrope” (meaning that water-water interactions dominate over ion-water interactions¹⁵⁰).

5.4.4 Chain and ion diffusion

The chain rotational (D_{Θ}), longitudinal (D_z) and transverse (D_{xy}) diffusion constants as well as the ion three-dimensional ($D_{i,r}$), longitudinal ($D_{i,z}$) and transverse ($D_{i,xy}$) diffusion constants (the latter averaged over all ions of a given type) are reported in Table 5.6 for the 14 simulations.

Considering the chain, the rotational and longitudinal diffusion constants are found to be the highest in simulations Na, K, MgCl, CaCl and CaMgCl (the corresponding lateral diffusion constants do not evidence clear systematic trends). A more rapid chain diffusion suggests that these systems present either weaker cohesive forces between the chain and its counter-ion atmosphere or a weaker extent of hydration of the chain-ion system (leading to a smaller “effective” radius of the chain in solution). However, the observation of faster diffusion for these five specific systems does not obviously correlate with the trends observed previously in the two-dimensional radial distribution functions (Section 5.4.2) or ion-binding events (Section 5.4.3).

Considering the ions, the three-dimensional diffusion coefficients typically (but not entirely systematically) decrease in the order $Cl^- > K^+ > Na^+ > Ca^{2+} > Mg^{2+}$, which is also the trend observed experimentally for the free ions in solution (2.03, 1.96, 1.33, 0.79 and 0.71 $nm^2 ns^{-1}$, respectively, as calculated from limiting molar conductivities of single ions¹⁵² at 298.15 K; see Eq.6.7.23 in the quoted reference), a faster diffusion correlating with a weaker hydration strength. The main exception to this sequence is an inversion between Ca^{2+} and Mg^{2+} for the systems involving the two ions simultaneously or involving ions of different valences (simulations CaMgCl, MgNaCl, MgKCl, CaNaCl and CaKCl). This inversion can be related to the stronger interaction of Ca^{2+} with the chain in this case (Sections 5.4.2 and 5.4.3). Note also another inversion between Na^+ and K^+ in the simulations involving the corresponding monovalent ion along with Cl^- anions (simulations NaCl and KCl). Since the comparison also reveals a significantly decreased mobility of the Cl^- ions in the presence of K^+ (compared to Na^+), this inversion might result from a stronger interaction of K^+ with Cl^- (compared to Na^+ with Cl^-). Note finally that the longitudinal ionic diffusion coefficients tend to be higher than the transverse ones, although this

trend is by no means systematic.

| system | chain/ ion | D_{Θ} [deg ² ns ⁻¹] | $D_{i,r}$ [nm ² ns ⁻¹] | $D_z(D_{i,z})$ [nm ² ns ⁻¹] | $D_{xy}(D_{i,xy})$ [nm ² ns ⁻¹] | system | chain/ ion | D_{Θ} [deg ² ns ⁻¹] | $D_{i,r}$ [nm ² ns ⁻¹] | $D_z(D_{i,z})$ [nm ² ns ⁻¹] | $D_{xy}(D_{i,xy})$ [nm ² ns ⁻¹] | |
|--------|------------------|--|--|---|---|------------------|------------------|--|--|---|---|-------------|
| Na | Chain | 8.93 | | 1.22 | 0.08 | CaCl | Chain | 5.42 | | 0.36 | 0.06 | |
| | Na ⁺ | | 1.54 (0.08) | 1.63 (0.16) | 1.49 (0.34) | | Ca ²⁺ | | | 2.47 (0.40) | 0.84 (0.10) | 3.29 (2.42) |
| K | Chain | 12.27 | | 0.89 | 0.12 | MgCaCl | Cl ⁻ | | 2.44 (0.08) | 2.39 (0.23) | 2.47 (0.41) | |
| | K ⁺ | | 1.65 (0.04) | 2.32 (0.17) | 1.32 (0.19) | | Chain | 3.60 | | | 1.06 | 0.19 |
| NaCl | Chain | 1.53 | | 0.22 | 0.24 | Mg ²⁺ | Mg ²⁺ | | 2.01 (0.12) | 1.89 (0.27) | 2.06 (0.55) | |
| | Na ⁺ | | 2.69 (0.08) | 3.42 (0.34) | 2.33 (0.27) | | Ca ²⁺ | | 1.08 (0.11) | 0.67 (0.06) | 1.28 (0.68) | |
| | Cl ⁻ | | 3.98 (0.14) | 3.49 (0.30) | 4.22 (0.81) | | Cl ⁻ | | 3.72 (0.15) | 3.83 (0.53) | 3.67 (0.52) | |
| KCl | Chain | 0.53 | | 0.22 | 0.09 | MgNaCl | Chain | 0.61 | | 0.36 | 0.19 | |
| | K ⁺ | | 1.98 (0.04) | 2.44 (0.15) | 1.75 (0.15) | | Mg ²⁺ | | 0.51 (0.02) | 0.81 (0.05) | 0.36 (0.12) | |
| | Cl ⁻ | | 2.55 (0.05) | 2.47 (0.26) | 2.59 (0.29) | | Na ⁺ | | 1.36 (0.03) | 1.46 (0.13) | 1.31 (0.21) | |
| NaKCl | Chain | 0.42 | | 0.12 | 0.26 | MgKCl | Cl ⁻ | | 2.30 (0.05) | 2.68 (0.20) | 2.11 (0.30) | |
| | Na ⁺ | | 1.47 (0.03) | 1.49 (0.10) | 1.46 (0.16) | | Chain | 0.42 | | | 0.34 | 0.11 |
| | K ⁺ | | 2.08 (0.07) | 1.89 (0.33) | 2.18 (0.32) | | Mg ²⁺ | | 0.71 (0.03) | 0.90 (0.19) | 0.61 (0.19) | |
| | Cl ⁻ | | 2.39 (0.07) | 1.91 (0.18) | 2.63 (0.39) | | K ⁺ | | 2.49 (0.09) | 3.17 (0.38) | 2.15 (0.24) | |
| Mg | Chain | 0.59 | | 0.17 | 0.09 | CaNaCl | Cl ⁻ | | 2.50 (0.06) | 2.81 (0.23) | 2.34 (0.32) | |
| | Mg ²⁺ | | 0.22 (0.01) | 0.36 (0.04) | 0.15 (0.02) | | Chain | 1.16 | | | 0.13 | 0.21 |
| Ca | Chain | 1.23 | | 0.12 | 0.14 | Ca ²⁺ | Ca ²⁺ | | 0.49 (0.03) | 0.66 (0.08) | 0.40 (0.17) | |
| | Ca ²⁺ | | 0.79 (0.07) | 0.33 (0.04) | 1.02 (0.38) | | Na ⁺ | | 1.82 (0.04) | 1.53 (0.15) | 1.97 (0.34) | |
| MgCl | Chain | 2.28 | | 0.32 | 0.04 | CaKCl | Cl ⁻ | | 2.41 (0.07) | 2.93 (0.24) | 2.14 (0.29) | |
| | Mg ²⁺ | | 1.36 (0.06) | 0.89 (0.10) | 1.59 (0.32) | | Chain | 1.45 | | | 0.17 | 0.11 |
| | Cl ⁻ | | 3.42 (0.16) | 2.33 (0.19) | 3.97 (0.96) | | Ca ²⁺ | | 0.65 (0.03) | 0.86 (0.08) | 0.55 (0.19) | |
| | | | | | | K ⁺ | | 2.61 (0.07) | 1.80 (0.19) | 3.02 (0.39) | | |
| | | | | | | Cl ⁻ | | 2.19 (0.04) | 2.15 (0.18) | 2.22 (0.21) | | |

Table 5.6: Diffusion constants evaluated for different sets of degrees of freedom of the chain and of the counter-ions based on the 14 simulations (Table 5.2). The diffusion coefficients $D_{\mathbf{q}}$ are calculated from a linear least-square-fit of the corresponding mean-square displacements $Q_{\mathbf{q}}(t)$ over the time period from 0 to 3 ns, themselves calculated based on the entire simulation time (10 ns) considering all possible time origins. The chain diffusion is described in terms of the rotational (D_{Θ}), longitudinal (D_z) and transverse (D_{xy}) diffusion constants. The ionic diffusion is described in terms the ionic three-dimensional ($D_{i,r}$), longitudinal ($D_{i,z}$) and transverse ($D_{i,xy}$) diffusion constants, averaged over all the counter-ions of a given type (the corresponding standard deviation divided by the square root of the number of ions is indicated between parentheses as an error estimate). See Section 5.3.2 for the definition of $Q_{\mathbf{q}}(t)$ and $D_{\mathbf{q}}$.

5.4.5 Effective chain charge

The effective chain charge $Q(r)$ is displayed as function of the cylinder radius r in Figure 5.5 for all simulations. For the ease of discussion, the corresponding values at $r = 1.48$ and 2.33 nm (approximate locations of local maxima and minima, respectively, in the curves for systems involving more than one ion type) are reported in Table 5.7 (along with the contributions of the different ions to the corresponding values). For the systems involving the chain neutralized by a single ion type in the absence of chloride ions (simulations Na, K, Mg and Ca), the effective charge increases monotonically from -12 (bare chain) to $0 e$ (neutralized chain) over the entire distance range. The increase is steeper for the two simulations involving divalent cations compared to monovalent cations. In the former case, the quasi-complete ($>99\%$) neutralization of the chain charge is reached at about 1.2 nm, while it is only reached at about 2.5 nm in the later case.

This difference is expected on the basis of Debye-Hückel theory, the Debye screening lengths κ^{-1} corresponding to the two types of systems being 0.76 and 1.08 nm, respectively (Table 5.2). For the systems involving chloride ions, the effective charge curves are no longer monotonic and present a maximum at about 1.48 nm, followed by a minimum at about 2.33 nm, before reaching zero at longer distances. The effective charge at the maximum of the curve is positive (except for simulation NaCl), and particularly large for the systems involving divalent cations (simulations MgCl, CaCl and MgCaCl; effective charge of about 2.9-3.5 e) and, to a lesser extent, for the systems involving cations of different valences (simulations MgNaCl, MgKCl, CaNaCl and CaKCl; effective charge of about 1.3-2.3 e). Furthermore, considering these systems, the simulations involving Ca^{2+} ions systematically present a higher positive charge at the maximum compared to the simulations involving Mg^{2+} (a similar difference is also observed in simulation MgCaCl, where Ca^{2+} provides the dominant positive contribution to this effective charge).

| system | ion | Q(1.48 nm) [e] | Q(2.33 nm) [e] | system | ion | Q(1.48 nm) [e] | Q(2.33 nm) [e] |
|--------|------------------|-------------------|-------------------|---------------|------------------|-------------------|-------------------|
| Na | Na^+ | 10.13 | 11.25 | CaCl | Ca^{2+} | 18.56 | 21.16 |
| | Q | -1.87 | -0.75 | | Cl^- | -3.04 | -9.41 |
| K | K^+ | 10.08 | 11.24 | Q | 3.53 | -0.25 | |
| | Q | -1.92 | -0.76 | CaMgCl | Mg^{2+} | 7.88 | 10.07 |
| NaCl | Na^+ | 13.75 | 18.94 | | Ca^{2+} | 10.38 | 11.21 |
| | Cl^- | -2.05 | -8.10 | | Cl^- | -3.35 | -9.17 |
| KCl | Q | -0.31 | -1.17 | Q | 2.91 | 0.12 | |
| | K^+ | 14.43 | 19.02 | MgNaCl | Mg^{2+} | 9.31 | 10.31 |
| | Cl^- | -2.09 | -8.18 | | Na^+ | 6.51 | 9.47 |
| Q | 0.34 | -1.16 | Cl^- | -2.52 | -8.76 | | |
| NaKCl | Na^+ | 7.70 | 9.75 | Q | 1.30 | -0.98 | |
| | K^+ | 6.87 | 9.52 | MgKCl | Mg^{2+} | 9.69 | 10.36 |
| | Cl^- | -2.05 | -8.37 | | K^+ | 6.38 | 9.37 |
| Mg | Q | 0.52 | -1.10 | Cl^- | -2.29 | -8.38 | |
| | Mg^{2+} | 11.93 | 12.00 | Q | 1.78 | -0.64 | |
| | Q | -0.07 | 0.00 | CaNaCl | Ca^{2+} | 11.65 | 12.00 |
| Ca | Ca^{2+} | 11.94 | 12.00 | | Na^+ | 4.91 | 7.90 |
| | Q | -0.06 | 0.00 | | Cl^- | -2.23 | -8.33 |
| MgCl | Mg^{2+} | 18.24 | 20.91 | Q | 2.33 | -0.43 | |
| | Cl^- | -3.18 | -8.88 | CaKCl | Ca^{2+} | 11.33 | 11.81 |
| | Q | 3.06 | 0.03 | | K^+ | 5.18 | 8.40 |
| | | | | Cl^- | -2.26 | -8.65 | |
| | | | | Q | 2.26 | -0.44 | |

Table 5.7: Values of the effective chain charge $Q(r)$ at $r = 1.48$ or $r = 2.33$ nm based on the curves of Figure 5.5. For the different simulations, the last entry represents $Q(r)$, while the other entries report the contributions of each ion type to this value. See Table 5.2 for the simulations codes.

In summary, it appears that the charge screening afforded by the counter-ion atmosphere leads to a complete charge screening of the chain within about 1.0-2.5 nm of the chain axis (1.0-1.2 nm excluding simulations Na and K). Furthermore, in the presence of negative charges other than the sole chain charges (here, Cl^- anions), a charge reversal is actually observed close to 1.5 nm. This charge reversal is particularly strong in the presence of divalent (and especially Ca^{2+}) cations, where the magnitude of the (positive) effective chain charge at the maximum may be as high as about a quarter of that of the (negative) bare chain charge.

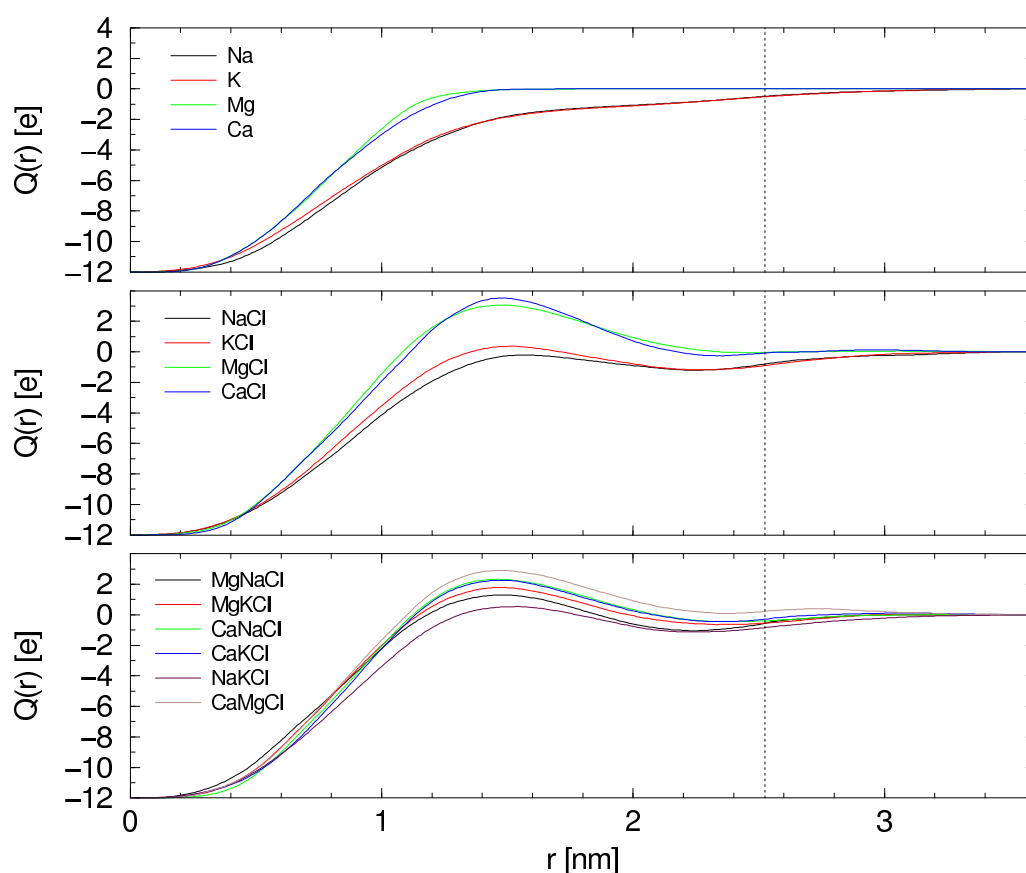


Figure 5.6: Effective chain charge $Q(r)$ as a function of the cylinder radius r for the 14 simulations (Table 5.2). This quantity represents the net-charge of a cylindrical volume of radius r centered on the helix axis (and of length corresponding to the simulated dodecameric unit), and was calculated based on the two-dimensional radial distribution function $g_{2D}(r)$ of Figures 5.1-5.4.

5.5 Conclusions

The goal of the present study was to investigate and compare the interaction of a set of monovalent (Na^+ , K^+) and divalent (Mg^{2+} , Ca^{2+}) cations with single-chain pGulU (2_1 -helical conformation) using explicit-solvent MD simulations, considering (overall neutral) systems involving one or two cation types, possibly along with chloride (Cl^-) anions. The main observations can be summarized as follows:

1. The chain conformation, flexibility and intramolecular hydrogen-bonding is insensitive to the ionic content of the surrounding solution.
2. The binding of counter-ions to the chain appears to be largely non-specific, irrespective of the ionic content of the surrounding solution. The counter-ion atmosphere is almost entirely confined within a cylinder of high ionic density around the chain axis, with no well defined binding sites. The carboxylate groups frequently (25-45%) bind cations, but nearly always remain in their second solvation shells. Events of tight binding, as well as of multiple coordination, are scarce and of short durations⁵.

3. The density (spanned distance range) and tightness (closeness to the helix axis) of the distributions of specific cations within the counter-ion atmosphere follow the approximate order $\text{Ca}^{2+} > \text{Mg}^{2+} > \text{K}^+ \sim \text{Na}^+$. In particular, when both monovalent and divalent ions are present in solution, a clear preferential affinity for the latter ones is observed. Preferential affinity for Ca^{2+} over Mg^{2+} is also apparent, while K^+ and Na^+ behave very similarly.
4. The frequency of ion binding events for a specific cation (predominantly solvent-separated binding to a single carboxylate group) is not systematically related to the density and tightness of the corresponding ionic distribution within the counter-ion atmosphere, and follows the approximate order $\text{K}^+ > \text{Na}^+ > \text{Ca}^{2+} > \text{Mg}^{2+}$.
5. The chain diffusion properties are not immediately correlated with the above properties (a faster diffusion is observed for simulations involving neutralizing monovalent ions alone, or divalent ions in the presence of chloride ions). The ionic diffusion properties essentially follow the trends expected from bulk diffusion ($\text{Cl}^- > \text{K}^+ > \text{Na}^+ > \text{Ca}^{2+} > \text{Mg}^{2+}$; correlated with the hydration strength), and reveal a (non-systematic) trend towards faster longitudinal (as opposed to transverse) diffusion.
6. The analysis of the effective chain charge reveals a nearly complete screening of the chain charge within 1.0-1.2 nm (excluding the simulations with a neutralizing amount of monovalent cations alone) and a charge reversal close to 1.5 nm (excluding the simulations with a neutralizing amount of cations alone; the effective chain charge at this distance being up to one quarter of the magnitude of the bare chain charge).

The second observation is in line with recent theoretical models suggesting that counter-ion binding to polyuronate single chains is (predominantly) a (non-specific) counter-ion condensation process^{5,34,113-115}. The third observation is also quantitatively consistent with experimental results on the relative affinities of alkali and alkali-earth cations for alginates (Section 5.2), except for the possible absence of difference between Na^+ and K^+ . Note, however, that the experimental data^{82,83} suggest that such differences could be (at least in part) due to egg-box like interchain interactions. Such interactions, if present, are obviously not accounted for in the present (single-chain) simulations. The preferential affinity for divalent over monovalent cations can easily be understood on the basis of simple arguments from Debye-Hückel or counter-ion condensation theory^{34,113-115}. The difference between Mg^{2+} and Ca^{2+} is probably connected with more complex effects related to the hydration strength of the bare cation (as well as, possibly, cation-cation and cation-chloride interactions). The fourth observation suggests that (solvent-separated) cation binding to single-chain alginates is predominantly correlated with the hydration strength of the cation. Finally, the sixth observation may have implications in the context of the kinetics of gel formation. Charge reversal effects^{34,113-115} may significantly accelerate the speed of dimer formation by inducing a (temporary) Coulombic attraction between chain segments, *i.e.* permitting the approach of a less screened segment (negative effective charge) with another more efficiently screened segment (positive apparent charge).

Taken together, the results of the present simulations provide a detailed (atomistic) picture of the interaction of ions with single-chain (pGulU segments in) alginates that is not directly accessible to experimental methods and is certainly more accurate than the description provided by simplified analytical theories. However, it should be stressed that a complete characterization

of the nature of alginate-cation interactions will require an analogous investigation in the context of associated chains (*e.g.* junction zones in gels). Such a study will be scope of a forthcoming study.

Acknowledgments

The author is grateful to Maria Reif for sharing her knowledge concerning ionic solvation properties and ion-solvent parameters.

5.6 References

- [1] Ovodov, Y.S. *Pure Appl. Chem.* **42** 351-369 (1974)
- [2] Rees, D.A. & Welsh, E.J. *Angew. Chem. Int. Ed.* **16** 214-224 (1977)
- [3] Morris, E.R. *Brit. Polym. J.* **18** 14-21 (1986)
- [4] Rees, D.A. *Pure Appl. Chem.* **53** 1-14 (1981)
- [5] Peric, L., Pereira, C.S., Perez, S. & Hünenberger, P.H. *Molecular Simulation* **34** 421-446 (2008)
- [6] Ramesh, H.P. & Tharanathan, R.N. *Crit. Rev. Biotech.* **23** 149-173 (2003)
- [7] Smidsrød, O. & Draget, K.I. *Carbohydr. Eur.* **14** 6-13 (1996)
- [8] Thakur, B.R., Singh, R.K. & Handa, A.K. *Crit. Rev. Food Sci. Nutrition* **37** 47-73 (1997)
- [9] Ridley, B.L., O'Neill, M.A. Mohnen, *Phytochemistry* **57** 929-967 (2001)
- [10] Willats, W.G.T., McCartney, L., Mackie W., Knox, *Plant Mol. Biol.* **47** 9-27 (2001)
- [11] Fang, Y., Al-Assaf, S., Philips, G.O., Nishinari, K., Funami, T. & Williams, P.A. *Struct. Chem.* **20** 317-324 (2009)
- [12] Voragen, A.G.J., Coenen, G.-J., Verhoef, R.P. & Schols, H.A. *Struct. Chem.* **20** 263-275 (2009)
- [13] Clarke, A.H & Ross-Murphy, S.B. *Adv. Polym. Sci.* **83** 56-192 (1987)
- [14] Rinaudo, M. *Macromol. Biosci.* **6** 590-610 (2006)
- [15] Rinaudo, M. *Struct. Chem.* **20** 277-289 (2009)
- [16] Williams, P.A. *Struct. Chem.* **20** 299-308 (2009)
- [17] Pérez, S., Mazeau, K. & Herve du Penhoat, C. *Plant Physiol. Biochem.* **38** 37-55 (2000)
- [18] Onsoyen, E. *Carbohydr. Eur.* **14** 26-31 (1996)

- [19] Heyraud, A., Dantas, L., Courtois, J., Courtois, B., Helbert, W. & Chanzy, H. *Carbohydr. Res.* **258** 275-279 (1994)
- [20] Braccini, I., Heyraud, A. & Pérez, S. *Biopolymers* **45** 165-175 (1998)
- [21] Dantas, L., Heyraud, A., Courtois, B., Courtois, J. & Milas, M. *Carbohydr. Polym.* **24** 185-191 (1994)
- [22] Walter, R.H. Polysaccharide association structures in food. Marcel Dekker Inc., New York, U.S.A. (1998)
- [23] Shilpa, A., Agrawal, S.S. & Ray, A.R. *J. Macromol. Sci. C.* **43** 187-221 (2003)
- [24] Skjåk-Braek, G. & Espevik, T. *Carbohydr. Eur.* **14** 19-25 (1996)
- [25] Draget, K.I., Skjåk-Braek, G. & Smidsrød, O. *Int. J. Biol. Macromol.* **21** 47-55 (1997)
- [26] Anal, A.K., *Trends Food Sci. Tech.* **18** 240-251 (2007)
- [27] Davis, T.A., Volesky, B. & Mucci, A. *Water Res.* **37** 4311-4330 (2003)
- [28] Rinaudo, M. *Polym. Int.* **57** 397-430 (2008)
- [29] Rees, D.A. & Samuel, J.W.B. *J. Chem. Soc. C* **22** 2295-2298 (1967)
- [30] Haug, A., Myklestad, S., Larsen, B. & Smidsrød, O. *Acta Chem. Scand.* **21** 768-778 (1967)
- [31] Fujihara, M. & Nagumo, T. *Carbohydr. Res.* **224** 343-347 (1992)
- [32] Dentini, M., Rinaldi, G., Barbeta, A., Risica, D., Anselmi, C. & Skjåk-Braek, G. *Carbohydr. Polymers* **67** 465-473 (2007)
- [33] Williams, M.A.K., Marshall, A., Haverkamp, R.G. & Draget, K.I. *Food Hydrocolloids* **22** 18-23 (2008)
- [34] Siew, C.K. & Williams, P.A. *Biomacromolecules* **6** 963-969 (2005)
- [35] Emmerichs, N., Wingender, J., Flemming, H.-C. & Mayer, C. *Int. J. Biol. Macromol.* **34** 73-79 (2004)
- [36] Donati, I., Holtan, S., Morch, Y.A., Borgogna, M., Dentini, M., *Biomacromolecules* **6** 1031-1040 (2005)
- [37] Morris, E.R., Rees, D.A., Thom, D. & Boyd, J. *Carbohydr. Res.* **66** 145-154 (1978)
- [38] Thibault, J.-F. & Rinaudo, M. *Biopolymers* **25** 455-468 (1986)
- [39] Tibbits, C.W., MacDougall, A.J. & Ring, S.G. *Carbohydr. Res.* **310** 101-107 (1998)
- [40] Grant, G.T., Morris, E.R., Rees, D.A., Smith, P.J.C. & Thom, D. *FEBS Lett.* **32** 195-198 (1973)

- [41] Jarvis, M.C. & Apperley, D.C. *Carbohydr. Res.* **275** 131-145 (1995)
- [42] Stokke, B.T., Draget, K.I., Smisrod, O., Yuguchi, Y., Urakawa, H. & Kajiwara, K. *Macromolecules* **33** 1853-1863 (2000)
- [43] Braccini, I. & Pérez, S. *Biomacromolecules* **2** 1089-1096 (2001)
- [44] Rees, D.A. *Carbohydr. Polym.* **2** 254-263 (1982)
- [45] Morris, D.F.C. *Electrochim. Acta* **27** 1481-1486 (1982)
- [46] Fang, Y., Al-Assaf, S., Philips, G.O., Nishinari, K., Funami, T., Williams, P.A. & Li, L. *J. Phys. Chem. B* **111** 2456-2462 (2007)
- [47] Atkins, E.D.T., Mackie, W. & Smolko, E.E. *Nature* **225** 626-628 (1970)
- [48] Atkins, E.D.T., Mackie, W., Parker, K.D. & Smolko, E.E. *Polymer letters* **9** 311-316 (1971)
- [49] Atkins, E.D.T., Nieduszynski, H.H., Mackie, W., Parker, K.D. & Smolko, E.E. *Biopolymers* **12** 1879-1887 (1973)
- [50] Sikorski, P., Mo, F., Skjåk-Braek, G. & Stokke, B.T. *Biomacromolecules* **8** 2098-2103 (2007)
- [51] Arnott, S., Bian, W., Chandrasekaran, R. & Mains, B.R. *Fibre Diffr. Rev.* **9** 44-51 (2000)
- [52] Mackie, W. *Biochem. J.* **125** 89P-89P (1971)
- [53] Dheu-Andries, M.L & Pérez, S. *Carbohydr. Res.* **124** 324-332 (1983)
- [54] Yuguchi, Y., Urakawa, H., Kajiwara, K., Draget, K.I. & Stokke, B.T. *J. Mol. Struct.* **554** 21-34 (2000)
- [55] Li, L., Fang, Y., Vreeker, R. & Appelquist, I. *Biomacromolecules* **8** 464-468 (2007)
- [56] Powell, D.A., Morris, E.R., Gidley, M.J. & Rees, D.A. *J. Mol. Biol.* **155** 517-531 (1982)
- [57] Morris, E.R., Powell, D.A., Gidley, M.J. & Rees, D.A. *J. Mol. Biol.* **155** 507-516 (1982)
- [58] Mimmo, T., Marzadori, C., Montecchio, D. & Gessa, C. *Carbohydr. Res.* **340** 2510-2519 (2005)
- [59] Lamelas, C., Benedetti, M., Wilkinson, K.J. & Slaveykova, V.I. *Chemosphere* **65** 1362-1370 (2006)
- [60] Davis, T.A., Kalis, E.J.J., Pinheiro, J.P., Town, R.M. & van Leeuwen, H.P. *Environ. Sci. Technol.* **42** 7242-7247 (2008)
- [61] Kalis, E.J.J., Davis, T.A.A., Town, R.M. & van Leeuwen, H.P. *Environ. Sci. Technol.* **43** 1091-1096 (2009)

- [62] Hung, Chin-Chang, Tang D., Warneken, K.W., Santschi, *Marine Chem.* **73** 305-318 (2001)
- [63] Davis, T.A., Llanes, F., Volesky, B. & Mucci, A. *Environ. Sci. Technol.* **37** 261-267 (2004)
- [64] Li, F.T., Yang, H., Zhao, Y. & Xu, R. *Chinese Chem. Lett.* **18** 325-328 (2007)
- [65] Lagoa, R. & Rodrigues, J.R. *Appl. Biochem. Biotechnol.* **143** 115-128 (2007)
- [66] Balaria, A. & Schiewer, S. *Separation and Purification Technology* **63** 577-581 (2008)
- [67] Maureira, A. & Rivas, B.L. *Eur. Polymer J.* **45** 573-581 (2009)
- [68] Haug, A. & Smidsrød, O. *Nature* **215** 757-757 (1967)
- [69] Abrahamse, A.J., Lipreau, C. & Heijman, S.G.J. *J. Membrane Sci.* **323** 153-158 (2008)
- [70] Jin, X., Huang, X. & Hoek, E.M.V. *Environ. Sci. Technol.* **43** 3580-3587 (2009)
- [71] McConaughty, S.D., Kirkland, S.E., Treat, N.J., Stroud, P.A. & McCormick, C.L. *Biomacromolecules* **9** 3277-3287 (2008)
- [72] Thibault, J.F. & Rinaudo, M. *Biopolymers.* **24** 2131-2143 (1985)
- [73] Haug, A. *Acta Chem. Scand.* **15** 1794-1795 (1961)
- [74] Haug, A. & Smidsrød, O. *Acta Chem. Scand.* **19** 341-351 (1965)
- [75] Haug, A. & Smidsrød, O. *Acta Chem. Scand.* **19** 1221-1226 (1965)
- [76] Smidsrød, O. & Haug, A. *Acta Chem. Scand.* **22**. **1989-1997** - (1968)
- [77] Haug, A. & Smidsrød, O. *Acta Chem. Scand.* **24** 843-854 (1970)
- [78] Smidsrød, O. & Haug, A. *Acta Chem. Scand.* **26** 2063-2074 (1972)
- [79] Seely, G.R. & Hart, R.L. *Macromolecules* **7** 706-710 (1974)
- [80] Bartmess, J.E. *J. Phys. Chem.* **98** 6420-6424 (1994)
- [81] Mørch, Y.A., Donati, I., Strand, B.L. & Skjåk-Braek, G. *Biomacromolecules* **7** 1471-1480 (2006)
- [82] Seale, R., Morris, E.R. & Rees, A. *Carbohydr. Res.* **110** 101-112 (1982)
- [83] Draget, K.I., Steinsvag, K., Onsoyen, E. & Smidsrød, O. *Carbohydr. Polym.* **35** 1-6 (1998)
- [84] Smidsrød, O. & Haug, A. *Acta Chem. Scand.* **19** 329-340 (1965)
- [85] Kohn, R., Furda, I., Haug, A. & Smidsrød, O. *Acta Chem. Scand.* **22** 3098-3102 (1968)
- [86] Lunde, G., Smidsrød, O. & Haug, A. *Acta Chem. Scand.* **26** 3421-3426 (1972)
- [87] Cescutti, P. & Rizzo, R. *J. Agric. Food Chem.* **49** 3262-3267 (2001)

- [88] Lamelas, C., Avaltroni, F., Benedetti, M., Wilkinson, K.J. & Slaveykova, V.I. *Biomacromolecules* **6** 2756-2764 (2005)
- [89] Rodrigues, J.R. & Lagoa, R. *J. Carbohydr. Chem.* **25** 219-232 (2006)
- [90] Li, Y., Shi, T., An, L., Lee, J., Wang, X. & Huang, Q. *J. Phys. Chem. B* **111** 12081-12087 (2007)
- [91] Mimmo, T., Cavani, L., Reggiani, R., Marzadori, C. & Gessa, C. *Biol. Fertil. Soils* **44** 521-526 (2008)
- [92] Dunstone, J.R. *Biochem. J.* **85** 336- (1962)
- [93] Woodward, C. & Davidson, E.A. *Biochemistry.* **60** 201-205 (1968)
- [94] Rochas, C. & Rinaudo, M. *Biopolymers.* **19** 1675-1687 (1980)
- [95] Ueda, K., Sato, S., Ochiai, H., Imamura, A. & Yamaoka, K. *Chem. Lett.* **4** 763-766 (1994)
- [96] Cantoni, C., Zennaro, F., Bertocchi, C., Mariotti, P. & Rizzo, R. *Biopolymers* **45** 157-163 (1998)
- [97] Rudd, T.R. et al. *Glycobiology* **17** 983-993 (2007)
- [98] Furth, G., Knierim, R., Buss, V. & Mayer, C. *J. Biol. Macromol.* **42** 33-40 (2008)
- [99] Rotureau, E. & Leeuwen, H.P. *J. Phys. Chem. A* **112** 7177-7184 (2008)
- [100] Braccini, I., Grasso, R.P. & Pérez, S. *Carbohydr. Res.* **317** 119-130 (1999)
- [101] Anderson, C.F. & Record Jr., M.T. *Annu. Rev. Biophys. Biophys. Chem.* **19** 423-465 (1990)
- [102] Hill, T.L. *Archives Biochem. Biophys.* **57** 229-239 (1955)
- [103] Wagner, K., Keyes, E., Kephart, T.W. & Edwards, G. *Biophys. J.* **73** 21-30 (1997)
- [104] Rau, D.C. & Charney, E. *Biophys. Chem.* **14** 1-9 (1981)
- [105] Le Bret, M. & Zimm, B. *Biopolymers* **23** 287-312 (1984)
- [106] Stigter, D. *Biophys. J.* **69** 380-388 (1995)
- [107] Grochowski, P. & Trylska, J. *Biopolymers* **89** 93-113 (2007)
- [108] Manning, G.S. *J. Chem. Phys.* **51** 924-933 (1969)
- [109] Manning, G.S. *J. Chem. Phys.* **51** 934-938 (1969)
- [110] Manning, G.S. *J. Chem. Phys.* **51** 3249-3252 (1969)
- [111] Manning, G.S. *Biophys. Chem.* **7** 95-102 (1977)
- [112] Manning, G.S. *Macromolecules* **40** 8071-8081 (2007)

- [113] Donati, I., Cesaro, A. & Paoletti, S. *Biomacromolecules* **7** 281-287 (2006)
- [114] Donati, I., Benegas, J.C., Cesaro, A. & Paoletti, S. *Biomacromolecules* **7** 1587-1596 (2006)
- [115] Donati, I., Benegas, J.C. & Paoletti, S. *Biomacromolecules* **7** 3439-3447 (2006)
- [116] van Gunsteren, W.F. & Berendsen, H.J.C. *Angew. Chem. Int. Ed.* **29** 992-1023 (1990)
- [117] Manunza, B. Deiana, S., Pintore, M., Gessa, *Glycoconjugate J.* **15** 297-300 (1998)
- [118] Eriksson, M., Lindhorst, T.K. & Hartke, B. *J. Chem. Phys.* **128** 105105/1-105105/10 (2008)
- [119] van Gunsteren, W.F., Billeter, S.R., Eising, A.A., Hünenberger, P.H., Krüger, P., Mark, A.E., Scott, W.R.P. & Tironi, I.G. *Biomolecular simulation: The GROMOS96 manual and user guide.* Verlag der Fachvereine, Zürich, Switzerland. (1996)
- [120] Scott, W.R.P., Hünenberger, P.H., Tironi, I.G., Mark, A.E., Billeter, S.R., Fennen, J., Torda, A.E., Huber, T., Krüger, P. & van Gunsteren, W.F. *J. Phys. Chem. A* **103** 3596-3607 (1999)
- [121] Schuler, L.D. & van Gunsteren, W.F. *Mol. Simul.* **25** 301-319 (2000)
- [122] Schuler, L.D., Daura, X. & van Gunsteren, W.F. *J. Comput. Chem.* **22** 1205-1218 (2001)
- [123] Chandrasekhar, I., Kastenholz, M.A., Lins, R.D., Oostenbrink, C., Schüler, L.D., Tieleman, D.P. & van Gunsteren, W.F. *Eur. Biophys. J.* **32** 67-77 (2003)
- [124] Soares, T.A., Hünenberger, P.H., Kastenholz, M.A., Kräutler, V., Lenz, T., Lins, R.D., Oostenbrink, C. & van Gunsteren, W.F. *J. Comput. Chem.* **26** 725-737 (2005)
- [125] Börjesson, U. & Hünenberger, P.H. *J. Phys. Chem. B* **108** 13551-13559 (2004)
- [126] Lins, R.D. & Hünenberger, P.H. *J. Comput. Chem.* **26** 1400-1412 (2005)
- [127] Pereira, C.S., Kony, D., Baron, R., Müller, M., van Gunsteren, W.F. & Hünenberger, P.H. *Biophys. J.* **90** 4337-4344 (2006)
- [128] Kräutler, V., Müller, M. & Hünenberger, P.H. *Carbohydr. Res.* **342** 2097-2124 (2007)
- [129] Hansen, H.S. & Hünenberger, P.H. *J. Comput. Chem.*, in press (2009)
- [130] Peric, L., Hansen, H.S., Baron, R. & Hünenberger, P.H. *J. Phys. Chem.* **B** to be submitted- (2009)
- [131] Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F. & Hermans, J. Interaction models for water in relation to protein hydration. In: *Intermolecular Forces.* Pullman, B., Ed. Reidel, Dordrecht, The Netherlands. pp 331-342 (1981)
- [132] Hockney, R.W. *Methods Comput. Phys.* **9** 136-211 (1970)

- [133] Ryckaert, J.-P., Ciccotti, G. & Berendsen, H.J.C. *J. Comput. Phys.* **23** 327-341 (1977)
- [134] Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., Di Nola, A. & Haak, J.R. *J. Chem. Phys.* **81** 3684-3690 (1984)
- [135] Barker, J.A. & Watts, R.O. *Mol. Phys.* **26** 789-792 (1973)
- [136] Tironi, I.G., Sperb, R., Smith, P.E. & van Gunsteren, W.F. *J. Chem. Phys.* **102** 5451-5459 (1995)
- [137] Heinz, T.N., van Gunsteren, W.F. & Hünenberger, P.H. *J. Chem. Phys.* **115** 1125-1136 (2001)
- [138] Manunza, B., Deiana, S. & Gessa, C. *J. Mol. Struct.* **368** 27-29 (1996)
- [139] Manunza, B., Delana, S., Pintore, B. & Gessa, C. *J. Mol. Struct.* **419** 169-172 (1997)
- [140] Manunza, B., Deiana, S., Pintore, M. & Gessa, C. *Carbohydr. Res.* **300** 85-88 (1997)
- [141] Perry, T.D., Cygan, R.T. & Mitchell, R. *Geochim. Cosmochim. Acta* **70** 3508-3532 (2006)
- [142] Larwood, V.L., Howlin, B.J. & Webb, G.A. *J. Mol. Model.* **2** 175-182 (1996)
- [143] Noto, R., Martorana, V., Bulone, D., *Biomacromolecules* **6** 2555-2562 (2005)
- [144] Kony, D., Damm, W., Stoll, S., van Gunsteren, W.F. & Hünenberger, P.H. *Biophys. J.* **93** 442-455 (2007)
- [145] Plashchina, I.G., Semenova, M.G., E.E. Braudo, Tolstoguzov, *Carbohydr. Polym.* **5** 159-179 (1985)
- [146] Cros, S., Garnier, C., Axelos, M.A.V., Imberty, A. & Pérez, S. *Biopolymers* **39** 339-352 (1996)
- [147] Lide, D.R. CRC Handbook of Chemistry and Physics. Edition 88 (Internet version). CRC Press, Boca Raton, Florida (2008)
- [148] Glättli, A., Daura, X. & van Gunsteren, W.F. *J. Chem. Phys.* **116** 9811-9828 (2002)
- [149] Insight II Molecular Modeling Package. Molecular Simulations Inc. (2000)
- [150] Collins, K.D. *Biophys. J.* **72** 65-76 (1997)
- [151] Abraham, M.H., Liszi, J. & Papp, E. *J. Chem. Soc. Faraday Trans. 1* **78** 197-211 (1982)
- [152] Fawcett, W.R. Liquids, solutions and interfaces. Oxford Univ. Press, Oxford, U.K. (2004)

Chapter 6

Outlook

Molecular dynamics (MD) simulation methods have proven to be a powerful tool complementary to experiment in investigating the properties of (bio)molecular systems at atomic resolution. They provide insight into the thermodynamics and dynamics of molecular systems and contribute to the understanding of the relationship between structure, flexibility, dynamics and function. Steadily increasing computing power along with the iterative refinement of force-fields allows for continuously improved accuracy. However, at a given time point, limitations in the reachable system sizes and sampling times always remain major obstacles.

Chapter 2 describes and applies the LEUS method as a powerful technique to enhance the sampling along conformationally relevant degrees of freedom, and permit the evaluation of important properties such as free energy barriers and estimated transition timescales. Two important directions of future research concerning the application of the LEUS scheme to carbohydrates are: (i) the consideration of other (non-glucose) disaccharides (to further investigate the generality of the topological symmetries observed for the glucose-based disaccharides; (ii) the application of this scheme to oligosaccharides (with a two-dimensional biasing potential applied separately to each linkage; to further investigate the possibility of a sampling enhancement in more complex carbohydrate systems). Chapter 3 features simulations of membrane systems which demand constant improvement in terms of both simulation timescales and force-field accuracy. Two important directions of future research concerning the investigation of membrane-carbohydrate interactions (in the context of bioprotection) are : (i) the design of a lipid force-field that can properly describe the liquid-crystal to gel phase transition temperature of the membrane (so as to investigate the direct effect of sugars on the phase transition temperature at various extents of hydration; as most relevant in the context of anhydrobiosis) (ii) the formulation of quantitative definitions for the various bioprotection mechanisms (*i.e.* amenable to an unambiguous assessment based on experimental or theoretical data). Chapters 4 and 5 are concerned with the properties of single-chain polyuronates. Obviously, the most important direction of future research is here the extension of the work to associated chains, in view of the characterization of junction zones in gels.

As a final note, it is also well recognized that the accuracy of MD simulations is also affected by the approximate treatment of electrostatic interactions. In this respect, the simulations presented here would certainly benefit from the inclusion of explicit electronic polarization and the use of more accurate interaction methods (*e.g.* lattice-sum methods). However, all these current shortcomings of MD simulation techniques are probably secondary if one keeps in mind that the principal strength of these methods resides in the correct quantitative insight they provide into molecular processes rather than in the exact quantitative prediction of system properties.

Curriculum Vitæ

Personal Data

| | |
|----------------|-----------------------|
| Name | Lovorka Perić-Hassler |
| Date of birth | April 6, 1979 |
| Place of birth | Rijeka, Croatia |
| Nationality | Croatian |

Education

| | |
|-------------------|---|
| 09/1985 – 07/1993 | Osnovna škola Nikola Tesla, Rijeka |
| 09/1993 – 07/1997 | Prva hrvatska sušačka gimnazija, Rijeka |
| 09/1997 – 07/1998 | Kemijsko tehnonloški fakultet, Zagreb |
| 09/1998 – 11/2004 | Prirodoslovno-matematički fakultet (PMF), Zagreb |
| 12/2003 – 11/2004 | Diploma thesis supervised by Prof. Z. Maksić, Institut Ruđer Bošković, Zagreb |
| 08/2005 – 08/2009 | Ph.D. studies with Prof. W.F. van Gunsteren supervised by Dr. P.H. Hünenberger, ETH Zurich, Zürich |