DISS. ETH NO. 23356

Solvent properties and their influence on carbohydrate conformation investigated using molecular dynamics simulations.

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

ALICE LONARDI

M. Sc. University of Trento

born on July 24, 1983

citizen of Italy

accepted on the recommendation of

Prof. Dr. Philippe Henry Hünenberger, examiner

Prof. Dr. Wilfred F. van Gunsteren, co-examiner

Prof. Dr. Xavier Daura Ribera, co-examiner

2016

Contents

Summary

Scientific modelling is a fundamental step towards understanding natural phenomena, as a model is required to define theoretical representations of reality and to interpret experimental observations.

In the analysis of bio-chemical processes, models and computer simulations are tools of increasing importance. The general purpose of the model as well as the type of system and phenomenon that are under study define the kind of model and simulation performed. A model is determined by the choice of the degree of freedom considered, the type of interactions included, the simulation method employed and the boundary conditions applied to the system. An overview of these four fundamental aspects is given in Chapter 1.

Bio-molecular simulations typically aim at reproducing the behaviour of bio-chemical systems under experimental conditions. This generally implies the consideration of a solvent. Due to its key role and to the complexity of the solvation interactions, the parametrization of solvent models represents a crucial and a highly non-trivial task.

In Chapter 2, the sensitivity of the macroscopic properties of one of the most fundamental solvents, water, to specific molecular parameters of the model will be analyzed, focusing in particular on characteristics related to polarity, namely the dielectric permittivity and hydrogen-bonding propensity.

Intramolecular hydrogen-bonding is commonly regarded as a major determinant of the conformation of (bio-)molecules. However, in different environments, solvent-exposed hydrogen-bonds vary in significance, possibly representing only marginal conformational driving as well as steering forces in aqueous media. The analysis of the effect of the solvent on intramolecular hydrogen-bonding in a solute is the subject of Chapter 3 and Chapter 4. Carbohydrates are chosen for this investigation because of their huge conformational diversity, a fundamental aspect in determining their great variety and flexibility. Moreover, the very large differences between properties in vacuum and in water are a clear sign of the importance of the solvent for these molecules.

Chapter 3 focuses on the rotameric preferences of the hydroxymethyl group in two of the most common monosaccharides, glucose and galactose. Chapter 4 evaluates the importance of the solvent effect on the conformational properties of the glycosidic linkage in the context of disaccharides. Both studies also attempt to disentangle the specific effects

of the solvent dielectric permittivity and hydrogen-bonding capacity by using artificial solvents that allow for a separate modulation of these two properties of the model.

Finally, concluding remarks and possible further developments are provided in Chapter 5.

Riassunto

A livello scientifico, la creazione di modelli è un passo fondamentale per capire i fenomeni naturali. Un modello e la sua verifica sono spesso usati per definire una rappresentazione teorica della realtà e per interpretare osservazioni sperimentali.

Modelli e simulazioni a computer stanno diventando strumenti di sempre maggior importanza nello studio di meccanismi biochimici. Lo scopo generale di un modello e il tipo di sistema e di fenomeno che si vogliono analizzare definiscono il tipo di modello e di simulazione che vanno effettuati. Il miglior modello di un sistema è determinato dalla scelta dei gradi di libertà considerati, dal tipo di interazioni incluse nel modello, dal metodo di simulazione usato e dal tipo di condizioni al contorno imposte al sistema stesso.

Una panoramica di questi quattro aspetti fondamentali viene fornita nel capitolo 1.

Le simulazioni biomolecolari tendono di frequente a riprodurre il comportamento di sistemi biochimici in condizioni sperimentali. In generale, questo implica l'uso di un solvente.

A causa dell'importanza del ruolo del solvente e della varietà e complessità delle interazioni incluse nella solvatazione, la parametrizzazione dei modelli di solvente rappresenta allo stesso tempo un compito cruciale e altamente non triviale.

Nel capitolo 2 viene analizzata la sensibilità delle proprietà macroscopiche di uno dei solventi principali, l'acqua, a specifici parametri molecolari del modello, concentrandosi in modo particolare sulle caratteristiche connesse alla polarità, cioè la permittività dielettrica, la tendenza al legame idrogeno.

I legami idrogeno intramolecolari sono comunemente considerati determinanti per la conformazione delle (bio)molecole. Tuttavia, in ambienti diversi, i legami idrogeno esposti al solvente possono avere un'importanza diversa, arrivando a rappresentare solo una forza marginale verso specifiche conformazioni, così pure come una linea guida non determinante in acqua. Il soggetto dei capitoli 3 e 4 è l'analisi dell'effetto del solvente sui legami idrogeno intramolecolari nel soluto. I carboidrati sono stati scelti per questo studio a causa della loro enorme diversità conformazionale, aspetto fondamentale nella determinazione della loro grande varietà e flessibilitá. Inoltre, le discrepanze tra le proprietà in vuoto e in acqua sono un chiaro segno dell'importanza del solvente per queste molecole.

Il capitolo 3 pone l'attenzione sulle preferenze rotameriche del gruppo idrossimetile di

due dei più comuni monosaccaridi, glucosio e galattosio. Il capitolo 4, invece, valuta l'importanza dell'effetto del solvente sulle proprietà conformazionali del legame glicosidico nel contesto dei disaccaridi. Entrambi i capitoli cercano di distinguere specifici aspetti della polarità del solvente, facendo uso di solventi artificiali che permettono una modulazione separata della permettività dielettrica e della capacità della formazione di legami idrogeno del modello.

Infine, conclusioni e possibili sviluppi futuri riguardanti questi argomenti saranno menzionati nel capitolo 5.

Publications

This thesis has led to the following publications:

Chapter 3

A. Lonardi, P. Oborský and P. H. Hünenberger.
"Solvent-modulated influence of intramolecular hydrogen-bonding on the conformational properties of the hydroxymethyl group in glucose and galactose: A molecular dynamics simulation study."
In preparation.

Related Publications

W. Plazinski, A. Lonardi and P. H. Hünenberger.
"Revision of the GROMOS 56A6 CARBO Force Field for Hexopyranose-Based Carbohydrates: Improving the Description of Ring-Conformational Equilibria in Oligo- and Polysaccharide Chains"
Journal of Computational Chemistry 37, 354â365, 2016.

Chapter 1

Introduction

1.1 Modelling

Modelling means generating a representation of reality and providing the set of rules that govern this modelled reality. This process always implies simplification and abstraction, and discernment between fundamental aspects and unimportant factors. Whenever a theoretical explanation of a phenomenon is given, it implicitly refers to a specific model. Different models may account for the same phenomenon while remaining essentially different, as they may be based on different assumptions. Different models generally also have different domains of application, as the purpose of the representation influences the aspects of interest, and thus the assumptions that are made during the abstraction.

Beyond a certain level of complexity, models are implemented in simulations, nowadays most of the time computer simulations. Computer simulations are fundamental to validate a model, because the simulation results can be directly compared with experimental observations in order to check the robustness of the assumptions made, and the approximations involved.

When a model has been thoroughly validated, it can yield more than a simple interpretation and understanding of reality. Simulations can be used in complement to experiment, driving and directing them, or they can even sometimes substitute to experiments, in situations where experiment is too dangerous, too expensive, or simply impossible.

Models and simulations are used in all areas of science, from physics to social sciences. All the aspects of modelling and simulation presented in this introduction concern molecular systems, as this thesis focuses on the molecular simulation of chemical and biochemical systems.

1.1.1   Molecular modelling

August Wilhelm von Hofmann is credited with the first real-life model of a molecule (balls connected by sticks) around 1860. The atomic model, with distinction between the nucleus and the electrons, is due to Bohr[1] in 1913. The first quantum calculations date back to the 1930s, while the first simulation of a liquid[2] required the use of a computer, and was not possible until 1953.

Since then, the steady increase of computer power has allowed for the refinement of more and more accurate models, especially at the quantum-chemical level, and the consideration of ever larger systems, mainly at the classical level. Additionally, along with this increase in the computing power, the main purpose and focus of molecular simulation has changed, evolving from the static analysis of single atoms or dimers to more complex chemically and biologically relevant systems and processes.

To define a molecular model, four major choices have to be made, in particular regarding:

- The elementary particles (degrees of freedom)[3–8] considered in the model.

- The basic interactions[3–6,9,10] between the selected elementary particles, *i.e.* the Hamiltonian operator or function of the system.

- The method to generate configurations[3–6,9,11–15], defining the statistical and dynamical properties of the configuration ensemble and sequence, respectively.

- The boundary conditions[3–6,16] associated with the system under consideration, such as size, shape and thermodynamic parameters.

All basic decisions regarding a model have to be taken considering the type of system and processes one is interested in. Sections 1.2-1.5 provide a short overview of the possibilities available regarding the treatment of these four aspects, focusing mainly on the models and methods used in the context of this thesis.

1.2 Degrees of freedom

The first decision to be taken when building a model concerns which and how many degrees of freedom of the system are relevant to the phenomena of interest. In other words, the nature of a specific target information determines the type and number of independent variables required to estimate it. In the simulation of chemical or biochemical processes, the level of resolution required to model a chemical reaction between two molecules is, for example, different from the one needed to model the folding of a protein. Different elementary particles and, consequently, different resolutions should be considered. The choice of the degrees of freedom of the model and of the size of the system, together with the type of interactions, intrinsically determine the computational cost of the simulation. It is clear that, on a practical level, any model involves a compromise between the resolution and the feasibility of the simulations that can be carried out with the available computing power.

A rough hierarchy[3–8] of models usually considered for molecular systems can be provided starting from basic quantum-mechanical models, for which the elementary particles are nuclei and electrons, and progressively removing (averaging out) specific degrees of freedom. Already for quantum-chemical simulations, some degrees of freedom are automat-

ically excluded (nucleons), but what is omitted at the nuclear level is more the concerns of nuclear and particle physics than of chemistry. Only relatively small systems can be simulated at quantum-mechanical resolution.

When electrons are also removed, atoms become the elementary particles and larger systems can be considered with affordable computational costs[3–5,17]. The simulation of these systems allows for the study of solvation, as well as of structural and thermodynamic properties of large molecules. The removed degrees of freedom are included in an effective way into the interatomic interaction function, but the electronic resolution is lost. A further approximation can be taken by treating the solvent in an implicit fashion[18–27]. In this case, its mean effect is included into the interaction potential energy for the solute, and one generally adds stochastic and frictional forces to reproduce its effect on the dynamics. Longer timescales become accessible and larger systems can be studied, but an accurate description of the short-range solute-solvent interactions is no longer possible.

When modelling larger molecules involved in complex biochemical processes, the size of the system can be too large to be studied at atomistic resolution. In this case, multiple atoms, polymer residues, or even entire molecules are grouped into single supra-particles, generating so called coarse-grained models[28–39]. With these models, the correct dynamic of the system is lost, as well as all the intra-bead flexibility.

Further removal of degrees of freedom is performed in fluid dynamics (continuum representation), but these simulations are beyond the scope of the present work.

Sometimes, combined levels of resolutions, so-called hybrid models, can be convenient for the description of specific systems. In this approach, some parts of the system, the most crucial ones for the phenomenon studied, are considered at a higher level of resolution, while the rest is approximated at a coarser level. Hybrid models can be used at all levels previously mentioned, *e.g.* active sites of enzymes can be studied at the quantum-

mechanical level in a classical environment (QM/MM)[40], atomic models of proteins can be simulated in coarse-grained solvents or in fine grained/coarse grained solvent mixtures, etc.

1.3   Interactions

The classical particles chosen to define the degrees of freedom of the model are interacting with each other. Each type of interaction has to be represented by an appropriate functional form, which also includes the specification of a number of parameters. The set of functional forms defining the potential energy of a system of particles, together with their parameters, is called a force field.

1.3.1   Approximations

The introduction of the concept of force-field to describe the interactions between particles is only valid in the classical limit. This means that a series of approximations has been introduced. Simplified particles (nuclei and electrons) in a constant environment (time-independent Hamiltonian, *i.e.* isolated system) within the Born-Oppenheimer approximation (nuclei quasi motionless from the point of view of the electrons and electrons relaxing instantaneously from the point of view of the nuclei), are assumed to be in the electronic ground state. Furthermore, the nuclei are assumed to be heavy enough and subject to sufficiently smoothly varying forces that they behave as classical particles (atoms), the motion of which can be described by the classical Newtonian equations of motion. In practice, these approximations impose some restrictions on the domain of application of the model: the detailed motion of the lightest atoms must be irrelevant (*e.g.* the description of protons is quantum-mechanically inaccurate), high-frequency oscillations (*e.g.* the

bond stretching) should be approximated by constraints, and the temperature must be sufficiently high to ensure the appropriateness of the classical limit of statistical mechanics.

Additionally, for most practical applications, the classical potential energy function that determines the motion of the particles must be approximated by a simple analytical and differentiable function. Moreover, the number of parameters required is commonly limited by transferability assumptions (of the force-field parameters for specific atoms in specific environment to the same atoms in similar chemical environments) and by combination rules (for pair parameters based on single-atom parameters).

1.3.2   Types of force-field

Several types of force-field are available. Their differences come mainly from the purpose of the model and the goal of the simulations they are used for. A complete description of all the force-fields available is beyond the scope of this introduction, and only those that are most relevant for the present work will be mentioned.

Condensed-phase (bio)molecular force-fields (*e.g.* OPLS[41,42], AMBER[43–45], CHARMM[46–48] and GROMOS[49,50]) are empirical force-fields, meaning that they rely on a simple functional form owing to the assumptions listed in Section 1.3.1 and that their parameters are derived primarily based on experimental data.

Introducing different approximations, different kinds of force-fields can be developed, that address simulations with different aims. As an example, reactive force fields[51,52] introduce the possibility of modelling chemical reactions, and they are based on different approximations for the classical limit.

1.3.3 Force-field terms

In the approximation frame described above and in most of the empirical force-fields, the potential energy function is represented by a simple sum of analytical force-field terms $V^{(t_\alpha)}$. Each force-field term can be written as a function of one or more internal coordinates q_α of the system (*i.e.* any function of the Cartesian coordinates of the set of atoms involved in the interaction), and of a set of parameters s_α specific to the term. The potential energy function is therefore represented as

$$V(\mathbf{r}) = \sum_{\alpha=1}^{N_{terms}} {}^{(n_\alpha)}V^{(t_\alpha)}(q_{\alpha,1}, \dots; s_{\alpha,1}, \dots), \tag{1.1}$$

where n_α is the order of the term α, that is, the number of particles involved in the interaction.

Different kinds of force-field terms can be introduced, each of them describing either a physical interaction found in nature (physical terms) or an artificial one included to modify the dynamics of the system in unphysical simulations (unphysical terms). In the following sections, a brief description of the most commonly employed force-field terms is given, with a special focus on the physical terms. The unphysical terms depend strongly on the kind of artificial dynamics and on the purpose of the simulation. A general classification can therefore not be given, and a complete description of these terms is beyond the scope of this introduction. Merely, a short list of the most typically used artificial terms is given, while a more detailed presentation of the specific terms relevant in the context of this thesis is provided in Section 1.4.2.

PHYSICAL TERMS

Physical force-field terms describe specific physical interactions between the atoms as found in nature. Two categories of physical terms can be defined: covalent terms and non-bonded terms.

Covalent terms

Covalent interactions occur between atoms within the same molecule that are close neighbors, *i.e.* separated by 1, 2, or 3 bonds. At the quantum-mechanical level, the electron density between the nuclei drives them to adopt specific bond distances, angles and torsional preferences. In a classical description, molecules tend to adopt specific geometries (including stereochemistry) depending on the bonds and angles defined between the atoms, and on the possibility of torsional rotation of certain groups. Covalent or bonded force-field terms represent these types of interactions.

Bonds between atoms can be treated classically either by constraining the bond distance to a fixed reference value (rigid models) or by accounting for their vibrations (flexible models). In the first case, no force-field term is included in the potential energy function for the bonds, and constraint algorithms, such as SHAKE[53] or LINCS[54] must be applied to keep the bond length between the atoms at its reference value. This approach is justified by the consideration that the vibrations associated with bond oscillations are typically not excited at room temperature. For flexible models, a typical potential energy term describing bond stretching between two atoms is the harmonic function

$$^{(2)}V^{(b)}\left(b_n; K_{b_n}, b_{0_n}\right) = \sum_{n=1}^{N_b} \frac{1}{2} K_{b_n} \left(b_n - b_{0_n}\right)^2,$$

(1.2)

where b_{0_n} is the reference distance of the bond n and K_{b_n} is the harmonic force constant.

A computationally cheaper form is a quartic term

$$^{(2)}V^{(b)}\left(b_n; K'_{b_n}, b_{0_n}\right) = \sum_{n=1}^{N_b} \frac{1}{4} K'_{b_n} \left[b_n^2 - b_{0_n}^2\right]^2, \tag{1.3}$$

where K'_{b_n} is the quartic force constant. More complex forms can also be employed, *e.g.* when a more precise description of the vibrational frequencies (Taylor expansion) or when bond dissociation (*e.g.* Morse function) is desired.

Very similar considerations can be made for the force-field terms that describe the bond-angle bending. Angles can be treated either with constraints, with the same justification as for the bonds, or in a flexible way. However, the use of bond-angle constraints in molecules that are not fully rigid is not recommended due to metric-tensor effects [55,56]. Therefore, most of the force-fields consider the potential energy function for the angles explicitly. The form typically used for this force-field term is the harmonic functional form,

$$^{(3)}V^{(\theta)}\left(\theta_n; K_{\theta_n}, \theta_{0_n}\right) = \sum_{n=1}^{N_\theta} \frac{1}{2} K_{\theta_n} \left[\theta_n - \theta_{0_n}\right]^2, \tag{1.4}$$

where θ_{0_n} is the reference bond angle for the angle n and K_{θ_n} the harmonic force constant. A computationally cheaper form is a harmonic form in the cosine of the angle

$$^{(3)}V^{(\theta)}\left(\theta_n; K'_{\theta_n}, \theta_{0_n}\right) = \sum_{n=1}^{N_\theta} \frac{1}{2} K'_{\theta_n} \left[\cos\theta_n - \cos\theta_{0_n}\right]^2, \tag{1.5}$$

where K'_{θ_n} is the quartic force constant. However, the latter is not the best choice for molecules involving linear groups. Other more complex forms can be employed, when higher precision in the description of the vibrational frequencies is desired (Taylor expansion or Urey-Bradley function).

The relative rotation around specific bonds considering four atoms is described by tor-

sional dihedral angles. In contrast to bond stretching and bond-angle bending, the torsional dihedral angles cannot be harmonically confined to one specific value, since the whole rotational range $[0 : 2\pi[$ is generally sampled at room temperature. A cosine series is commonly used for the force-field term representing this interaction

$$^{(4)}V^{(\phi)}\left(\phi_n; K_{\phi_n}, \delta_n, m_n\right) = \sum_{n=1}^{N_\phi} K_{\phi_n}\left[1 + \cos\left(m_n\phi_n - \delta_n\right)\right], \tag{1.6}$$

where m_n is the multiplicity of term, δ_n the phase shift (0 or π) and K_{ϕ_n} is the force constant. Note that terms of different multiplicities are sometimes employed simultaneously for the same dihedral angle. The choices that have to be made with respect to this type of force-field term concern how many terms of different multiplicities are included, how many dihedral angles are considered around a common bond, and how the non-bonded interactions between the third-neighbor atoms that define the end of a torsional dihedral angle have to be handled (excluding the pair from the non-bonded interactions, including it, or including it in a reduced or scaled form).

The specific stereochemistry of a molecule can be enforced by a special force-field term, the improper dihedral angle term. This term describes how difficult it is to distort the specific geometry, *e.g.* planar or tetrahedral, around an atom. Improper dihedral angles are used normally to enforce the chirality or the planarity around a central atom, and they are needed in particular when a specific geometry has to be enforced around a carbon described by a CHR_3 united atom. A harmonic functional form is typically employed, namely

$$^{(4)}V^{(\xi)}\left(\xi_n; K_{\xi_n}, \xi_{0_n}\right) = \sum_{n=1}^{N_\xi} \frac{1}{2}K_{\xi_n}\left[\xi_n - \xi_{0_n}\right]^2. \tag{1.7}$$

where ξ_{0_n} is the reference value for the improper dihedral angle n and K_{ξ_n} is the harmonic

force constant.

Non-bonded terms

Interactions between atoms within different molecules or atoms within the same molecule that are not close covalent neighbors, *i.e.* separated by 3 or more bonds, are described by non-bonded force-field terms. In principle, one should use a series of terms that depend on the positions of an increasing number of atoms

$$V(\mathbf{r}) = \sum_i^N V_1(\mathbf{r}_i) + \sum_{i<j}^N V_2(\mathbf{r}_i, \mathbf{r}_j) + \sum_{i<j<k}^N V_3(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \cdots \tag{1.8}$$

The one-body term is only relevant when an external field is applied. In absence such a field, the potential energy does not depend on the absolute positions of the atoms, but only on their relative positions. The pairwise terms represent the most important non-bonded interaction terms. Because these terms are summed over all atom pairs, for a system of N atoms, there are about $(1/2)N(N-1)$ pairwise terms to be included.

All terms of order higher than two represent so-called many-body interactions. These terms are seldom explicitly included in simulations, since they are computationally very expensive (*e.g.* $(1/6)N(N-1)(N-2)$ terms for the three-body interactions). They can be seen as accounting for the indirect influence of a third, fourth, fifth, ... atom on the pairwise interaction between a specific pair. For the van der Waals interactions, several potential energy functional forms have been suggested to consider many-body interactions, either explicitly (*e.g.* Axilrod-Teller three-body term[57], EAM-like functionals[58], Tight-Binding Second Moment Approximation (TBSMA)[59] functionals) or as effective pairwise interactions (Baker-Fischer-Watts functional[60]). For the electrostatic interactions, the explicit inclusion of electronic polarizability[61–64] formally accounts for an N-body term. The treatment of these potential energy terms remains one of the most challenging theoretical issue in molecular simulations. A complete description of many-body terms is beyond

the scope of this introduction, and the focus of the following paragraphs will be on the effective pairwise terms that are mostly used for simulations of (bio)molecular systems.

The calculation of non-bonded interactions, even if just pairwise, is the most computationally demanding component of a simulation. Because of their non-local nature, they involve (at least feeble) interactions between all particles in the system. Several numerical schemes or approximations have been developed to reduce their computational cost. The most commonly used are cutoff truncation, reaction-field methods[60,65], and the particle mesh Ewald[66,67] or Particle-Particle Particle Mesh[68] (P3M) methods. In the context of this thesis, a twin-range cutoff scheme[69] with a reaction-field correction is normally employed.

Three interaction types are generally considered in biomolecular simulations to account for pairwise non-bonded interactions: electrostatic interactions, van der Waals interactions, and hydrogen-bonding interactions.

The electrostatic functional form is commonly given in the monopole (atomic partial charges) approximation by Coulomb's law

$$^{(2)}V^{(C)}\left(r_{ij}; q_i, q_j, \epsilon_{cs}\right) = \sum_{\substack{not\ excluded \\ pairs\ i,j}} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_{cs}} \frac{1}{r_{ij}}. \tag{1.9}$$

where q_i and q_j are the charges of atoms i and j, ϵ_0 is the permittivity of vacuum, ϵ_{cs} is the relative permittivity of the medium, and r_{ij} is the distance between the two atoms considered.

The choice of the partial charges attributed to the atoms is not trivial. Several different approaches can be used, from electrostatic potential fits based on quantum-mechanical calculations to pure effective parameter-fitting procedures based on experimental thermodynamic data (typically concerning liquids).

Van der Waals interactions account for the fact that also uncharged atoms (*e.g.* rare-

gas atoms) interact, so that their behaviour in the gas phase deviates from the ideal-gas model. Qualitatively, they include two effects:

1. A very-short range repulsion, consequence of the Pauli exclusion principle.

2. An intermediate range attraction, arising from the dispersion or London forces, and consequence of electron correlation.

A wealth of different representations of van der Waals interactions have been suggested[70], all of them trying to account for the two above effects. Many of them rely on the two-parameters reduced form

$$^{(2)}V^{(vdW)}(r_{ij}) = \epsilon(i,j)\eta(\rho_{ij}) \quad \text{with } \rho_{ij} = \frac{r_{ij}}{r^*(i,j)} \tag{1.10}$$

where $\epsilon(i,j)$ and $r^*(i,j)$ are the depth and position, respectively, of the potential energy minimum. The choice of the function $\eta(\rho_{ij})$ determines in particular the steepness of the repulsion component and the magnitude of the dispersion component. These are in principle of fundamental importance for several properties of the simulated systems (*e.g.* pressure, compressibility). However, ultimately, the parameters involved in all these functional forms are effective empirical parameters that can be tuned to reproduce the experimental data in the condensed phase, reducing the importance of a specific choice of functional form.

The most commonly used van der Waals interaction function is the Lennard-Jones func-

tion[71–73]

$$^{(2)}V_{(LJ)}\left(r_{ij};C_{12}(i,j),C_6(i,j)\right) = \sum_{\substack{nonbonded \\ pairs\ i,j}} \epsilon(i,j)\left(\rho^{-12} - 2\rho^{-6}\right)$$

$$= \sum_{\substack{nonbonded \\ pairs\ i,j}} \left(\frac{C_{12}\left(i,j\right)}{r_{ij}^{12}} - \frac{C_6\left(i,j\right)}{r_{ij}^6}\right). \qquad (1.11)$$

It includes a 12th power repulsion component (*ad hoc* but computationally advantageous) and a 6th power dispersion component (correct leading term according to the quantum mechanics of dispersion). Other common choices for these non-bonded force-field terms are the 9-6 function[74], the exp-m functions[75,76], the Double-Morse function[77,78], the n-m buffered function[79].

The Lennard-Jones function depends on two parameters ($\epsilon(i,j)$ and $r^*(i,j)$ or, equivalently, $C_{12}(i,j)$ and $C_6(i,j)$) for each atom pair, determined by the nature of the two atoms involved in the interaction. To reduce the number of parameters involved, they are often written as a function of the corresponding homoatomic parameters, as

$$\epsilon(i,j) = f[\epsilon(i,i),\epsilon(j,j)]$$

$$r^*(i,j) = f[r^*(i,i),r^*(j,j)], \qquad (1.12)$$

where the function f defines a so-called combination rule[80] (*e.g.* geometric, arithmetic, or cubic mean).

Sometimes, the potential energy function also includes an explicit term to model hydrogen-bonding interactions, that depend on the nature of donor and acceptor atoms. Alternatively, hydrogen-bonding interactions can be accounted for by adjusting the combination rules for the van der Waals interaction parameters to account for the H-bonding propensity of the pair considered (*e.g.* in GROMOS[81–84], different parameters are combined for non-

hydrogen-bonding, uncharged-hydrogen-bonding and charged interactions). In this case the H-bonding is determined by a balance between electrostatic and van der Waals interactions.

UNPHYSICAL TERMS

Artificial force-field terms can be included in the potential energy function to modify the dynamics of the system in unphysical simulations. This is commonly done to enforce the agreement between the simulation and experimental results (*e.g.* NOE-derived distance restraints[85,86], [3]J-value restraints[87], structure factor restraints[88,89]), to bias or improve the sampling of the conformational space (*e.g.* local elevation[90], local elevation umbrella sampling[14]), to restrict the sampling to specific areas of the conformational space (*e.g.* position restraints, ball&stick local elevation umbrella sampling[15]), or to define specific pathways between to states of the system (*e.g.* free-energy calculation methods, like thermodynamic integration[91], ball&stick local elevation umbrella sampling[15], λ-dynamics[92]). Artificial terms can also be introduced for simulations performed under special physical conditions, *e.g.* under the effect of an external field[5]. When doing simulations that include unphysical force-field terms, the statistical mechanics of the system is biased and this can sometimes be corrected for when analyzing the results (reweighting).

1.3.4 Parametrization

Once the functional form of each force-field terms has been chosen, a set of parameters has to be defined. For effective force-fields, these are calibrated by comparison with data available from theoretical (quantum mechanical calculations) and experimental sources. Quantum mechanics can provide a lot of information, but this information is often of limited accuracy, in particular concerning intermolecular interactions and solvation, due to

restrictions in system size and level of theory. Experimental measurements can sometimes provide (quasi) direct information about specific parameters (primary data; *e.g.* bond lengths from crystallography, force constants from infrared spectroscopy). Alternatively, they can often be compared to the results of short simulations (secondary data; *e.g.* thermodynamic properties of small organic molecules in the liquid phase). When they are not reliable enough to be used for parametrization purposes, they can still give qualitative information for the validation of the force-field (tertiary data). The evaluation of the sources of data that can be used for the parametrization of a force field is a crucial part of the parametrization itself, as simulations performed with a potential energy function based on incorrect data may give meaningless results.

The parametrization procedure is a high-dimensional and non-linear optimization problem. A universal strategy does not exist for this task, and the choice of the best possible approach depends on the scope of the force-field and on the available data. The choice of the target properties used in the parametrization also influences the procedure, in addition to being determinant for the quality of the resulting force-field.

1.3.5   Calculation of the forces

The simplest potential energy function (without artificial terms) accounting for the sum of force-field terms introduced in Section 1.3.3 is given by

$$
\begin{aligned}
V\left(\mathbf{r};\mathbf{s}\right) &= V_{bonded}\left(\mathbf{r};\mathbf{s}\right) + V_{non-bonded}\left(\mathbf{r};\mathbf{s}\right) \\
&= V_{bond}\left(\mathbf{r};\mathbf{s}\right) + V_{angle}\left(\mathbf{r};\mathbf{s}\right) + V_{improper}\left(\mathbf{r};\mathbf{s}\right) \\
&\quad + V_{dihedral}\left(\mathbf{r};\mathbf{s}\right) + V_{LJ}\left(\mathbf{r};\mathbf{s}\right) + V_{elec}\left(\mathbf{r};\mathbf{s}\right).
\end{aligned}
\tag{1.13}
$$

Once the interactions between all the atoms in the system are defined, along with their specific parameters s_α, the corresponding force $\mathbf{F}_i(\mathbf{r})$ on each atom i can be evaluated as

$$\mathbf{F}_i(\mathbf{r}) = -\frac{\partial V(\mathbf{r})}{\partial \mathbf{r_i}} \tag{1.14}$$

As the potential energy function $V(\mathbf{r})$ is a sum over force-field terms, the force $\mathbf{F}_i(\mathbf{r})$ can also be written as a sum over force-field terms

$$\mathbf{F}_i(\mathbf{r}) = \sum_{\alpha=1}^{N_{terms}} {}^{(n_\alpha)}\mathbf{F}_{i,\alpha}^{(t_\alpha)} \tag{1.15}$$

1.4 Generating configurations

The interactions described in Section 1.3 determine the potential energy surface (PES) of the system, $V(\mathbf{r})$. Such a hypersurface describes the potential energy of the system in terms of selected coordinates. For isolated molecules in the absence of external fields, the potential energy is invariant upon translation or rotation in space. Thus, the potential energy only depends on the internal coordinates of the molecule. These internal coordinates may be represented *e.g.* by simple stretch-bend-torsion coordinates, by symmetry-adapted linear combinations, by redundant coordinates, or by normal mode coordinates, etc. The minimal dimensionality of these spaces matches the $3N$ Cartesian coordinates for an N-atom molecule, minus three translations and three rotations. The PES is therefore generally a high dimensional hypersurface, depending on the size of the system considered. On this surface, stationary points (points where the gradient with respect to all the coordinates is zero) are of particular interest. They represent either energy minima, *i.e.* stable or meta-stable conformations, energy maxima, or saddle-points of various orders.

Exploring the PES and generating configurations for a specific system, so as to be able

to calculate its relevant properties (structural, thermodynamic or dynamical) is the main aim of molecular simulations. For large systems, the visualization of the PES and the enumeration of all possible minimum-energy conformations is impossible. The configuration-generation scheme should therefore automatically focus on the relevant configurations. Different methods have been developed to this purpose. A basic classification can be provided in terms of searching methods, sampling methods and simulation methods.

Searching methods focus on searching the PES for low-energy minima. This includes methods like systematic or random search, genetic algorithms, or homology modelling, typically followed by energy minimization[93] of the discovered configurations. Enhanced-search molecular dynamics method can be also applied for searching purposes. The list of these methods is long. Among many others, one may mention simulated annealing[94,95], PEACS, 4D molecular dynamics[96], local elevation[90] and essential dynamics[97].The main problem associated with searching methods is that the energy is by no means the only criterion for the determination of relevant conformations. A proper generation of relevant statistical information should actually be performed according to the free energy, which includes entropic effects, rather than only the energy.

Sampling methods generate a Boltzmann-weighted ensemble of configurations that can be used to calculate thermodynamic properties of the system. Monte Carlo sampling[2], replica-exchange[98,99] and molecular dynamics with altered masses are just examples, and also in this case a complete list would be longer.

Simulation methods generate a sequence of configurations by mean of a physically motivated equation of motion, and allow for the analysis of the dynamical properties of the system. Molecular dynamics is the most commonly used simulation method, based on Newton's equation of motion. Other equations and methods can be applied, *e.g.* stochastic dynamics[100,101] or Brownian dynamics.

In the following paragraphs, a more extensive description of the methods that are most relevant for this thesis is given. This description is neither complete nor exhaustive, and we refer to the original references for a more detailed presentation.

1.4.1 Molecular dynamics simulation

Molecular dynamics (MD) simulation is a deterministic and time-reversible simulation method based on Newton's second law connecting force with acceleration and mass, $\mathbf{F} = m\mathbf{a}$. It allows to generate a dynamical sequence of configurations for a molecular system by integrating these classical equations of motion. In Hamiltonian formulation and considering a Cartesian coordinate system, these equations read

$$
\begin{aligned}
\frac{\partial H(\mathbf{r}, \mathbf{p})}{\partial \mathbf{p}} &= \underline{\mathbf{M}}^{-1}\mathbf{p} = \dot{\mathbf{r}} \\
\frac{\partial H(\mathbf{r}, \mathbf{p})}{\partial \mathbf{r}} &= \frac{\partial V(\mathbf{r})}{\partial \mathbf{r}} = -\mathbf{F} = -\dot{\mathbf{p}},
\end{aligned}
\tag{1.16}
$$

where $\mathbf{r}$ is the $3N$-dimensional position vector, $\mathbf{p}$ is the corresponding momentum vector, $\underline{\mathbf{M}}^{-1}$ is the diagonal mass matrix, $V(\mathbf{r})$ is the potential energy of the system and the dot over a symbol indicates its time derivative. For systems behaving ergodically, MD simulations can also be used to determine macroscopic thermodynamic properties, becoming a sampling method as well. Note also that MD conserves the total momentum and energy of a system.

Several algorithms (integrators) have been proposed to integrate the classical equations of motion based on a finite timestep, *e.g.* Euler, Verlet[102], leap-frog[103] and velocity-Verlet[104]. The choice of a specific integrator, together with its parameter (integration timestep), is fundamental in determining the cumulative numerical errors. In particular, the timestep should be sufficiently small that even the fastest motion of the system can

be captured.   Still, it should be also sufficiently large to enable an efficient sampling given a certain computational budget. The starting conditions may be equally important only when the feasible equilibration time of a simulation is longer than the characteristic configuration and velocity relaxation times of the system.

1.4.2   Enhanced sampling

For large systems, the potential energy surface typically presents a huge number of local minima.   The barriers separating them may be so high that a normal MD simulation becomes unable to sample a significant extent of the configurational space. The system remains most of the simulation time trapped in one or a few of the minima.   Numerous methods to extend the space visited during a simulation have been proposed. These methods can be implemented either as searching methods or as sampling methods. The extension can rely on the deformation or smoothing of the potential energy surface, in particular of the barriers between the minima (*e.g.* 4D-MD[96], local elevation[90], umbrella sampling[105], EDS[13], solute potential scaling[106], coarse-grained models), or on the scaling of the parameters of the systems, such that an enhanced dynamics is achieved (*e.g.* altered masses MD[107,108], simulated annealing[94,95], adiabatic decoupling).   Additionally the searching or sampling can be performed in a more efficient way using multi-copy techniques (replica exchange[98,99], SWARM-MD[109]).

Most of the work undertaken during this thesis is based on the use of one specific enhanced sampling method, local elevation umbrella sampling[14] (LEUS). In this approach, a memory based biasing potential energy function is applied to penalize previously visited configurations, so that the system is locally elevated and pushed into areas of the configurational space that have not yet been considered. This potential energy function is an artificial time-dependent function $V_{bias}(t)$ that is added to the physical force-field terms

(see Section 1.3). Once the conformational space of interest is homogeneously covered, the time-dependence of the potential energy function is removed (the biasing function is frozen) and V_{bias} is used as an umbrella function which allows for an extensive, yet statistical-mechanically well characterized, sampling of the conformational space. Clearly, the resulting statistics is non-Boltzmann and all the average properties of interest must be evaluated by use of a reweighting procedure. Local elevation umbrella sampling is an extremely powerful method. Yet, the difficulty of identifying appropriate internal coordinates for the biasing as well as the high computational and memory costs when using a large number of such coordinates make it effectively convenient for biases involving low dimensionality spaces only.

1.5 Choice of the boundary conditions

The boundary conditions of a simulation are global constraints enforced on the entire system. This type of constraints can be imposed on instantaneous observables, such that they are enforced at all times during the simulation (hard boundary conditions), or on average observables, such that the average value over a given time window is imposed (soft boundary conditions). Boundary conditions can be applied to the sample surface (spatial), *e.g.* imposing size and shape, or determining how the surroundings and the interface are treated, or to the simulated ensemble (thermodynamic), choosing which thermodynamic properties should be kept constant during the simulation. Internal coordinate (geometric) constraints or restraints, *e.g.* using rigid molecules or enforcing agreement with specific experimental data, can also be viewed as a form of boundary conditions.

In the following, a general description of spatial and thermodynamic boundary conditions is given, focusing for both on methods that are commonly used in molecular simulations.

1.5.1   Spatial boundary conditions

When simulating a macroscopic system by modelling it as a microscopic sample, major artifacts can be introduced. These mainly arise from the fact that the small system size leads to the omission of long-range effects that would be included in the macroscopic system (finite-size effect[110,111]) and that approximations in the treatment of the surface must be introduced (surface effect). The factor that contributes most significantly to finite-size effects is the long-ranged nature of electrostatic interactions. For surface effects, the problems are related to the surface tension appearing at an interface with vacuum.

Different approaches can be used to handle surface effects, and they can in certain cases be completely eliminated, for example by using periodic boundary conditions. Finite-size effects are a more challenging issue in molecular simulations. In principle they can be avoided using implicit solvent models that formally extend the solvation range to infinity, but these models typically lack short-range accuracy and are practically difficult to design and parametrize in a transferable way.

1.5.2   Thermodynamic boundary conditions

When performing a simulation, principles of statistical mechanics are used to extract information about average macroscopic observables from the sampled microscopic instantaneous configurations. The instantaneous observable corresponding to a given property is connected to the corresponding macroscopic observable through ensemble-averaging using the probability distribution ρ_{ens} of the configurations in the simulated ensemble, as

$$A = \int \int d\mathbf{r} d\mathbf{p} \; \rho_{ens}(\mathbf{r}, \mathbf{p}) \mathcal{A}(\mathbf{r}, \mathbf{p}) = \langle \mathcal{A} \rangle_{ens}, \tag{1.17}$$

where $\langle ... \rangle_{ens}$ denotes averaging. Clearly the probability distribution of the configurations depends on the type of ensemble simulated, as defined by its independent thermodynamic parameters. In a simulation, independent extensive quantities (whose values are proportional to the quantity of substance under study) are strictly conserved in time, while for independent intensive quantities (which correspond to local properties) only the average matches the corresponding imposed values, the associated fluctuations only becoming negligible when increasing the system size to the macroscopic limit. Specific combinations of independent thermodynamic parameters can be considered, such as number of particles N, total energy E, volume V, temperature T, pressure P or enthalpy H, defining specific ensembles, *e.g.* microcanonical (NVE), canonical (NVT), isoenthalpic-isobaric (NPH) or isothermal-isobaric (NPT).

Most frequently, molecular simulations are performed at constant (average) pressure and temperature (isothermal-isobaric ensemble), so as to match the most common experimental conditions, and different algorithms have been developed for this purpose.

Constant temperature algorithms (thermostats)

The instantaneous temperature is proportional to the kinetic energy $\mathcal{K}$ of the system, following from the equipartition theorem,

$$\mathcal{T} = \frac{2}{\mathcal{N}_{dof} k_B} \mathcal{K} \quad \text{with } \mathcal{K} = \frac{1}{2} \sum_{i=1}^{\mathcal{N}} m_i \dot{\mathbf{r}}_i^2, \tag{1.18}$$

with $\mathcal{N}_{dof}$ the number of degrees of freedom of the system, k_B the Boltzmann constant, m_i the mass of the particle i and $\dot{\mathbf{r}}_i$ its velocity.

To control the temperature, an algorithm (thermostat) should act on the velocities of the particles. Several methods can be used to keep the (average) temperature constant. Some are constraining the temperature (Hoover-Evans thermostat[112,113], Woodcock ther-

mostat[114]), thus incorrectly setting the fluctuations to zero even for a microscopic system. Others are coupling the temperature to an external bath, allowing for fluctuations of the instantaneous value, and generating a probability distribution that corresponds either approximately (Berendsen thermostat[115]) or exactly (Nosé-Hoover[116–118], Nosé-Hoover chain[119]) to a canonical ensemble. The details of the possible thermostats are not shown here and we refer to Ref.[16] for a more detailed description.

Constant pressure algorithms (barostats)

The instantaneous pressure depends on the kinetic energy $\mathcal{K}$, virial $\mathcal{W}$ and volume $\mathcal{V}$ of the system, as

$$\mathcal{P} = \frac{2\left(\mathcal{K} - \mathcal{W}\right)}{3\mathcal{V}} \quad \text{with } \mathcal{W} = \frac{1}{2}\sum_{i=1}^{\mathcal{N}-1}\sum_{j=1,j>i}^{\mathcal{N}} \mathbf{r_{ij}} \cdot \mathbf{F}_{ij}, \tag{1.19}$$

with $\mathbf{F}_{ij}$ the force by particle on particle i and $\mathbf{r}_{ij}$ the vector connecting the two particles (for simplicity this expression only considers pairwise forces).

   An algorithm that controls the pressure (barostat) must act on the atomic coordinates, the velocities or the volume of the system. Algorithms developed to impose a specific (average) pressure on the system can constrain the pressure (Evans-Moriss barostat[120,121], Abrahams barostat[122]) or can weakly couple the pressure (Berendsen barostat[115]), producing approximate fluctuations for an isobaric ensemble. Other algorithms are available to ensure fluctuations corresponding more closely (Nosé-Hoover-Andersen[118,123]) or exactly (MTK[124]) to the isobaric ensemble. A detailed description of these algorithms can be found in the cited references.

1.6 Solvent effects

When modelling a bio-chemical system and trying to remain as close as possible to experimental conditions, an additional aspect that must be considered is the treatment of the solvent. Solvent molecules play a fundamental role in the stability of macro-molecules[125]. Additionally, the solvent can be a major determinant of conformational[126–128] and reactivity properties of biomolecular solutes.

Solvation involves different types of intermolecular interactions, *e.g.* hydrogen bonding, ion-dipole, and dipole-dipole interactions plus van der Waals forces. It influences the properties of a solute molecule in a very complex way, involving very different phenomena. These include *e.g.* the dielectric screening of the intramolecular electrostatic interactions, the direct competition with the solute regarding these interactions, as well as the preferential solvation of certain parts of the solute molecule (hydrophobic/hydrophilic effects).

Solvents can have very diverse characteristics. One of the fundamental aspects considered in the evaluation of their properties is polarity, used for the classification into polar and apolar solvents. A rough measure of the polarity of a solvent is commonly given by its dielectric permittivity, even though polarity includes other aspects of the electrostatic and van der Waals interactions between solute and solvent (*e.g.* H-bonding).

The "basic" solvent for most of the chemical and all (bio)chemical processes is water, also one of the most polar solvents (relative dielectric permittivity $\varepsilon = 80$ and a strong propensity towards formation of hydrogen bonds). In fact, water is one of the most extraordinary substances on Earth, and its properties are far from being exhaustively understood.

Due to both the important role of the solvent and the variety and complexity of the solvation interactions, the parametrization of solvent models represents at the same time

a crucial and a highly non-trivial task in MD simulations. As a striking illustration of this point, after over four decades of computer simulations involving explicit water molecules[129], there is no universally valid water model. For example, the models that reproduce most accurately the properties of the pure liquid (*e.g.* TIP5P[130]) are not automatically showing the best ionic solvation properties[131], so that apparently less accurate models are comparable (sometimes better) for solvation purposes. In Chapter 2 of this thesis, the sensitivity of the macroscopic properties of water to specific molecular parameters of the model will be analyzed, focusing in particular on characteristics related to polarity (dielectric permittivity, hydrogen bonding propensity).

In the remainder of the thesis (Chapter 3 and Chapter 4), special attention is dedicated to the analysis of the effect of the solvent on intramolecular hydrogen bonding in the solute.

A hydrogen bond (H-bond) results from the electrostatic attraction between a hydrogen atom bound to an electronegative atom (*e.g.* oxygen or nitrogen) and another electronegative atom. This kind of interaction can occur between distinct molecules or within a single molecule, and it is commonly regarded as a major driving force in (bio)chemical processes[126,127]. However, in an aqueous environment, the conformational influence of H-bonds is probably limited to buried H-bonds, and represents in this case a steering (as opposed to driving) force[132]. In contrast, solvent-exposed H-bonds are likely to provide only a marginal (possibly adverse) conformational driving as well as steering force, as the solvent-exposed H-bonded interaction is screened by the solvent dielectric response as well as subject to H-bonding competition by the solvent molecules. As the polarity of the solvent is decreased, however, solvent-exposed intramolecular H-bonding is expected to evolve from a negligible (possibly adverse) to a very significant (favorable) driving force.

1.7 Carbohydrates

In this thesis, carbohydrates are used as model systems to analyze the solvent effect on intramolecular hydrogen bonding.

Carbohydrates represent one of the most important components of the biochemical structure and activity of the cell, including rigidity (*e.g.* cellulose in plant cell walls), energy processing (*e.g.* amylose in starch), cell recognition and cell signaling. Additionally, they play a fundamental role in many biochemical and technological applications.

A key feature of carbohydrate molecules is their great variety and flexibility. Unlike proteins, already at the basic unit level (monosaccharides), carbohydrates present a huge conformational diversity. The large number of possible isomers is due to the presence of several chiral centers and to a great variety of possible functionalizations. This huge number of possible monosaccharides can be connected in chains, again with a great diversity of possible sequences, linkages and extents of branching. The chain flexibility of carbohydrates is principally related to variations of the ϕ and ψ glycosidic dihedral angles, while a more detailed conformational analysis must also consider the dihedral angles defining the orientations of the exocyclic substituents. Several effects play a role in determining the conformational properties of carbohydrate molecules. In particular steric, stereoelectronic, electrostatic and solvent effects are of fundamental importance for understanding both variety and flexibility.

While being one of the most complex and challenging type of biomolecules, carbohydrates also represent a perfect system to study the importance of hydrogen bonding in determining conformational preferences. Each hydroxyl group in a carbohydrate acts as a possible H-bond donor or acceptor. This means that even for monosaccharides, the potential formation of hydrogen bonds has to be considered in evaluations regarding conformational changes.

The carbohydrate basic units are experimentally well characterized and important conformational aspects are well defined. In Chapter 3 of this thesis the focus is on the evaluation of the solvent-modulated influence of intramolecular hydrogen bonding on the hydroxymethyl group rotation, considering two of the most common monosaccharides, glucose and galactose. Then, in Chapter 4, the influence of intramolecular hydrogen bonding is analyzed in the context of disaccharides, evaluating its relevance in the conformational properties of the glycosidic linkage.

Influence of molecular geometry and charge distribution on the collective properties of the liquid water.

Abstract

Water-like models are generated starting from the parameters of the simple point charge (SPC) water model, to study the impact of the solvent geometry and charge distribution on the collective properties of the liquid. In a NVT series of models, the simulations are performed at constant volume and only the OH bond length and the partial charges are modified. In a NPT series of models, the simulations are performed at constant pressure and the Lennard-Jones repulsion coefficient is adjusted as well, to reach a density appropriate for liquid water at room temperature and atmospheric pressure. The systems are analyzed in terms of thermodynamic properties (pressure, total potential energy, dielectric permittivity, hydrogen-bonding capacity), structural properties (radial distribution function and dipole-dipole orientation correlation function), dynamic properties (diffusion coefficient, rotational correlation time and the Debye relaxation time) and solvation properties (for model hydrophobic, polar and ionic solutes). The correlation between these properties gives insight into the nature of certain characteristics of water and the sensitivity of its macroscopic properties to the above molecular properties. Additionally, these water-like models can be used as liquid environment in the study of specific solutes, to investigate separately the role of the dielectric permittivity and hydrogen-bonding capacity on the conformational properties of the solute molecule.

2.1 Introduction

Water is ubiquitous on earth and the "basic solvent" for many chemical and all biochemical processes. As a pure liquid it has a number of unique properties[133–135]. Among these, one may cite the volume expansion upon freezing, the density maximum in the liquid state at 4° C, the particularly high heat capacity and viscosity. As a solvent, water is a major determinant of the conformational and reactivity properties of molecular solutes. Many of these features are related to the high dielectric permittivity and strong hydrogen-bonding capacity that characterize water.

The connection between molecular and macroscopic properties is given by statistical mechanics. For simple model systems, like ideal gases or harmonic crystals, this connection can be established analytically. However, in most cases, the chemical systems considered are too complex and numerical methods must be used instead. Molecular dynamics (MD) simulation is a unique tool that can be employed for this purpose. In addition, MD allows to investigate the sensitivity of macroscopic properties on molecular structure in unphysical (*i.e.* experimentally inaccessible) situations.

Numerous MD studies of water have been reported in the literature and many different water models have been proposed[130,136–143], from simple classical three-charge models[136,137], to more complex ones with up to five interaction sites[130] and possibly including explicit electronic polarizability[144]. In many cases, the parametrization of these water models focuses on the properties of water, trying to reproduce experimental data over a broad range of different properties for the pure liquid, and sometimes the solid. In other cases, the water models are engineered specifically to be used as solvent in biochemical simulations. For these models some properties are of higher relevance than others, because they are more important in terms of the solvation properties of the model.

The impact of specific parameters of a water model on its collective liquid properties is

generally analyzed during the parametrization of that model. Considering models derived from the simple point charges (SPC) water model[136], the influence of the partial charges and of the geometry of the model (bond length and bond angle) has been investigated by Glättli[138,139]. For SPC and polarizable SPC-like models, the influence of the partial charges and of the polarization parameters has been investigated by Kuntz[145] and Bachmann[146]. These studies focused mainly on a small number of possible candidates for the development of a new model, and the range of alternative values of the parameters that were considered were close to the optimal choices.

The present study considers water-like models derived from the SPC model[136] by altering the geometry (OH bond distance) and atomic partial charges (O and H charges), keeping exactly or approximately the same molecular size (same or similar Lennard-Jones interaction parameters). These models are simulated at constant number of molecules and temperature, under either isochoric (NVT) conditions close to the experimental water density, or isobaric (NPT) conditions at atmospheric pressure. When NVT conditions are employed, the Lennard-Jones interaction parameters are kept identical to those of SPC water and the pressure in the system may significantly differ from the atmospheric pressure. When NPT conditions are employed, the coefficient $C_{12}^{1/2}$ defining the repulsive term of the Lennard-Jones interaction function is slightly altered, so as to reach the experimental density of water at equilibrium. The main goal of the study is to examine the sensitivity of the collective liquid properties of the fluid to the molecular parameters considered, in order to provide insight into what makes water so special. In addition, the wide spectrum of water-like models with tunable dielectric permittivity and hydrogen-bonding capacity engineered in this way can be used as liquid environment for specific solutes, to investigate the role of these two collective properties on the conformational equilibrium of a solute molecules. Based on the NVT water-like models, this approach will be used in the context of mono- and disaccharides in Chapter 3 and Chapter 4 of this thesis.A simple application

of the NPT water-like models is also given here. The solvation free energies of different chemical species (hydrophobic, polar or ionic) in solvents of different permittivities and hydrogen-bonding capacities are calculated using thermodynamic integration (TI). In this way, the importance of the two properties on the solvation is analyzed, considering different possible chemical natures of the solute.

2.2 Computational procedure

2.2.1 Water-like models

Water-like models were generated starting from the parameters of SPC water[136] to study the impact of the solvent geometry and charge distribution on the collective properties of the liquid. All the models considered were fully rigid. Two series of models were considered, further referred to as the NVT and the NPT series. In both series, the oxygen-hydrogen (OH) bond length b as well as the charges on the oxygen and hydrogen atoms $q_O = -2q$ and $q_H = q$ were systematically scaled using factors s_b and s_q, respectively, with values in a range between 0.1 and 1.5 by increments of 0.1.

In the NVT series, all the other parameters (bond angle, Lennard-Jones dispersion coefficient $C_6^{1/2}$ and repulsion coefficient $C_{12}^{1/2}$) were kept fixed at the values corresponding to SPC water. In that way, a set of 225 models was generated. A total of 195 models were further analyzed, the other 30 presenting combinations of s_b and s_q too extreme to be successfully simulated with a normal timestep (see Section 2.2.3).

In the NPT series, a scaling factor s_c was also applied to the Lennard-Jones repulsion coefficient $C_{12}^{1/2}$. This factor was optimized so as to reproduce the experimental density of liquid water at room temperature under atmospheric pressure. Out of the starting 225 models, the calibration of an appropriate scaling factor s_c was only possible for 62

combinations of the scaling factors s_b and s_q (see Sections 2.2.4 and 2.3.2).

The 195 plus 62 water-like models are labeled with the letter W (NVT series) or Y (NPT series) followed by the values of the two scaling factors s_b and s_q as superscript and subscript, respectively, *i.e.* as $W_{s_q}^{s_b}$ and $Y_{s_q}^{s_b}$, respectively. The scaling factor s_c in the NPT series is not explicitly indicated in the symbol since it depends implicitly on s_b and s_q *via* calibration against the experimental water density. In the NVT series, one has $s_c = 1$. Note that the original SPC water model is equivalent to $W_{1.0}^{1.0}$ but not to $Y_{1.0}^{1.0}$. The latter model relies on a s_c value that slightly differs from 1.0 (namely 0.97352), as the equilibrium density of SPC water is not exactly equal to the experimental value for liquid water.

2.2.2   Simulation parameters

All MD simulations were performed using the GROMOS simulation package[81–84] considering systems of 1000 (NVT series) or 3000 (NPT series) molecules simulated in cubic boxes under periodic boundary conditions. The OH bond lengths and the HH distance were constrained by application of the SHAKE algorithm[53] with a relative geometric tolerance of 10^{-4}. For simulations involving models of both the NVT and NPT series, the temperature was maintained close to its reference value of 298.15 K by weakly coupling[115] the systems to an external bath using a relaxation time of 0.1 ps. For simulations involving models of the NVT series, the box volume was kept constant with dimensions appropriate for a density $\rho_{\text{SPC}} = 968$ kg m^{-3} (box edge length 3.25 nm), which is close to the equilibrium value[146] for SPC water at 298.15 K and 1 bar. For simulations involving models of the NPT series, the pressure (calculated based on an atomic virial) was maintained close to its reference value of 1 bar using the weak coupling[115] method (isotropic coordinate scaling) with a relaxation time of 0.5 ps. For this coupling, the isothermal compressibility of

the system was set to the experimental value of $7.51 \cdot 10^{-4}$ (kJ mol^{-1}nm^{-3})$^{-1}$ for liquid water[147]. This parameter was not adjusted for the different artificial solvents, because it is combined with the arbitrary choice of a pressure relaxation time.

The equations of motion were integrated using the leap-frog algorithm[103] with a timestep of 1 fs. For the models involving $s_b = 0.1$, $s_b < 0.5$ together with $s_q < 0.5$, or $s_b = 1.5$ together with $s_q > 1.1$, the timestep was reduced to 0.5 fs to avoid SHAKE failures. As detailed in Section 2.2.3, 30 water-like models with extreme geometries and charge distributions could not be simulated even with a 0.5 fs timestep, and were not considered further. The non-bonded interactions were calculated using a twin-range scheme[69], with short- and long-range cutoff distances set to 0.8 and 1.4 nm, respectively, and an update frequency of 5 timesteps for the short-range pairlist and intermediate-range interactions. A reaction-field correction[60,65] was applied to account for the mean effect of electrostatic interactions beyond the long-range cutoff distance. The relative dielectric permittivity $\varepsilon_{\mathrm{RF}}$ used for this correction was set equal (self-consistency) to the actual permittivity ε of the specific solvent model (see Sections 2.2.3 and 2.2.4).

2.2.3 Simulations of the NVT models

The models W$_{s_q}^{s_b}$ of the NVT series rely on a scaling of the OH bond length by s_b and of the charges by s_q without any other parameter change relative to SPC water (in particular, $s_c = 1.0$). They are investigated in simulations of 1000 molecules in cubic periodic boxes at 298.15 K and at a constant volume determined by the equilibrium density of $\rho_{\mathrm{SPC}} = 968$ kg m^{-3} corresponding to the SPC water model at 298.15 K and 1 bar (value obtained by independent simulations, comparable to *e.g.* 972 kg m^{-3} in Ref.[146]). Starting from a well equilibrated (0.5 ns equilibration) SPC water box containing 1000 molecules under the indicated conditions, the bonds of all molecules were scaled applying the specific scaling

factor s_b to the OH bonds while keeping the oxygen atom fixed. The resulting scaled configurations were then simulated under NVT conditions applying the scaling factor s_b (to OH bond lengths and HH distances) and s_q (to the charges) in the topology.

Simulations of 12 ns were performed under NVT conditions for the calculation of the dielectric permittivity, starting with $\varepsilon_{\mathrm{RF}} = 61$ and iterating[148] until the relative error between the ε value calculated from the simulation and the $\varepsilon_{\mathrm{RF}}$ value used as simulation parameter was below 1%. A number of iterations between 4 and 10 was necessary to reach the convergence, depending on the model considered. Some models with very short bond lengths ($s_b = 0.1$), with short bond lengths and low charges ($s_b < 0.5$ and $s_q < 0.5$) and with very long bond lengths and high charges ($s_b = 1.5$ and $s_q > 1.1$) could not be simulated even with a 0.5 fs timestep due to SHAKE failures, and were not considered further. A list of the 195 feasible NVT models is given in Suppl. Mat. Table 2.S.1. The not feasible models are indicated in black in Figures 2.3, 2.5, 2.7, 2.9 and 2.10.

2.2.4   Simulations of the NPT models

The models $\mathrm{Y}_{s_q}^{s_b}$ of the NPT series rely on a scaling of the OH bond length by s_b and of the charges by s_q, along with a scaling of $\mathrm{C}_{12}^{1/2}$ by s_c. They are investigated in simulations of 3000 molecules in cubic periodic boxes at 298.15 K and at a constant pressure of 1 bar. The value of s_c is preoptimized to reproduce an average density $\rho_{\mathrm{exp}} = 997$ kg m^{-3}, the experimental density of liquid water under these conditions[147]. Starting from a well equilibrated (0.5 ns equilibration) SPC water box containing 3000 molecules at density ρ_{SPC}, the bonds of all molecules were scaled applying the specific scaling factor s_b to the OH bonds while keeping the oxygen atom fixed. The resulting scaled configurations were then simulated under NPT conditions applying the appropriate scaling factors s_b and s_q in the topology. After 0.2 ns equilibration with $s_c = 1.0$, a series of successive 3 ns simulations

was performed with iterative update of s_c, until the density reached a value within 5% from the target experimental water density ρ_{exp}. The optimization of s_c was only possible for 62 models (see Section 2.3.2). The final boxes with the correct density were simulated multiple times during 8 ns for the self-consistent optimization of the dielectric permittivity (ε *vs.* ε_{RF}), following the same iterative procedure as employed for the NVT models (see Section 2.2.3).

2.2.5 Property analysis

The procedures of Sections 2.2.3 and 2.2.4 led to a set of 195 and 62 water-like models in the NVT and NPT series, respectively. In both series, the value of $\varepsilon_{\mathrm{RF}}$ is self-consistent with that of ε within 1%. In the NPT series, the value of ρ at 1 bar is within 5% of the experimental density ρ_{exp} of liquid water. The corresponding final parameters (s_b, s_q, s_c), along with the self-consistent value of ε and the exact value of ρ (NPT series) are listed in Supp. Mat. Tables 2.S.1 and 2.S.2. Based on these final models, an extensive analysis of the collective properties of the system was performed. The trajectories considered for these analyses are described below, while the specific analyses are detailed in Section 2.2.6.

The NVT simulations relied on 21 ns production runs for the calculation of the different properties. Five repeats of these simulations with different starting configurations (coordinates and velocities) taken from the last 1 ns of the final calibration runs were performed to obtain an estimate of the statistical error on the calculated properties. During the production runs, configurations of the system were saved every 1 ps. Additional 30 ps simulations were performed saving the configurations every 1 fs to permit a better estimation of the short-time dynamic properties (rotational correlation times, see Section 2.2.6).

The NPT simulations relied on 12 ns production runs for the calculation of the different

properties. The statistical error was again estimated from five repeats of these simulations. During the production runs, configurations of the system were saved every 1 ps. Additional 30 ps simulations were also performed saving the configurations every 1 fs to permit a better estimation of the short-time dynamic properties (rotational correlation times, see Section 2.2.6).

Unless otherwise specified, the collective properties were evaluated from the final 12 ns of the 21 ns production runs (NVT series) or the final 6 ns of the 12 ns production runs (NPT series). The reported values are averaged over the five repeats and the error is estimated as the corresponding standard deviation.

2.2.6   Properties monitored

Multipoles

The multipoles of the models resulting from the 225 combinations of s_q and s_b were calculated. They depend on the geometric properties of the molecules (bond length and angle) and on the charges, but not on the Lennard-Jones interaction parameters. The values are therefore the same for equally scaled NVT and NPT models.

The dipole μ, the linear component Q_0 of the quadrupole, the square component Q_T of the quadrupole, the linear component Ω_0 of the octupole, and the square component Ω_T of the octupole are calculated using the expressions

$$\mu = 2qb \, \cos\alpha \, s_q s_b \tag{2.1}$$

$$Q_0 = -2qb^2 \left(\frac{1}{2} \, \sin^2\alpha - \cos^2\alpha \right) \, s_q s_b^2 \tag{2.2}$$
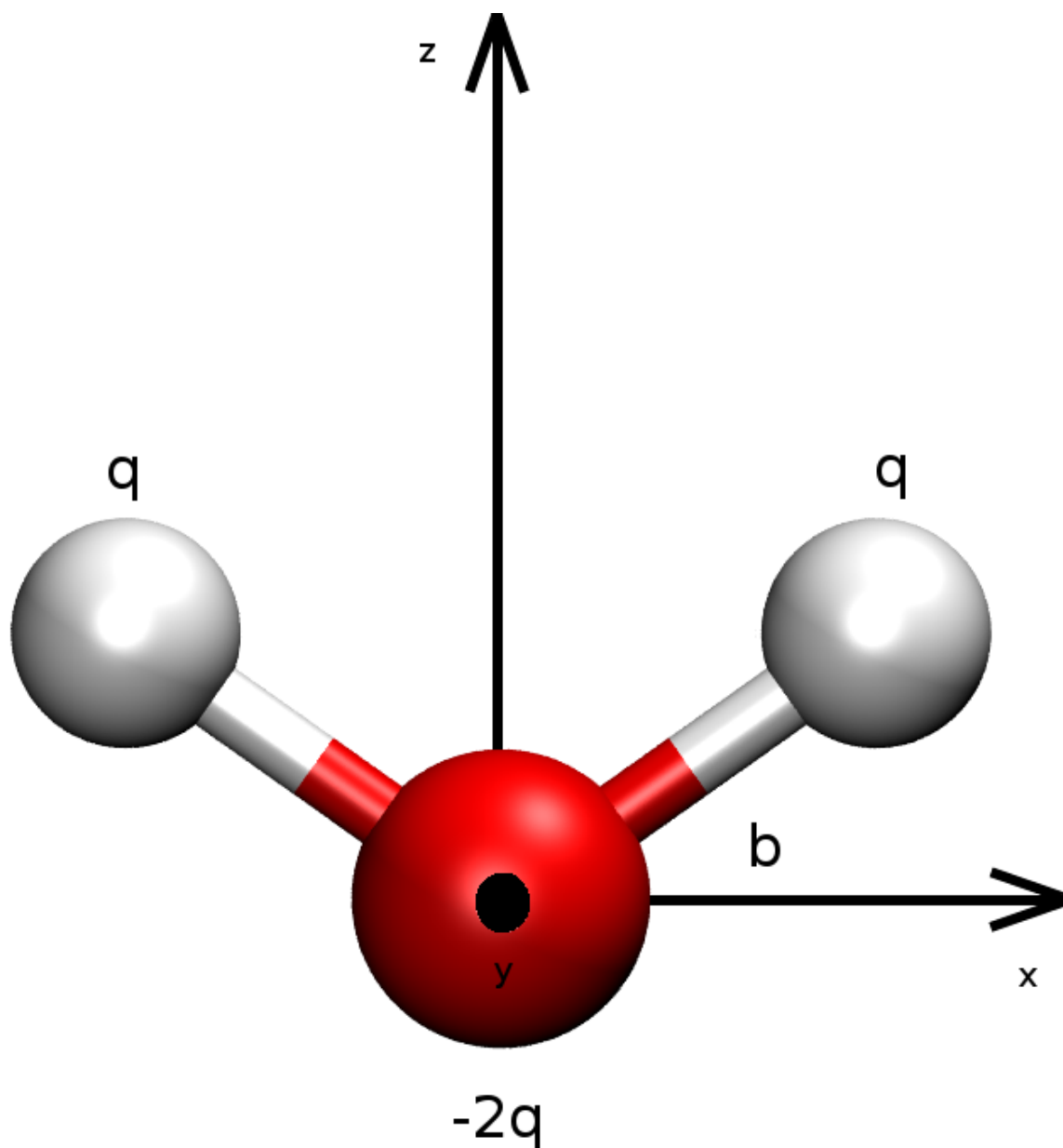
Figure 2.1: Schematic representation of a water molecule described by the SPC model with OH bond length $b = 0.1$ nm, HH distance $h = 0.1633$ nm, HOH angle $\theta = 109.47°$, partial charges $q = 0.41$ e on each hydrogen atom and $-2q = -0.82$ e on the oxygen atom. The coordinate system has its origin O at the oxygen atom.

$$Q_T = \frac{3qb^2 \, \sin^2\alpha}{2} \, s_q s_b^2 \tag{2.3}$$

$$\Omega_0 = -2qb^3 \left(\frac{3}{2} \, \sin^2\alpha \, \cos\alpha - \cos^3\alpha\right) \, s_q s_b^3 \tag{2.4}$$

$$\Omega_T = \frac{5qb^3 \, \sin^2\alpha \, \cos\alpha}{2} \, s_q s_b^3 \tag{2.5}$$

In these equations, q is the charge of the hydrogen atom in the SPC model (0.41 e), b is the OH bond length in this model (0.1 nm), and $\alpha = \theta/2$ is half of the angle between the two OH bonds in this model ($\theta = 109.47°$), as illustrated in Figure 2.1. A derivation of these equations can be found in Appendix 2.A.

Pressure

For the simulations performed under NVT conditions, the pressure P of the system (based on an atomic virial) was monitored during the 12 ns analysis period. The pressure is calculated as

$$P = \frac{2(K - W)}{3V}, \tag{2.6}$$

where K is the kinetic energy, W the virial and V the volume of the system. For the NPT simulations, the pressure is on average 1 bar owing to the barostat coupling.

Density

For the simulations performed under NPT conditions, the density ρ of the system was monitored during the 6 ns analysis period. The values are all in a range of 5% around the

experimental value $\rho_{\mathrm{exp}} = 997$ kg m^{-3} for liquid water, as this was the target for the choice of the Lennard-Jones scaling factor s_c (see Section 2.2.4). For the NVT simulations, the density is always $\rho_{\mathrm{SPC}} = 968$ kg m^{-3} owing to the fixed box volume.

Potential Energy

The total potential energy U per molecule for both NVT and NPT models was monitored over the analysis period (12 ns NVT, 6 ns NPT). If desired, these values can be related to the molar heat of vaporization ΔH_{vap} according to

$$\Delta H_{vap} = -U + P\Delta V + Q_{int} + Q_{ext} \approx -U + RT + Q, \tag{2.7}$$

where P is the pressure, ΔV the molar volume change between the liquid and the gas, R the gas constant and T the absolute temperature (298.15 K). The quantities Q_{int} and Q_{ext} are quantum corrections[133]. The first correction Q_{int} is due to the rigid treatment of bonds and bond angles, and accounts for the difference in vibrational energy of a water molecule between the liquid and the gas phases. The second correction Q_{ext} is due to the intermolecular interactions in the liquid, and is the difference between the intermolecular vibrational energies calculated quantum-mechanically and classically. This treatment assumes that the intramolecular and intermolecular vibrational modes are decoupled. Both corrections are calculated approximating quantum mechanically the single modes as harmonic oscillators and assuming classically an energy contribution RT for each degree of freedom that is not constrained. At room temperature, one has[133,149] $Q_{int} = 3.54$ kJ mol^{-1} and $Q_{ext} = -3.77$ kJ mol^{-1}, which adds up to a total quantum correction $Q = -0.23$ kJ mol^{-1}. This relation allows for a comparison with experimental data.

Static Relative Dielectric Permittivity ε

The static relative dielectric permittivity ε of the systems is obtained from the fluctuation in the total dipole moment of the system over the analysis period (12 ns NVT, 6 ns NPT) using the Kirkwood-Fröhlich-type equation derived by Neumann[150]

$$(\varepsilon - 1)\left(\frac{2\varepsilon_{RF} + 1}{2\varepsilon_{RF} + \varepsilon}\right) = \frac{<\mathbf{M}^2> - <\mathbf{M}>^2}{3\varepsilon_0 V k_B T}, \tag{2.8}$$

where ε_{RF} is the relative dielectric permittivity of the reaction-field continuum that is used in the simulation, $\mathbf{M}$ is the total dipole moment of the computational box, V is the volume of the box, k_B is the Boltzmann constant, T is the absolute temperature, and ε_0 is the dielectric permittivity of vacuum. For both NVT and NPT simulations, the calculated value of ε is within 1% of ε_{RF} by construction.

Hydrogen-bonding Capacity n_H

The average number n_H of hydrogen bonds per molecule was calculated using a geometric criterion. A hydrogen bond is considered to be present if the distance between a hydrogen atom H connected to the donor atom O is within 0.25 nm of the acceptor atom O, and the OHO angle is larger than 135°. To evaluate this property, the water models for which $s_b \neq 1.0$ were first rescaled to the normal bond length at fixed position of the oxygen atom. For each system, 10 different molecules were chosen and the number of hydrogen bonds was calculated from five separate sections of 0.3 ns each (1.5 ns total time) selected along the 21 ns (NVT) or the 12 ns (NPT) production period.

Radial Distribution Function $g(r)$

The structure of liquid water is characterized by short-range order and long-range disorder. This is reflected in the radial distribution function $g(r)$, which is experimentally (although indirectly) available through x-ray diffraction. The pair distribution function is calculated over the analysis period (12 ns NVT, 6 ns NPT) as

$$g(r) = \frac{1}{4\pi\eta r^2 dr} \frac{1}{N} \left\langle \sum_i^N \sum_{j\neq i, r<r_{ij}(t)\leq r+dr}^N 1 \right\rangle, \tag{2.9}$$

where N is the total number of molecules, η the number density, dr the window size (set to 0.0017 nm), $r_{ij}(t)$ the minimum image distance between the oxygen atoms of molecules i and j at time t, and $\langle ... \rangle$ denotes averaging over time. This function gives the probability of finding a pair of molecules at certain distance relative to the corresponding probability for a completely uniform distribution at the same density.

Dipole-dipole orientation correlation function $c(r)$

The structure of liquid water is characterized by a strong correlation of the dipole orientations among neighboring molecules, corresponding to hydrogen-bonding orientations. At longer range this correlation disappears. Such a behaviour is reflected in the dipole-dipole orientation correlation function $c(r)$, calculated over the analysis period (12 ns NVT, 6 ns NPT) as

$$c(r) = \frac{\frac{1}{N} \left\langle \sum_i^N \sum_{j\neq i, r<r_{ij}(t)\leq r+dr}^N \mathbf{e}_i(t) \cdot \mathbf{e}_j(t) \right\rangle}{4\pi\eta r^2 dr g(r)}, \tag{2.10}$$

where $\mathbf{e}_i(t)$ is a unit vector along the dipole moment of water molecule i at time t.

Self-Diffusion Coefficient D

The self-diffusion coefficient was calculated according to the Einstein relation from the long-time limit of the mean-square displacement over the analysis period (12 ns NVT, 6 ns NPT), as

$$D = \lim_{t \to \infty} \frac{\langle (\mathbf{r_i}(\tau + t) - \mathbf{r_i}(\tau))^2 \rangle}{2dt}, \qquad (2.11)$$

where the $\mathbf{r}(t)$ corresponds to the position vector of the center of mass of molecule i at time t, and the averaging $\langle ... \rangle$ is performed over both time origins τ and water molecules i. In practice, a linear least-squares fit was performed considering the region of linearity of the mean-square displacement function, taking into account the maximum number of time-points resulting in a coefficient of determination (square of the correlation coefficient) R^2 of at least 0.9.

Rotational Correlation Times τ_l^α

Reorientational correlation functions $C_l^\alpha(t)$ are calculated for two different molecular axes, namely the HH vector and the OH vector, according to

$$C_l^\alpha(t) = \langle P_l(\mathbf{e_i}^\alpha(\tau + t) \cdot \mathbf{e_i}^\alpha(\tau)) \rangle, \qquad (2.12)$$

where P_l is the Legendre polynomial of order l, $\mathbf{e_i}^\alpha$ is a unit vector pointing along the rigid-body α-axis of molecule i (see Figure 2.1), and the averaging $\langle ... \rangle$ is performed over both time origins τ and water molecules i. Assuming an exponential form for the time-correlation function decay for $C_l^\alpha(t)$,

$$C_l^\alpha(t) = A_l^\alpha \, \exp\left(-\frac{t}{\tau_l^\alpha}\right), \tag{2.13}$$

a single-molecule correlation time τ_l^α can be obtained. For the calculation of these correlation times, simulations of 30 ps were performed saving the configurations every 1 fs. For correlation functions that decayed to 0.1 before 0.2 ps, the rotational times are indicated as $\tau_l^\alpha < 0.2$ ps. For correlation functions that did not show an exponential decay, either within 30 ps or in longer 21 ns simulations (indicative of a non-liquid state), the rotational times are indicated as $\tau_l^\alpha > 21$ ns. In all other cases, the least-squares fit for the calculation of the correlation time τ_l^α was performed considering the longest time range that gave a coefficient of determination R^2 of at least 0.9. The HH and the OH relaxation times can be compared to results from ^{1}H-^{1}H and ^{1}O-^{1}H dipolar relaxation NMR experiments[151].

Debye relaxation time

The Debye relaxation time τ_D can be obtained from the normalized autocorrelation function $C_M(t)$ of the total box dipole moment $\mathbf{M}$ of the system. Assuming exponential decay,

$$C_M(t) = \frac{\langle \mathbf{M}(\tau + t) \cdot \mathbf{M}(\tau) >}{< \mathbf{M}^2(\tau)\rangle} = \exp\left(-\frac{t}{\tau_\phi}\right), \tag{2.14}$$

where the averaging $\langle ... \rangle$ is performed over all time origins τ, the Debye relaxation time τ_D can be evaluated using the relation[152]

$$\tau_D = \frac{2\varepsilon_{\mathrm{RF}} + \varepsilon}{2\varepsilon_{\mathrm{RF}} + 1}\tau_\phi. \tag{2.15}$$

where $\varepsilon_{\mathrm{RF}}$ is the reaction-field permittivity and ε is the static relative dielectric permittivity of the liquid.

For the autocorrelation functions that decayed to 0.1 before 1 ps, the Debye relaxation

times are indicated as $\tau_D < 1$ ps. For the correlation functions that did not show an exponential decay in the 21 ns (NVT) or 12 ns (NPT) production simulations (indicative of a non-liquid state), they are indicated as $\tau_D > 21$ ns. In all other case, the fit for the calculation of the Debye relaxation times τ_D was performed considering the longest time range that gave a coefficient of determination R^2 of at least 0.9.

Correlation between properties

The correlation between the properties calculated as described in the previous sections was analyzed using the Pearson product-moment correlation coefficient r (linear regression coefficient), defined as

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}, \tag{2.16}$$

where the two sets of data $x_1, ..., x_n$ and $y_i, ..., y_n$ represent the two properties to be correlated, $\bar{x}$ and $\bar{y}$ being the corresponding averages. A correlation matrix was obtained in this way by forming all possible pairwise combinations of the calculated properties.

2.2.7   Solvation free energies of model solutes

The solvent models developed in this study can be used as a liquid environment for specific solutes, to investigate the dependency of their properties on specific characteristics of the solvent. These artificial solvents have some advantages over physical ones. First, in contrast to physical solvent series that show simultaneous variations in multiple properties, some properties of the artificial water-like models can be varied (almost) independently from the others. This allows for the disentanglement of different effects. Second, the tuning of the parameters of the artificial models can generate conditions impossible to

reach with physical solvents. These features will be taken advantage of in of this thesis, where the NVT series will be used to analyze the effect of the solvent permittivity and hydrogen-bonding capacity on the conformational properties of carbohydrates. Another simple example of application of these models, here the NPT models, is provided below, namely the calculation of the solvation free energy of three compounds (hydrophobic, polar or ionic), and its dependence on the dielectric permittivity and hydrogen-bonding capacity of the solvent.

The solvation free energy of three compounds was calculated using thermodynamic integration[91] (TI), considering a thermodynamic cycle, as

$$\Delta F_{solute}^{solvation} = \Delta F_{solute \rightarrow dummies}^{vacuum} + \Delta F_{dummies}^{solvation} + \Delta F_{solute \rightarrow dummies}^{solvated}. \tag{2.17}$$

where $\Delta F_{dummies}^{solvation} = 0$ (non-interacting sites) and $\Delta F_{solute \rightarrow dummies}^{vacuum} = 0$ (solutes with no intramolecular interactions; see below). The last term in the right-hand side was calculated as

$$\Delta F = F(B) - F(A) = \int_{\lambda_A}^{\lambda_B} F'(\lambda) d\lambda = \int_{\lambda_A}^{\lambda_B} \left\langle \frac{\partial H}{\partial \lambda} \right\rangle d\lambda \tag{2.18}$$

where the Hamiltonian H depends on the λ parameter.

The compounds considered are argon AR (hydrophobic species), SPC water[136] (polar species) and sodium ion NA+ (ionic species). The simulations involved computational boxes containing one solute molecule and 1000 solvent molecules. The three solutes were simulated in 23 artificial solvent. The artificial solvents considered are a subset of the 62 tunable solvent models $Y_{s_q}^{s_b}$ of the NPT series. The subset considered here correspond to four series S_p^H, S_p^h, S_h^P and S_h^p, that permit to investigate specifically the effect of the following trends: (i) in series S_p^H, the progressive decrease of the permittivity (p subscript) at water-like H-bonding capacity (H superscript); (ii) in series S_p^h, the progressive decrease of the permittivity (p subscript) at lower H-bonding capacity (h superscript);

Series	Solvent	s_b	s_q	s_c	ε	n_H
S_p^H	$W_{1.5}^{0.8}$	0.8	1.5	1.0474	140.0	3.6
	$W_{1.1}^{0.9}$	0.9	1.1	0.9489	74.5	3.3
	$W_{0.7}^{1.2}$	1.2	0.7	0.9242	38.8	3.4
	$W_{0.5}^{1.4}$	1.4	0.5	0.8626	22.6	3.2
S_p^h	$W_{1.4}^{0.6}$	0.6	1.4	0.7886	74.4	2.2
	$W_{1.2}^{0.6}$	0.6	1.2	0.7215	38.2	2.0
	$W_{0.7}^{0.9}$	0.9	0.7	0.7086	20.4	2.2
	$W_{0.6}^{0.9}$	0.9	0.6	0.6562	13.1	2.0
	$W_{0.5}^{1.1}$	1.1	0.5	0.6901	12.7	2.2
S_h^P	$W_{0.9}^{1.1}$	1.1	0.9	0.9858	59.3	3.7
	$W_{1.0}^{1.0}$	1.0	1.0	0.9735	66.7	3.5
	$W_{0.7}^{1.2}$	1.2	0.7	0.9242	38.8	3.4
	$W_{0.8}^{1.1}$	1.1	0.8	0.9242	45.7	3.3
	$W_{0.9}^{1.0}$	1.0	0.9	0.9091	53.4	3.2
	$W_{1.2}^{0.8}$	0.8	1.2	0.9091	82.4	2.9
	$W_{1.1}^{0.8}$	0.8	1.1	0.8565	62.8	2.7
	$W_{1.2}^{0.7}$	0.7	1.2	0.8072	63.3	2.4
	$W_{1.3}^{0.6}$	0.6	1.3	0.7517	54.1	2.1
S_h^p	$W_{0.4}^{1.5}$	1.5	0.4	0.8010	14.3	3.0
	$W_{0.5}^{1.3}$	1.3	0.5	0.7968	19.0	2.9
	$W_{0.5}^{1.2}$	1.2	0.5	0.7424	15.4	2.5
	$W_{0.7}^{0.8}$	0.8	0.7	0.6562	15.1	2.0
	$W_{1.3}^{0.4}$	0.4	1.3	0.6147	13.0	1.7
	$W_{1.4}^{0.4}$	0.4	1.4	0.6408	15.3	1.6

Table 2.1: Definition and simulated properties of the 23 artificial solvent models of the NPT series considered for the solvation free-energy calculations, at 298.15 K and 1 bar. The reported quantities are the scaling factors s_b and s_q applied to the oxygen-hydrogen bond length and to the atomic partial charges, respectively, relative to the SPC water model[136], the static relative dielectric permittivity ε of the liquid as a measure of its polarity, and the average number n_H of hydrogen-bonds per molecule in the liquid as a measure of its hydrogen-bonding capacity. The models are grouped into four series as described in section 2.2.7. The main thermodynamic, dynamic, dielectric and hydrogen-bonding properties being reported in Suppl. Mat. Table 2.S.2.

(iii) in series S_h^P,the progressive decrease of the H-bonding capacity (h subscript) at water-like permittivity (P superscript); (iv) in series S_h^p, the progressive decrease of the H-bonding capacity (h subscript) at lower permittivity (p superscript). The specific solvents considered for each series, along with the corresponding values of dielectric permittiv-

ity and hydrogen-bonding capacity are listed in Table 2.1. The simulation procedure is analogous to that provided in Section 2.2.2 and only the differences will be described here

After 100 ps of equilibration under NVT conditions and 100 ps under NPT conditions, simulations of 240 ps were carried out at 11 equally spaced λ-values between 0 and 1, changing the compound considered into a dummy atom or molecule for which all non-bonded interactions are zero. The perturbation was done in one step for electrostatic interactions and Lennard-Jones interactions, using a non-bonded interaction softness parameter $s_{ij}^{LJ} = 0.5$[153]. Configurations were written to file every 0.5 ps for further analysis.

2.3 Results

2.3.1 Construction of the NVT models

The construction of the NVT models does not require the optimization of s_c, which is kept equal to 1. Not all of the 225 initial s_b and s_q combinations led to a model that could be simulated in practice. Some models with very short bond lengths ($s_b = 0.1$), with short bond lengths and low charges ($s_b < 0.5$ and $s_q < 0.5$) and with very long bond lengths and high charges ($s_b = 1.5$ and $s_q > 1.1$) could not be simulated even with a 0.5 fs timestep due to SHAKE failures. A list of the 195 feasible NVT models is provided in Table 2.S.1. The not feasible models are marked in black in Figures 2.3, 2.5, 2.7, 2.9 and 2.10. Note that $W_{1.0}^{1.0}$ is SPC water.

2.3.2 Construction of the NPT models

To be able to perform NPT simulations at 1 bar close to the experimental water density, the Lennard-Jones repulsive coefficient $C_{12}^{1/2}$ of the NPT models were scaled by a factor
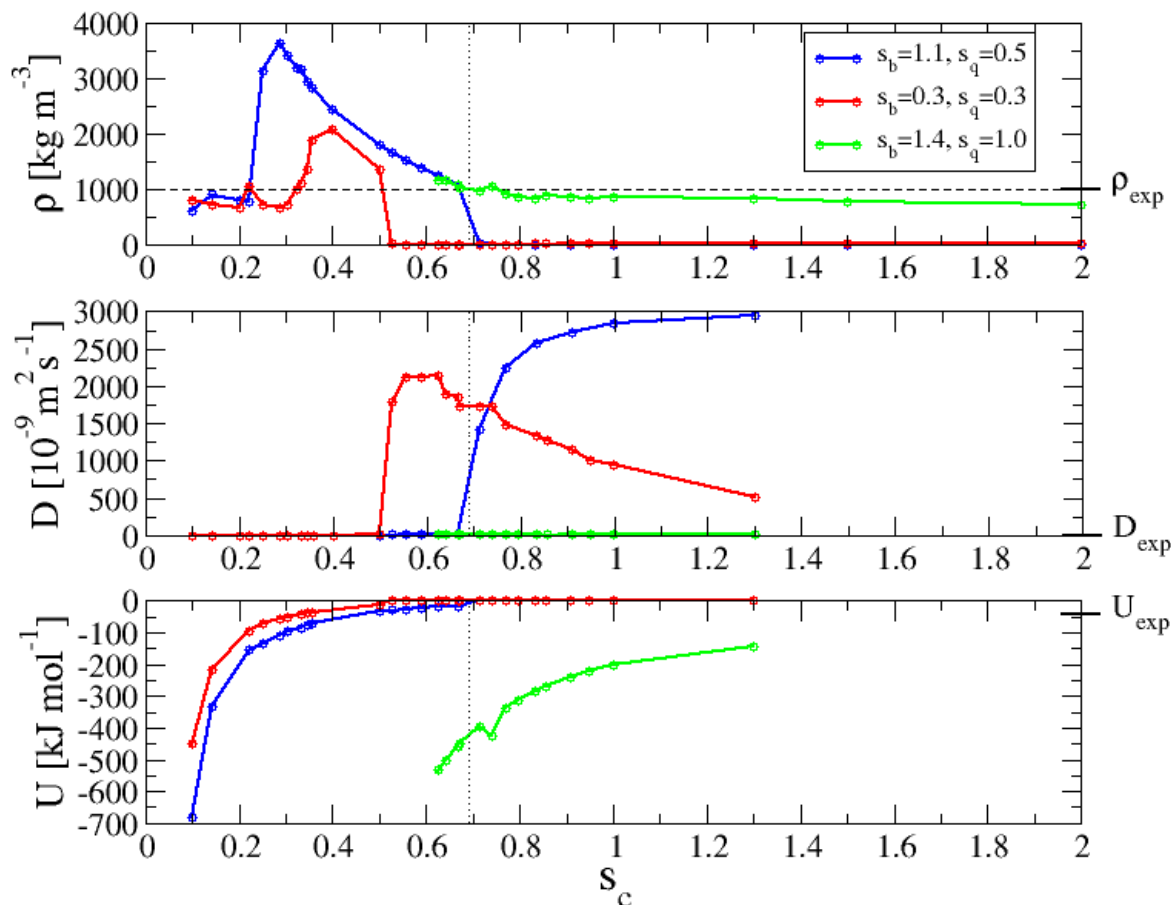
Figure 2.2: Density ρ, diffusion constant D and total potential energy U per molecule as a function of the scaling factor s_c for the repulsive term of the Lennard-Jones interaction energy. Three combinations of s_b and s_q are used to illustrate the three different trends of behaviour. For type A (in blue), a parametrization that gives a liquid density close to the experimental value for water is possible.For type B (in red), the change of phase from gas to liquid upon decreasing s_c occurs at too high density. For type C (in green), the change of phase upon decreasing s_c leads directly to a solid (glassy) state. The horizontal dashed line in the first graph shows the experimental value ρ_{exp} for the density of water. The vertical dotted line shows the value of s_c retained to define the $Y^{1.1}_{0.5}$ model of the NPT series. The reference bars on the right side of the three graphs mark the experimental values ρ_{exp}, D_{exp} and U_{exp}.

s_c. After selecting the value of s_b and s_q, the dependence of density, self-diffusion constant and total potential energy was monitored as a function of the scaling factor s_c in the range 0.1-1.0.

Not all of the 225 initial s_b and s_q combinations led to a density curve achieving the

experimental water density for a given s_c value in this range. In practice, three types of situations were encountered. These situations are illustrated by prototypical examples in Figure 2.2, while the s_b and s_q combinations leading to situations of those three types are shown in Figure 2.3b (using the same color code).

- The first group A (shown in blue in Figures 2.2 and 2.3b), corresponds to models similar to the illustrative example with $s_b = 1.1$ and $s_q = 0.5$. For very high s_c, the system is in a gaseous state. Decreasing s_c reduces the diffusion constant, having initially only a small effect on the density (values on the order of 10 kg m^{-3}) and on the total potential energy (values close to 0 kJ mol^{-1}). When s_c reaches a critical value (different for each model), a change of state occurs from gas to liquid. The density increases abruptly, with a drop of the diffusion constant. The potential energy becomes negative. Decreasing s_c further, the density increases progressively along with a decrease of the potential energy and of the diffusion constant. The systems ultimately ends in a solid (glassy) state. Because in the liquid phase, directly below the gas-to-liquid transition, the density is lower than or close to 997 kg m^{-3}, the experimental density of water can be achieved for a specific value of s_c. This group includes the 62 models that were further considered.

- The second group B (shown in red in Figures 2.2 and 2.3b), corresponds to models similar to the illustrative example with $s_b = 0.3$ and $s_q = 0.3$. The general behaviour of these systems is similar to the one of group A. The main difference is that the experimental density cannot be achieved for any choice of s_c. In the liquid phase, directly below gas-to-liquid transition, the system is already too dense, with densities between 1200 kg m^{-3} and 1600 kg m^{-3}.

- The third group C (shown in green in Figures 2.2 and 2.3b), corresponds to models similar to the illustrative example with $s_b = 1.4$ and $s_q = 1.0$. These systems
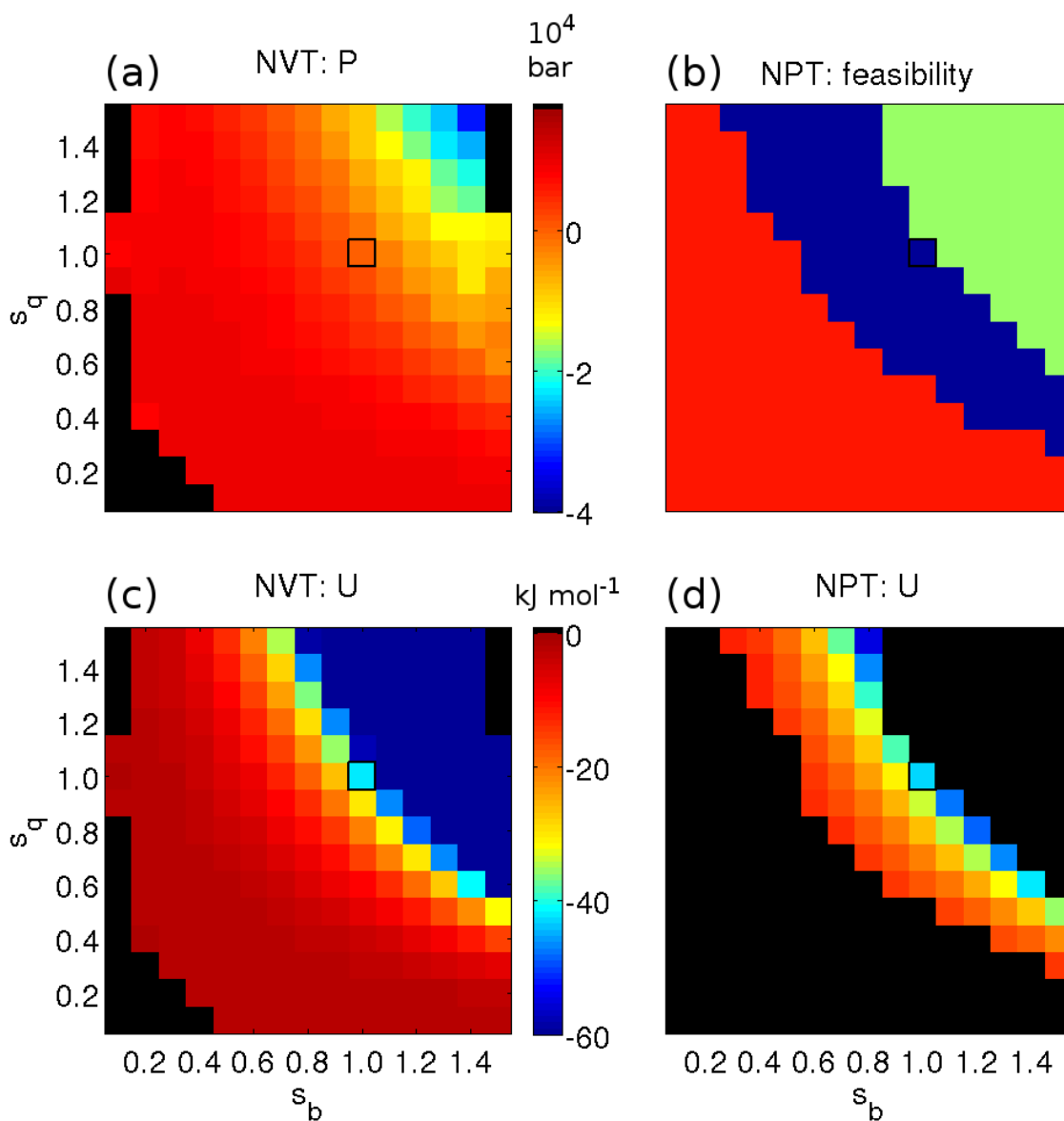
Figure 2.3: Upper panels: (a) pressure P as a function of the scaling factors s_b and s_q for the NVT models. (b) color code illustrate the feasibility of the parametrization of the NPT models: Type A in blue represents the combinations of s_b and s_q that lead to a final $Y_{s_q}^{s_b}$ model; Type B in red represents the combinations that present a liquid phase with a too high density; Type C in green represents the combinations that present a direct gas-to-solid transition (no liquid state). Lower panels: the total potential energy U as a function of the scaling factors s_b and s_q for the NVT models (c) and for the NPT models (d). The color scale is truncated at -60 kJ·mol^{-1} (dark blue). In all the maps, the model with $s_b = 1.0$ and $s_q = 1.0$ (SPC water in the NVT series, closest one in the NPT series) is highlighted.

change directly from the gas phase to a solid (glassy) phase upon decreasing s_c (resublimation). As a result the liquid phase cannot be reached for any value of s_c.

As shown in Figure 2.3b, models with low s_b and s_q typically present a behaviour of type B and models with high s_b and s_q a behaviour of type C. Note that the s_q and s_q combinations of type C are (almost) the same as those presenting a solid (glassy) state in the NVT series (see Section 2.3.4). Only for 62 intermediate combinations with behaviour of type A it is possible to define a final NPT model $Y_{s_q}^{s_b}$. We stress that the only criterion considered for the parametrization of these models was the value of the density, considered acceptable within a range of 5% around the experimental value ρ_{exp} of liquid water. A list of the 62 feasible NPT models is provided in Table 2.S.2. The not feasible models are shown in Figure 2.3b (red and green areas) and marked in black in the following figures. Note that $Y_{1.0}^{1.0}$ is not equivalent to SPC water, as it relies on a s_c value of 0.97352 that slightly differs from 1.0, leading to a density (1003 kg m^{-3}) closer to the experimental value compared to SPC water[146] (973 kg m^{-3}).

2.3.3 Multipole moments

Alterations of the OH bond length and the charges change the electrostatic properties of the molecule. These changes can be characterized by the analysis of the resulting multipole moments. Figure 2.4 shows the variation of the dipole μ, square quadrupole Q_T, and octopoles Ω_0 and Ω_T as a function of s_b and s_q. The expected linear dependence on the charge is evident for all multipole moments, while the dependence on the bond length varies from linear for the dipole to quadratic for the quadrupole and cubic for the octopoles. The multipoles have a determinant role on the properties of a water model[154–160,160–163]. This influence will be discussed in Section 2.3.5.

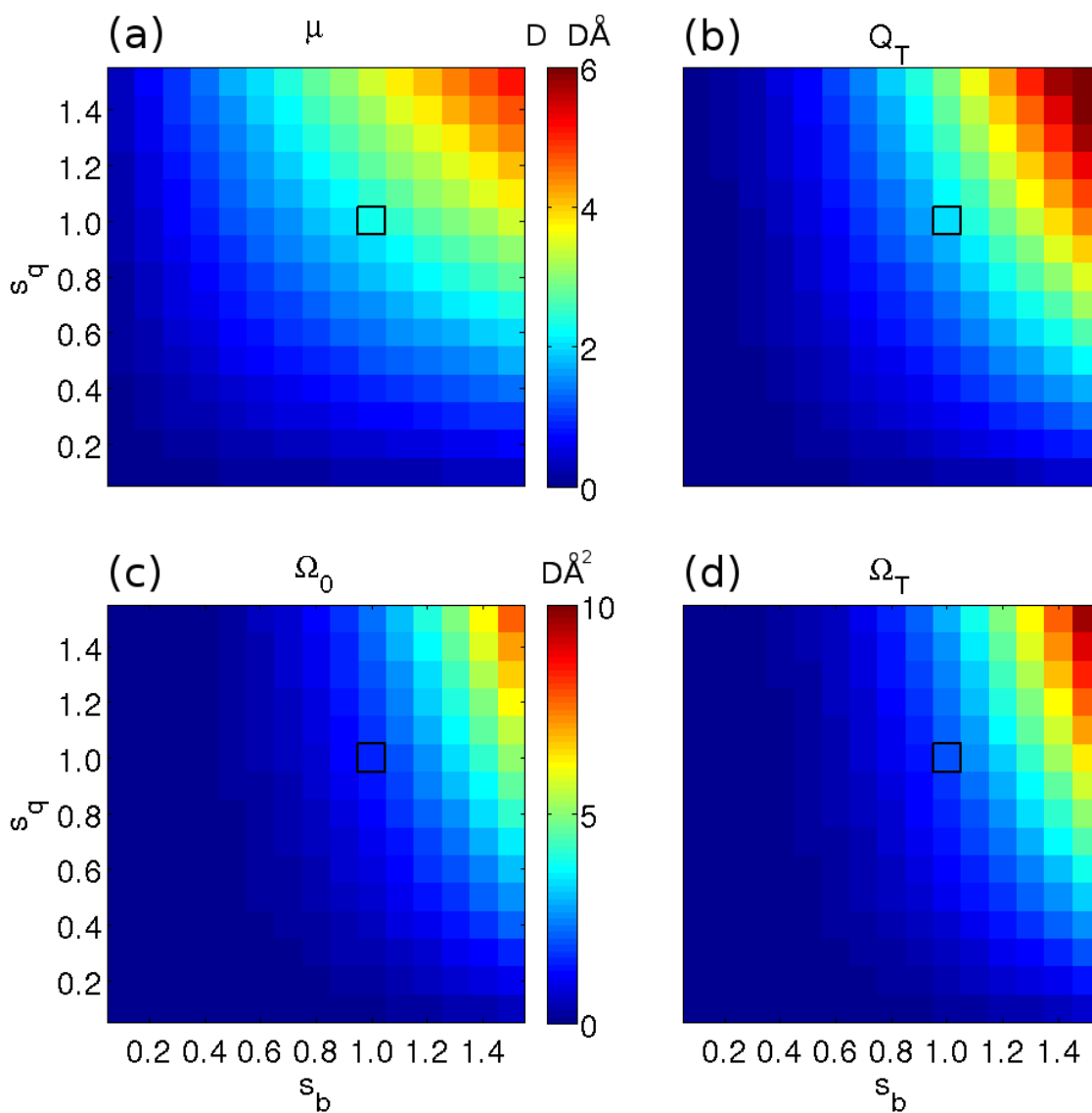**Figure 2.4:** Multipole moments as a function of the scaling factors s_b and s_q. In the upper panels, the dipole μ (a) and the square quadrupole Q_T (b) are shown. In the lower panels, the linear component of the octupole Ω_0 (c) and the square component of the octupole Ω_T (d) are shown. In all the maps, the model with $s_b = 1.0$ and $s_q = 1.0$ (SPC water) is highlighted.

2.3.4 Analysis of the liquid properties

Basic thermodynamic, structural and dynamic properties were analyzed for each of the 195 models of the NVT series and the 62 models of the NPT series. The results obtained are displayed in Figures 2.3-2.5 (thermodynamic properties), 2.6-2.9 (structural properties) and 2.10 (dynamic properties). All numerical values (including error estimates) can be found in Suppl. Mat. Tables 2.S.1 and 2.S.2. The properties corresponding to the SPC water model[146] are compared with experimental values in Table 2.2.

The change of the OH bond length and of the charges of the water molecule is altering the balance between repulsion and attraction between different molecules, as well as the strength and directionality of these interactions. As the collective properties of liquid water depend on this balance, changing these two molecular parameters is expected to create fluids with very different characteristics.

In the NVT series, the Lennard-Jones interactions between the molecules is not altered ($s_c = 1.0$). There, the behaviour of the models is expected to be dominated either by the

Properties	Exp.	Ref.	SPC
$\rho \; [\mathrm{kg \cdot m^{-3}}]$	997	[147]	973
$\Delta H_{vap} \; [\mathrm{kJ \cdot mol^{-1}}]$	44.05	[129]	43.9
$U^{pot} \; [\mathrm{kJ \cdot mol^{-1}}]$	−41.5	[164]	−41.6
$\mu \; [\mathrm{D}]$	1.855	[165]	2.27
$Q_T \; [\mathrm{D \cdot \mathring{A}}]$	2.57	[162]	−
ε	78.4	[166]	64.7
$\tau_D \; [\mathrm{ps}]$	8.3	[166]	6.9
$D \; [10^{-9}\mathrm{m^2 \cdot s^{-1}}]$	2.3	[167]	4.1
$\tau_1^{HH} \; [\mathrm{ps}]$	−		2.9
$\tau_1^{OH}[\mathrm{ps}]$	−		3.1
$\tau_2^{HH} \; [\mathrm{ps}]$	2.0	[151]	1.7
$\tau_2^{OH} \; [\mathrm{ps}]$	1.95	[168]	1.9

Table 2.2: Liquid properties of water at 298.15 K and 1 bar from experiment and from simulations using the SPC model (taken from Ref.[146]).

electrostatic interactions (high s_q and s_b) or the Lennard-Jones interactions (low s_q and s_b). In the first case the systems will be overstructured, possibly leading to a glassy state. Among the 195 NVT models, 41 indeed show characteristics typical for a glassy solid.

In the NPT series, the repulsion term $C_{12}^{1/2}$ of the Lennard-Jones interaction function is tuned ($s_c \neq 1.0$) in order to obtain densities comparable to the experimental value for liquid water. In this case the balance between electrostatics and Lennard-Jones interactions is largely restored, and the models are expected to show characteristics closer to those of liquid water.

In the following sections, the properties of the models of both the NVT and the NPT series are analyzed, focusing on the trends in these properties upon variation of the scaling factors s_b and s_q. Special attention is given to systems that show characteristics compatible with a liquid. The properties of the models that are in a glassy state are only described briefly.

Pressure P

For the NVT models, the pressure P of the systems was monitored. The results are displayed in Figure 2.3a. The value for the model $W_{1.0}^{1.0}$, which is equivalent to SPC water, is -71 bar. The slightly negative value is due to the small discrepancy between the density of the system simulated (968 kg m^{-3}) and the equilibrium density[146] of 973 kg m^{-1} for SPC water at 1 bar). To our knowledge ,the dependence of the pressure on the density was never systematically studied for SPC water, but values of the pressure at different densities are reported in Ref.[145], which were used for the calculation of the isothermal compressibility. The value of -71 bar obtained here at 968 kg m^{-3} leads to a compressibility of 54.0 10^{-6} bar^{-1}, which is compatible with the result reported for this model in Ref.[145]

The pressure shows a monotonic dependence upon variations of s_b and s_q. A decrease of

both scaling factors is increasing the pressure (up to about 10^4 bar), while an increase of both scaling factors is decreasing it (down to about $-3 \cdot 10^4$ bar). Decreasing or increasing one of the two scaling factors while keeping the other constant has about the same effect irrespective of the selected scaling factor, *i.e.* the map in Figure 2.3a is approximately symmetric around the diagonal. Decreasing one factor while equivalently increasing the other one keeps the value of the pressure P approximately constant, as long as s_q is not larger than 1.2. For larger values of s_q, a smaller decrease of s_b is needed in order to keep the pressure in the same range. Note that negative values of the pressure are perfectly acceptable physically. Based on equation 2.6, they merely indicate that the virial is larger than the kinetic energy, and that a p̈ulling f̈orce would have to be applied to the walls of the computational box to maintain the volume in a constant-pressure simulation. In practice, when the pressure is very negative ($P < -2 \ 10^{-3}$ bar), the statistical error calculated from the five different simulation repeats is larger than 100 bar (see Table 2.S.1) and the systems are typically found in a solid (glassy) state.

Total Potential Energy U

Figures 2.3c and 2.3d show the dependence of the potential energy U on s_b and s_q. Note that the scale is restricted to a range between -60 and 0 kJ·mol^{-1}, so as to focus on the characteristics of the liquid systems. The values of U for $W_{1.0}^{1.0}$ and $Y_{1.0}^{1.0}$ (SPC charges and bond lengths) are -41.5 kJ·mol^{-1} and -43.4 kJ·mol^{-1}, respectively. The value for the NVT model is in line with the experimental value and the typical value for SPC water[136] (see Table 2.2). The scaling of the repulsive Lennard-Jones parameter $C_{12}^{1/2}$ for the NPT model $Y_{1.0}^{1.0}$ ($s_c = 0.9735$) decreases the potential energy U by about 2 kJ·mol^{-1}. This effect is explained by the decreased repulsion parameter along with unaltered charges. The $Y_{1.0}^{1.0}$ water is denser than SPC water, so the potential energy is lower. The NVT models $W_{s_q}^{s_b}$ show a dependence of the total potential energy U on variations of s_b or s_q

analogous to that of the pressure P. The map in Figure 2.3c is also essentially symmetric around the diagonal. The decrease of both scaling factors simultaneously increases the potential energy up to values close to 0 kJ·mol^{-1} (*e.g.* -3.5 to -1.5 kJ·mol^{-1} for the models with scaling factor s_b or s_q equal to 0.1 or 0.2). The simultaneous increase of both scaling factors decreases the potential energy U to values smaller than -60 kJ·mol^{-1}. These systems correspond to the models with pressure P below -1000 bar and are in a solid (glassy) state. The lowest values of the potential energy U are around -950 kJ·mol^{-1}.

Static Relative Dielectric Permittivity ε

Figures 2.5a and 2.5b show the variations of the dielectric permittivity ε as a function of the two scaling factors s_b and s_q for the solvent of NVT and NPT series. The dielectric permittivity was calculated only for the systems considered liquid. In the case of solids, the fluctuations in the total dipole moment of the system lose their meaning and the Kirkwood-Fröhlich-type equation cannot be used. The value obtained for the $W_{1.0}^{1.0}$ model is 64.0 (in line with the standard value of SPC, see Table 2.2), while it increases to 68.1 for the $Y_{1.0}^{1.0}$ model.

In the NVT series, the maximum value obtained for the dielectric permittivity is 140.4 for the $W_{1.5}^{0.7}$ model. For the models with charges scaled by $s_q > 1.0$ and bonds scaled by $0.6 < s_b < 1.0$ (the value of s_b depending on scaling of the charges), the dielectric permittivity is larger than 65. A total of 13 NVT models have dielectric permittivities higher than that of SPC water. Note that models with high values of both scaling factors are solid, so that their dielectric permittivity is not considered here, meaning that the limit $s_b < 1.0$ for models with $s_q > 1.0$ is due to the glassy state of those systems. Lower values of ε, in a range from 20 to 64, are obtained for models with $s_q > 0.5$, with a minimum value of the scaling factor s_b depending on the scaling of the charges but always larger
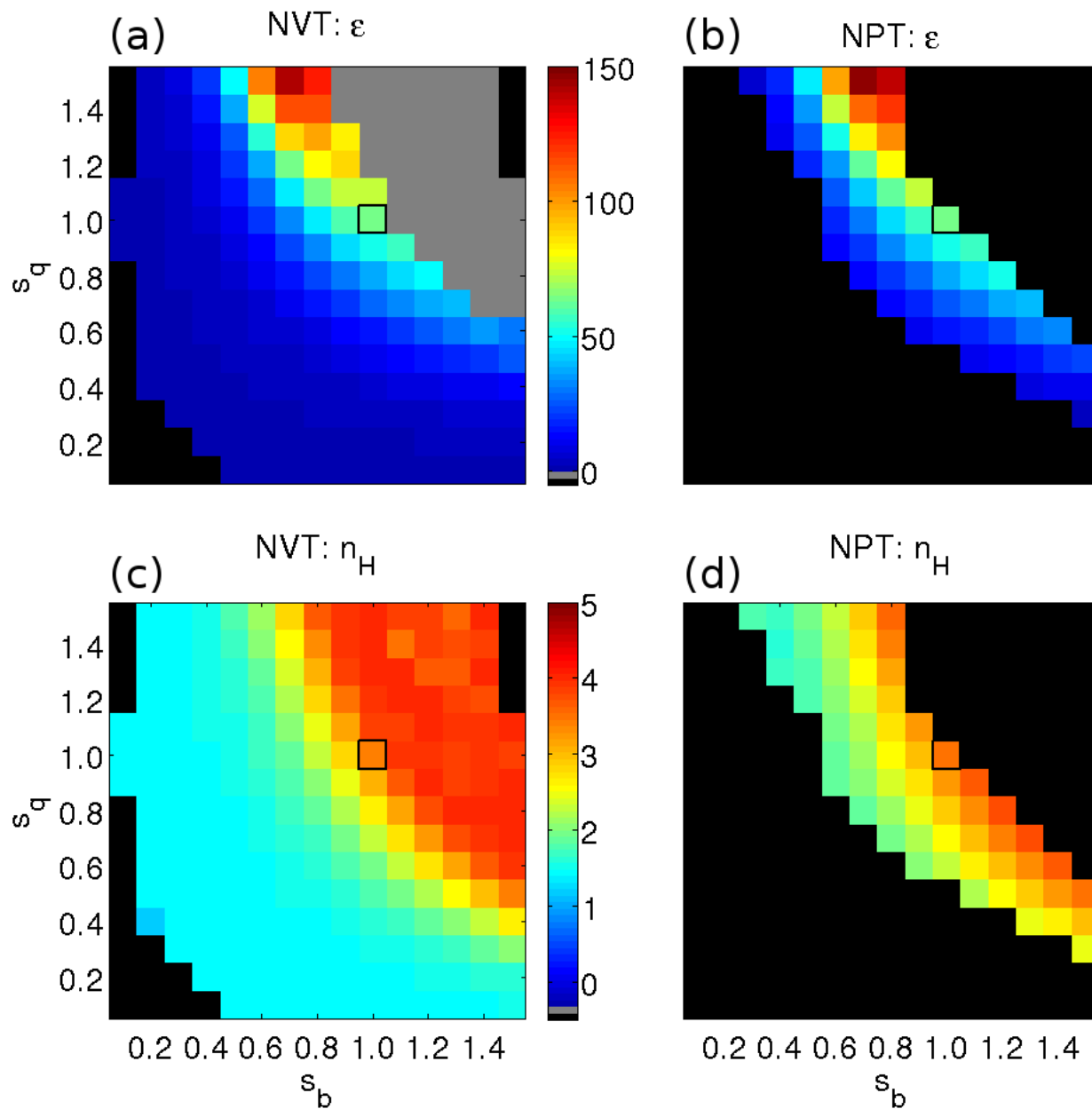
Figure 2.5: Upper panel: the dielectric permittivity ε as a function of the scaling factors s_b and s_q for the NVT models (a) and for the NPT models (b). Lower panels: number of the hydrogen bonds per molecule n_H as a function of the scaling factors s_b and s_q for the NVT models (c) and for the NPT models (d). In all the maps, the model with $s_b = 1.0$ and $s_q = 1.0$ (SPC water in the NVT series, closest one in the NPT series) is highlighted.

than 0.4. For models with $s_b < 0.4$ or $s_q < 0.4$, the dielectric permittivity is never higher than 20, independently of the other scaling factor. When both scaling factors are smaller than 0.5, ε is never higher than 2. The lowest value obtained is 1.

In the NPT series, the maximum value obtained for the dielectric permittivity is 147.5 for the $Y_{1.5}^{0.7}$ model. A total of 11 models have dielectric permittivities higher than that of SPC water, with similar combinations of s_b and s_q as in the NVT series. The NPT models with high charges and short bond lengths or low charges and long bond lengths have ε values smaller than 20. In this series, values of the dielectric permittivity lower than 7.3 cannot be found.

For both series, the dependence of the dielectric permittivity seems to be approximately symmetric with respect to the two scaling factors when these are smaller than 1.0. For higher factors, the behaviour is different depending on the factor that is changed. With high charges ($s_q > 1.0$), high values of the dielectric permittivity are obtained. Longer bond lengths seem unable to reach the same effect.

Hydrogen-bonding Capacity n_H

In Figures 2.5c and 2.5d, the average numbers of hydrogen bonds n_H per molecule are given as a measure of the hydrogen-bonding capacity of the models. The $W_{1.0}^{1.0}$ model and the $Y_{1.0}^{1.0}$ have $n_H = 3.4$ (as expected for SPC water) and $n_H = 3.5$, respectively. In the NVT series, 18 models have a n_H value equal to 4 (all of them in the solid state). For for the NPT series, the maximum value obtained is 3.7 (for the models $Y_{0.9}^{1.1}$, $Y_{0.8}^{1.2}$, $Y_{0.7}^{1.3}$, and $Y_{0.6}^{1.4}$). All the NVT models in a solid state have $n_H > 3.5$. Values of n_H smaller than 2 can be obtained with one of the two tuned parameters scaled to low values and the whole range of the other. The lowest values obtained are 1.4 for the NVT series and 1.7 for the NPT series. The dependence of n_H is again essentially symmetric relative to the two

scaling factors.

Radial Distribution Function $g(r)$

The overall structure of the systems was analyzed by calculating the radial distribution functions $g(r)$. In Figure 2.6, one illustrative example is shown for each of the two solvent series, considering the variation of the radial distribution function upon changing of the scaling factor s_b, while keeping $s_q = 1.0$ (panels (a) and (b) of the figure, $W_{1.0}^{s_b}$ and $Y_{1.0}^{s_b}$, respectively) or upon changing the scaling factor s_q, while keeping $s_b = 1.0$ (panels (c) and (d) of the figure, $W_{s_q}^{1.0}$ and $Y_{s_q}^{1.0}$, respectively). The radial distribution functions $g(r)$ on the top and on the bottom panels of each graph are those of the extreme cases accepted as liquid models. The trends in the position and height of the first $g(r)$ peak are also shown in Figure 2.7 for all the models of the two series.

For the NVT series and with $s_q = 1.0$, see Figure 2.6a, the extreme values of the scaling factor s_b considered for a liquid are 1.0 (normal SPC water) and 0.1. The $W_{1.0}^{1.0}$ model shows the typical radial distribution function of SPC water. The first relatively narrow peak is around 0.28 nm, with a height of 2.87, and the first minimum is at 0.31 nm. Also visible are a second and a third peak, broader than the first, at 0.45 nm and 0.68 nm, respectively. The function converges to a value of 1 in the long-distance limit. When the scaling factor s_b is decreased, the first peak becomes quickly (already for the model $W_{1.0}^{0.9}$) broader. Its height is reduced and it shifts to longer distances. A further decrease of the scaling factor s_b keeps shifting the position of the peak, up to about 0.32 nm for the $W_{1.0}^{0.1}$ model, close to the distance at the minimum of the Lennard-Jones interaction function for SPC water which is 0.35 nm. The height of this first peak initially decreases (from $s_b = 1.0$ to 0.6) but then increases again (from $s_b = 0.6$ to 0.1). The first minimum also becomes broader and shifts to the position of the second peak in the $g(r)$ of the SPC model, while
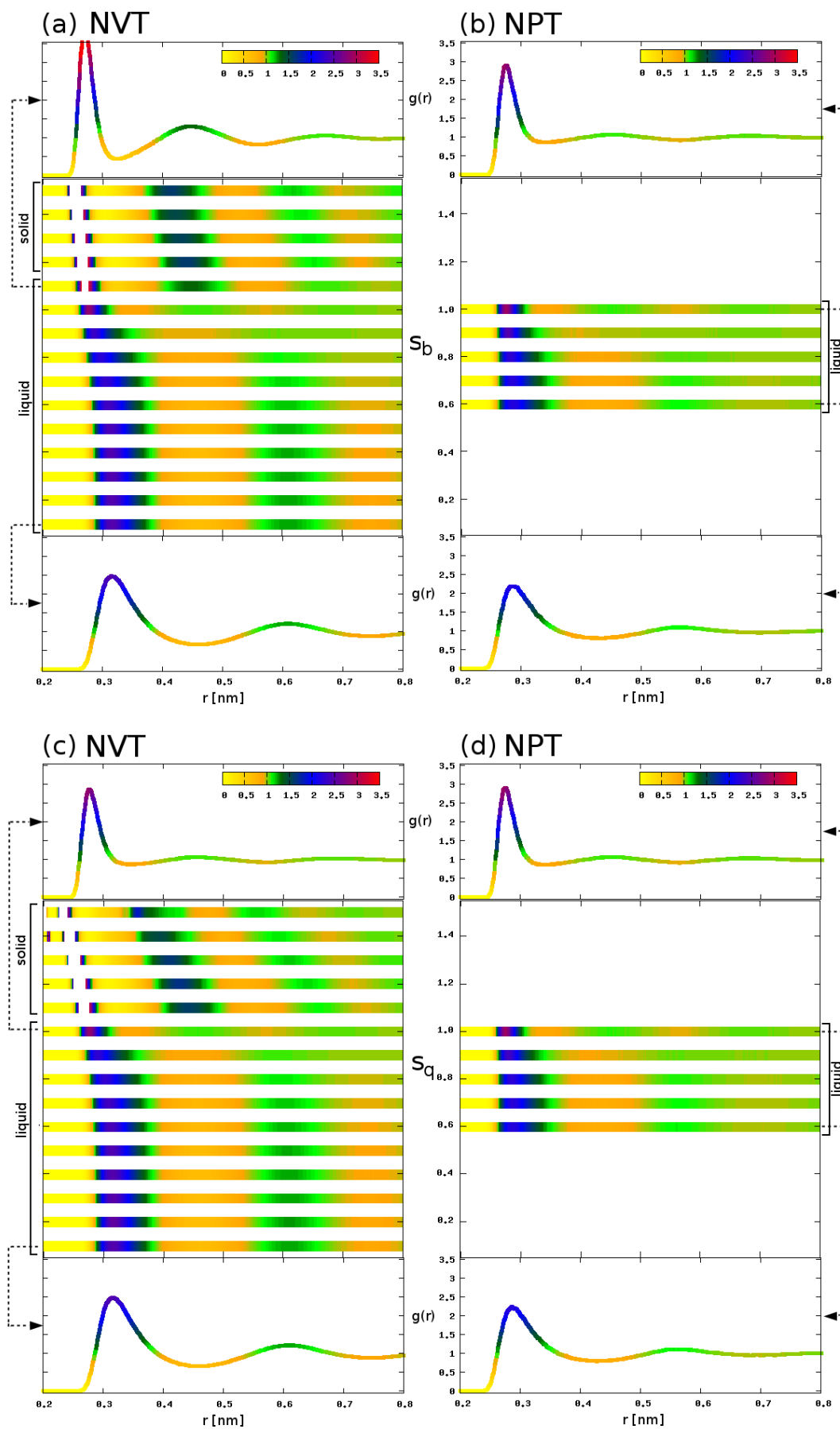
Figure 2.6: [Previous page] Upper panels: radial distribution function $g(r)$ for different values of the scaling factor s_b together with $s_q = 1.0$, for NVT simulations (a) and NPT simulations (b). In the central part, $g(r)$ is shown in the form of stripes. The upper and lower parts show the function $g(r)$ of $W_{1.0}^{1.0}$ and $Y_{1.0}^{1.1}$ (top), and $W_{1.0}^{0.1}$ and $Y_{1.0}^{0.6}$ (bottom). Lower panels: radial distribution function $g(r)$ for different values of the scaling factor s_q together with $s_b = 1.0$, for NVT simulations (c) and NPT simulations (d). In the central part, $g(r)$ is shown in the form of stripes. The upper and lower parts show the function $g(r)$ of $W_{1.1}^{1.0}$ and $Y_{1.0}^{1.1}$ (top), and $W_{1.0}^{0.1}$ and $Y_{1.0}^{0.6}$ (bottom).

Figure 2.7: Upper panels: position r_{max} of the first peak in the radial distribution function $g(r)$ as a function of the scaling factor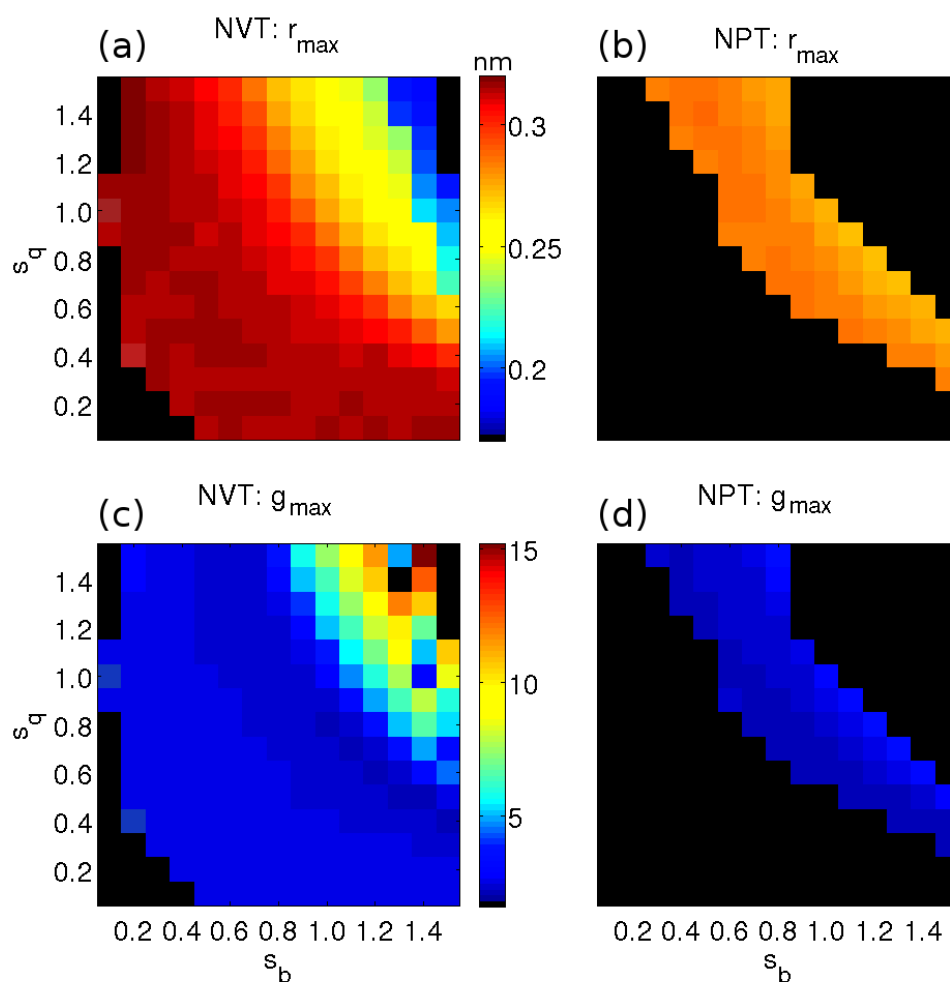s s_b and s_q for the NVT models (a) and the NPT models (b). Lower panels: the height g_{max} of the first peak in the radial distribution function $g(r)$ as a function of the scaling factors s_b and s_q for the NVT models (c) and the NPT models (d). In all the maps, the model with $s_b = 1.0$ and $s_q = 1.0$ (SPC water in the NVT series, closest one in the NPT series) is highlighted.

the second peak is now almost overlapping with the original third minimum and becomes broader and higher.

For the NVT series and with $s_b = 1.0$, see Figure 2.6c, the extreme values of the scaling factor s_q considered for a liquid are 1.1 and 0.1. The behaviour of the systems upon decreasing the scaling factor s_q is very similar to the one observed upon decreasing the scaling factor s_b. One observes a broadening and lowering of the first peak, followed by an increase, along with a shifting to larger distances. Simultaneously, one sees a broadening of the first minimum. The second and third peaks become more evident, broader and more defined, equally shifted to larger distances. Figures 2.6a and 2.6c also show cases characteristic of a glassy state ($s_q = 1.0$ with $s_b > 1.0$ or $s_b = 1.0$ with $s_q > 1.1$). The general behaviour is again similar for both scaling factors. Upon increasing s_b and s_q, the system becomes in this case overstructured, as a consequence of the change of state. The first peak becomes sharper and higher, and its position is shifted, this time to lower distances. The first minimum is lower and shifted. The second and third peaks are again sharper and higher, and shifted to lower distances. A fourth peak appears. The increase of the scaling factor s_b shows a much more pronounced shift (for the model $W_{1.0}^{1.5}$ the position of the first minimum is 0.20 nm, compared with 0.25 nm for the model $W_{1.5}^{1.0}$). Additionally, for the models with the two largest factors $s_b = 1.4$ and $s_b = 1.5$, a double first peak appears, and the following peak is destructured. This is because the system becomes inhomogeneous, including clusters, aggregates and bubbles. This behaviour is not seen for the largest values of the scaling factor s_q, where the glassy system remains homogeneous.

The trends observed and the considerations made for the NVT systems are essentially valid also for the NPT series, see Figure 2.6b and 2.6d. The structure of the $Y_{1.0}^{1.0}$ model is the same as that of $W_{1.0}^{1.0}$. The decrease of both s_b and s_q is progressively broadening the first peak, while shifting it. The second and the third peaks start to become less defined.

Compared to the NVT models, the decrease of the two scaling factors is not having the effect of initially lowering and then increasing the height of the first peak. The trend in this case shows a clear monotonic decrease.

Dipole-dipole orientation correlation function $c(r)$

The dipole-dipole orientation correlation function $c(r)$ gives a measure of the alignment of the water molecules as a function of distance. Figure 2.8 shows one illustrative example for each of the two solvent series, considering the variation of this function upon changing the scaling factor s_b, while keeping $s_q = 1.0$ (panels (a) and (b) of the figure, $W^{s_b}_{1.0}$ and $Y^{s_b}_{1.0}$, respectively) and upon changing the scaling factor s_q, while keeping $s_b = 1.0$ (panels (c) and (d) of the figure, $W^{1.0}_{s_q}$ and $Y^{1.0}_{s_q}$, respectively). The dipole-dipole orientation correlation functions $c(r)$ on the top and on the bottom panels of each graph represent the extreme cases accepted as liquid models. The trend in the position and the height of the first $c(r)$ peak are also shown in Figure 2.9 for all the models of the two series.

For the NVT series with $s_q = 1.0$, see Figure 2.8a, the extreme values of the scaling factor s_b considered for a liquid are 1.0 (normal SPC water) and 0.1. The $W^{1.0}_{1.0}$ model shows the typical dipole-dipole orientation correlation function of SPC water. The first neighbors of a water molecule strongly favor an alignment of the dipoles with the central molecule, as seen from the first correlation peak at 0.27 nm. The first minimum around 0.4 nm is slightly negative and the second peak follows at around 0.5 nm. The correlation is in general positive almost everywhere and levels off towards zero at long distances, showing the random orientation at long range. Decreasing the scaling factor s_b has initially the effect of broadening and lowering the first peak, sign of a weaker correlation of the first neighbors. The maximum width of the peak is reached with $s_b = 0.8$, then it starts to become narrower and lower, until it almost completely disappears for $s_b < 0.4$. The position of the first
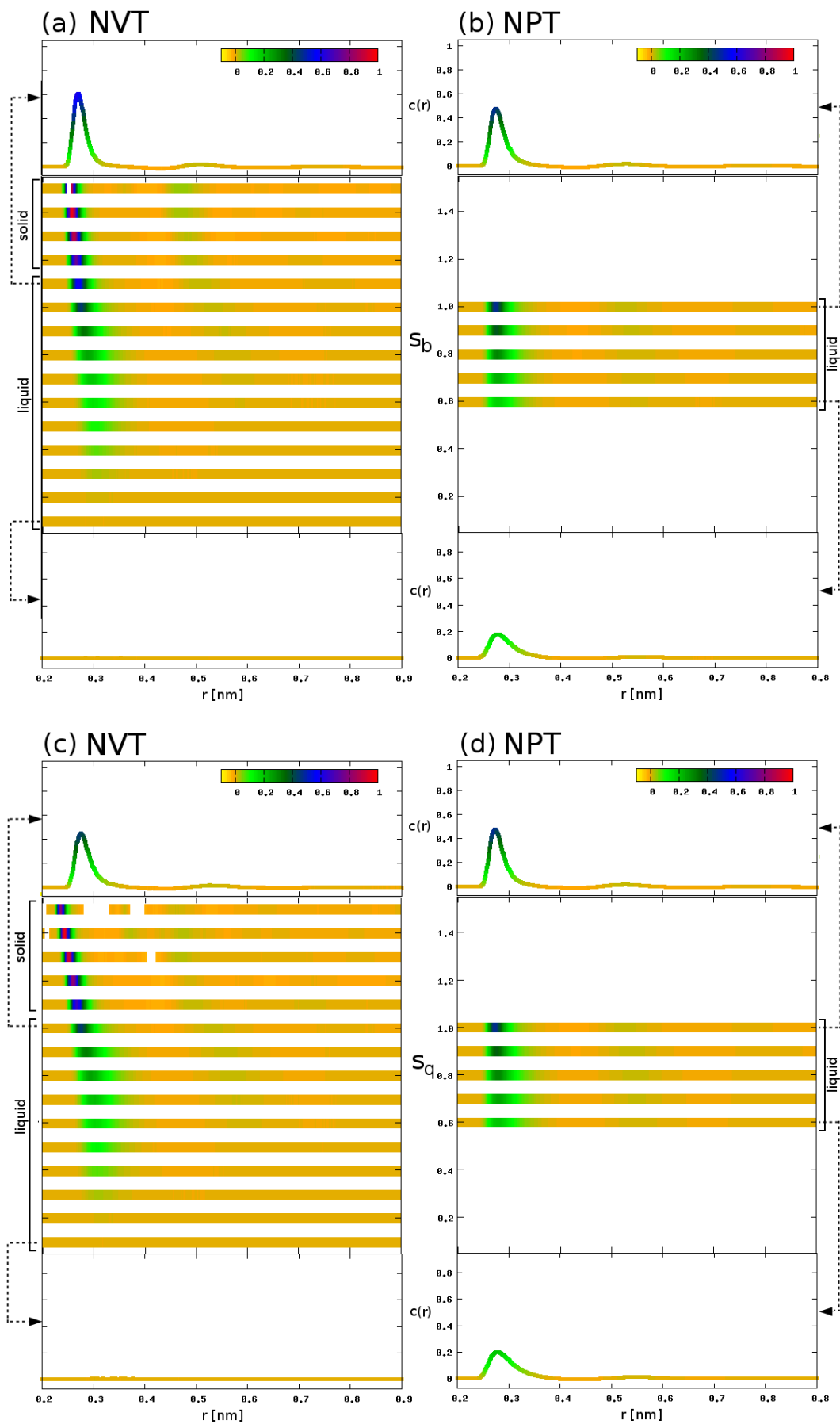
Figure 2.8: Upper panels: dipole-dipole orientation function $c(r)$ for different values of the scaling factor s_b together with $s_q = 1.0$, for NVT simulations (a) and NPT simulations (b). In the central part, $c(r)$ is shown in the form of stripes. The upper and lower parts show the function $c(r)$ of $W_{1.0}^{1.0}$ and $Y_{1.0}^{1.1}$ (top), and $W_{1.0}^{0.1}$ and $Y_{1.0}^{0.6}$ (bottom). Lower panels: dipole-dipole orientation function $c(r)$ for different values of the scaling factor s_q together with $s_b = 1.0$, for NVT simulations (c) and NPT simulations (d). In the central part, $c(r)$ is shown in the form of stripes. The upper and lower parts show the function $c(r)$ of $W_{1.1}^{1.0}$ and $Y_{1.0}^{1.1}$ (top), and $W_{1.0}^{0.1}$ and $Y_{1.0}^{0.6}$ (bottom).
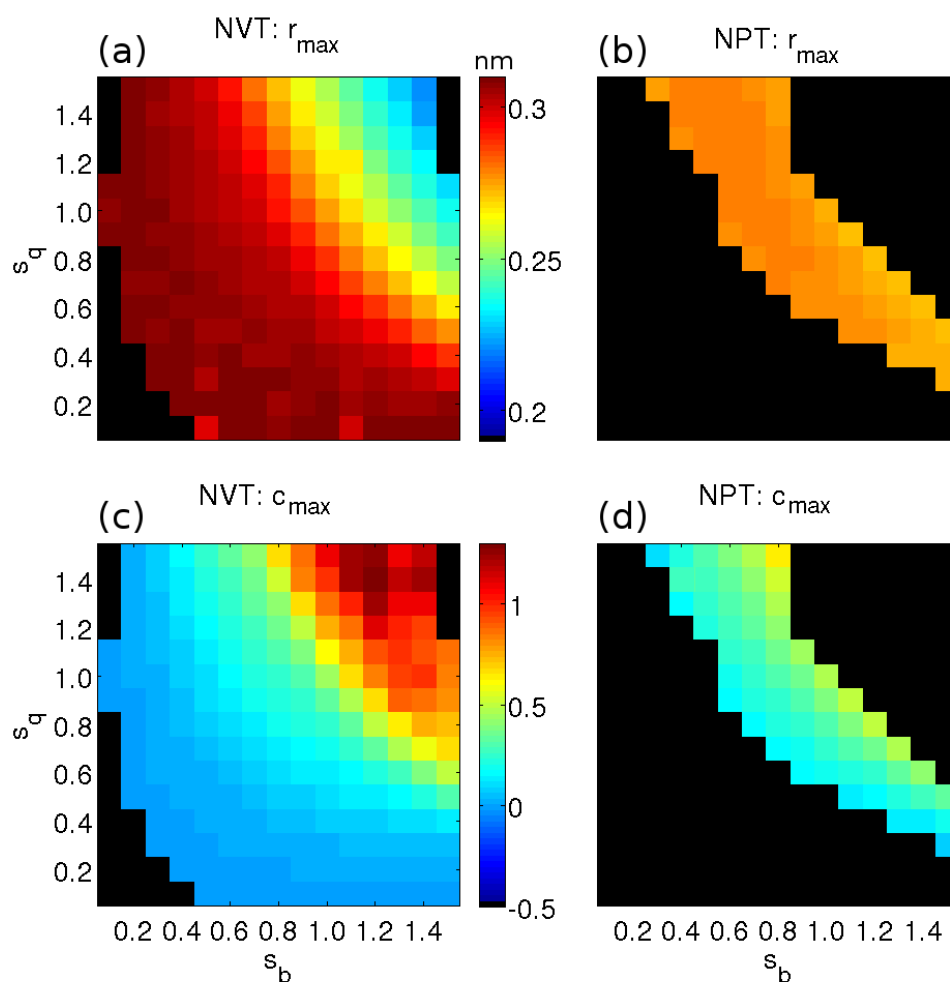
Figure 2.9: [Previous page] Upper panels: position r_{max} of the first peak in the dipole-dipole orientation function $c(r)$ as a function of the scaling factors s_b and s_q for the NVT models (a) and the NPT models (b). Lower panels: height c_{max} of the first peak in the dipole-dipole orientation function $c(r)$ as a function of the scaling factors s_b and s_q for the NVT models (c) and the NPT models (d). In all the maps, the model with $s_b = 1.0$ and $s_q = 1.0$ (SPC water in the NVT series, closest one in the NPT series) is highlighted.

peak also varies upon decreasing the scaling factor, shifting to long distances following
the trend of the radial distribution function $g(r)$, sign that the major contribution to the
dipole alignment comes from the first neighbors. Models with the lowest values for s_b
completely lose the preference for a specific alignment of the dipoles (random orientation),
as expected with a decrease of the importance of the electrostatic interaction. For these
models, the Lennard-Jones contribution is dominating and the correlation of the dipole-
dipole orientation is almost zero everywhere. When the scaling factor s_b is increased above
$s_b = 1.0$, the systems become glassy. The first peak becomes narrower and higher, and
shifts to shorter distances, indicating a higher correlation between the dipoles, consequence
of the overstructuring of the models due to the transition to a solid state. A second peak
starts to be evident, increasing with the increase of the scaling factor, further sign of the
change of state. The last two models ($s_b = 1.4$ and $s_b = 1.5$) show in the opposite a very
peculiar behaviour. A fist negative peak appears at around 0.2 nm, followed by a second
positive peak at similar distances compared with the first peaks of the models with smaller
scaling factors. A third and a fourth peak appear, even though not very defined, following
minima again with negative values. These two models display a double first peak in the
radial distribution function $g(r)$, corresponding to the same positions as the first negative
peaks in $c(r)$. The following peaks and their distorted shapes indicate an inhomogeneity in
the systems. These solids are forming aggregates and clusters, with empty areas (bubbles).

A similar behaviour is seen in case of the decrease of the scaling factor s_q with $s_b = 1.0$,
see Figure 2.8c. In this case the broadening is less pronounced, as well as the shift of the
first peak to longer distances, but the essential trends are the same. When the scaling
factor s_q is increased above 1.1 (glassy systems), the dipole-dipole orientation correlation
function $c(r)$ shows at first a trend very similar to the one observed with an increase of
the scaling factor s_b. The first peak becomes again narrow and higher, shifting to shorter
distances, and the second peak becomes more and more evident. The models with higher

charges are not showing the same peculiar characteristics of the models with the longest bond lengths. The increase of the charge values is not leading to inhomogeneity in the system.

In the NPT series, the Lennard-Jones interactions are tuned to obtain a density close to the experimental density of liquid water. In this case, the equilibrium between the electrostatic and the Lennard-Jones interactions is somewhat restored. Reflecting this balance, the dipole-dipole correlation functions of the NPT models have very similar shapes for all the models, see Figures 2.8b and 2.8d. In particular, the position of the first peak is in the same range (0.27-0.28 nm). The height of the first peak shows more pronounced variations, with values that vary between 0.07 and 0.53, increasing mainly with the increase of the scaling factor s_q, and to a lower extent with the increase of the scaling factor s_b. As expected, the variation of the charges has a stronger effect on the orientation of the dipoles in the system than that of the bond length. Additionally, it seems that the enhancement of the Lennard-Jones interaction (less repulsive $C_{12}^{1/2}$) is not sufficient to restore completely the influence of the electrostatics when too small charges are used.

Self-Diffusion Coefficient D

Figures 2.10a and 2.10b show the diffusion constant D as a function of the scaling factors s_b and s_q. The diffusion constant for the $W_{1.0}^{1.0}$ model is $4.17 \cdot 10^{-9}$ m$^2 \cdot$s^{-1} (typical value for SPC water, see Table 2.2), while the slight change in the Lennard-Jones repulsion is decreasing the value of the $Y_{1.0}^{1.0}$ model to $3.84 \cdot 10^{-9}$ m$^2 \cdot$s^{-1}. Reducing the Lennard-Jones repulsive term allows for the molecules to get closer and it increases the influence of the electrostatic, making the system less diffusive.

The only 8 systems with a diffusion coefficient lower than that of SPC but still in a reasonable liquid-like range (0.2-3.1$\cdot 10^{-9}$ m$^2 \cdot$s^{-1}) are $W_{0.6}^{1.4}$, $W_{0.7}^{1.3}$, $W_{0.8}^{1.2}$, $W_{0.9}^{1.1}$ (models with

$s_b > 1.0$ and $s_q < 1.0$) and $W^{1.1}_{1.0}$, $W^{0.9}_{1.2}$, $W^{0.9}_{1.3}$, $W^{0.8}_{1.5}$ (models with $s_b < 1.0$ and $s_q > 1.0$). Decreasing one of the two scaling factors while keeping the other constant or both factors simultaneously has the same effect independently of which parameter is reduced. The diffusion constant initially increases to values between $10 \cdot 10^{-9}$ m$^2 \cdot$s^{-1} and $11.8 \cdot 10^{-9}$ m$^2 \cdot$s^{-1} (for $W^{0.8}_{1.3}$), but then decreases to values between $7 \cdot 10^{-9}$ m$^2 \cdot$s^{-1} and $10 \cdot 10^{-9}$ m$^2 \cdot$s^{-1}. Values lower than $7 \cdot 10^{-9}$ m$^2 \cdot$s^{-1}, but still higher than the one of SPC water, are possible with $s_b < 0.3$ and $s_q > 1.0$.

When the scaling factors are both very high, the diffusion coefficient D of the models is decreasing to values close to zero. This, combined with the considerations on the properties analyzed before, confirms that these systems are in a solid (glassy) state.

For the NPT series, the behaviour is similar. In this case, models with both scaling factors simultaneously very low or very high were not feasible. The decrease of one of the two scaling factor while keeping the other constant or of both factors simultaneously increases the diffusion constant, initially to values between $10 \cdot 10^{-9}$ m$^2 \cdot$s^{-1} and $15 \cdot 10^{-9}$ m$^2 \cdot$s^{-1}, and further up to $20 \cdot 10^{-9}$ m$^2 \cdot$s^{-1}. A further decrease seems to evidence a lowering of the diffusion constant similar to the NVT models.

Rotational Correlation Time τ_2^{OH}

Two different rotational correlation times were calculated, namely τ_2^{OH} corresponding to the OH vector along the bonds of the molecules and τ_2^{HH} corresponding to the vector HH connecting the two hydrogen atoms. Only τ_2^{OH} is displayed in Figures 2.10c and 2.10d and will be discussed here. The other correlation time τ_2^{HH} presents similar trends (data not shown).

The value of τ_2^{OH} for the $W^{1.0}_{1.0}$ model is 1.4 ps, lower than what is found in the literature for SPC water (see Table 2.2). The $Y^{1.0}_{1.0}$ model shows a rotational correlation time of 1.8
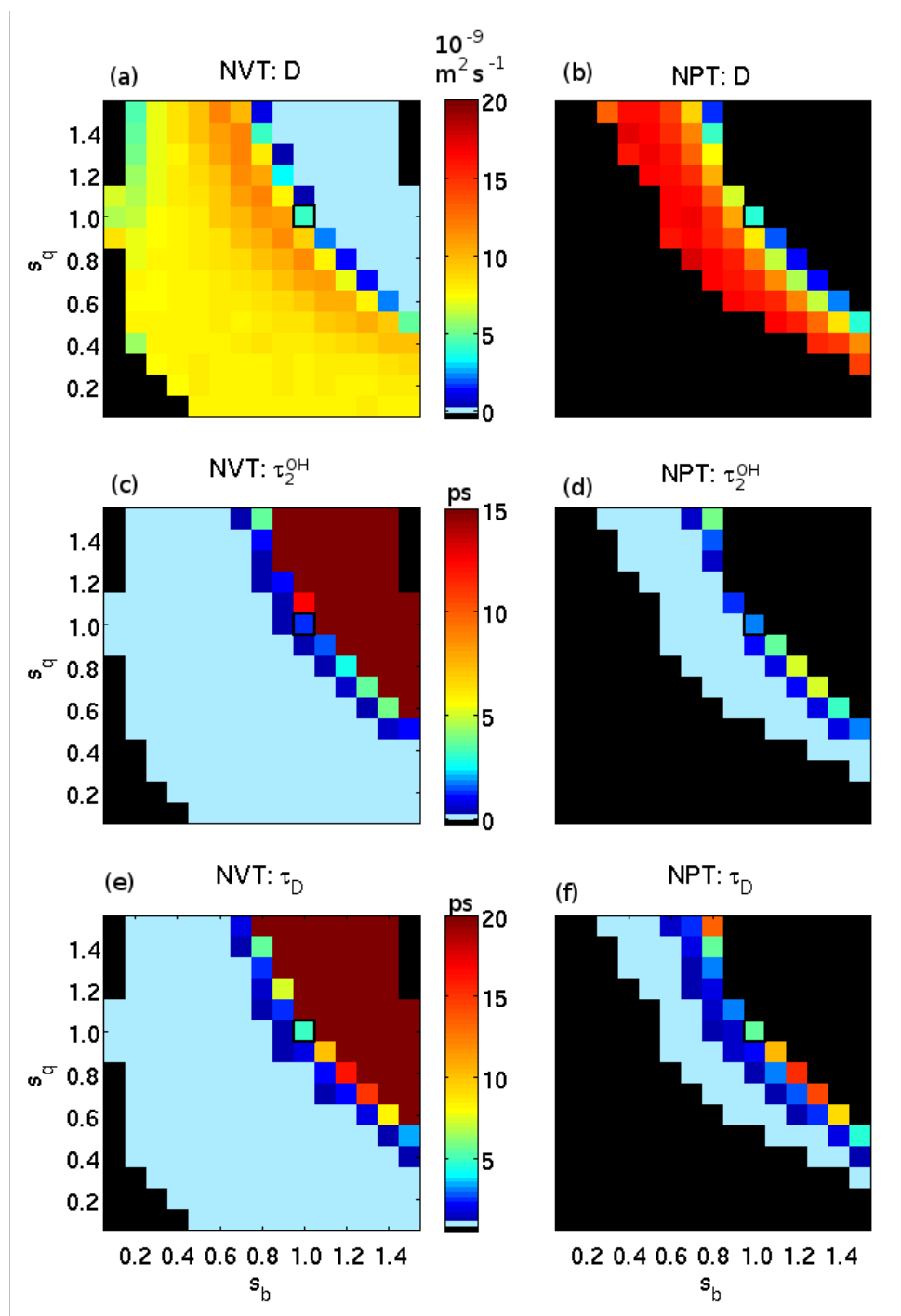
Figure 2.10: Upper panels: the diffusion constant D as a function of the scaling factors s_b and s_q for the NVT models (a) and the NPT models (b). Central panel: the rotational correlation time τ_2^{OH} as a function of the scaling factors s_b and s_q for the NVT models (c) and the NPT models (d). Lower panels: the Debye relaxation time τ_D as a function of the scaling factors s_b and s_q for the NVT models (e) and the NPT models (f). In all the maps, the model with $s_b = 1.0$ and $s_q = 1.0$ (SPC water in the NVT series, closest one in the NPT series) is highlighted.

ps. In the NVT series, an increase of the scaling factors is immediately increasing the rotational time. In some cases, the autocorrelation function is not decorrelating to 0 even after nanoseconds. When this happens, the rotational time is reported as > 21 ns. This behaviour is in line with the other properties calculated, giving further confirmation of the change of state of these models. For the other models which remain liquid, three cases can be found. In the first case, the autocorrelation function is relaxing so fast that it is already close to zero within 0.2 ps or less and assumes immediately a damped oscillatory behaviour. For these systems, the rotational correlation time is reported as $\tau < 0.2$ ps. In the second case, 5 models have a rotational correlation time longer than SPC water, namely $W_{0.6}^{1.4}$ (3.9 ps), $W_{0.7}^{1.3}$ (3.7 ps), $W_{0.8}^{1.2}$ (2.7 ps), $W_{1.1}^{1.0}$ (12.6 ps) and $W_{1.5}^{0.8}$ (3.6 ps). These models show a diffusion constant similar to the one of SPC water. In the third case, models have a rotational correlation time shorter than SPC water.

A very similar behaviour is observed for the NPT series, with the NPT models corresponding to a given NVT model having similar characteristics. For this series, the slowest model has a rotational correlation time of 5.1 ps ($Y_{0.8}^{1.2}$), still much shorter than the slowest NVT model (12.6 ps).

Note that the values of the rotational times are affected by very high statistical uncertainties, in particular because the fitting procedure of the autocorrelation function is prone to depend on arbitrary choices. For example, a change in the time interval considered for the exponential fit, keeping the same criterion for the coefficient of determination ($R^2 > 0.9$), can yield a variation in the rotational time of up to 50%. For this reason, the values reported for the rotational times have to be considered indicative and the trends qualitative.

Debye Relaxation Time τ_D

Figures 2.10e and 2.10f show the Debye relaxation time as a function of the scaling factors s_b and s_q. The value for the $W_{1.0}^{1.0}$ model is lower than expected for SPC water (5.1 ps instead of 6 ps). Also the value for the unmodified NPT model $Y_{1.0}^{1.0}$ is lower than the one of SPC water (5.6 ps instead of 6 ps). The Debye relaxation times shorter than 1 ps were not considered. In the NVT series, the same 5 models that evidenced a rotational time longer than SPC water, plus two additional ones ($W_{1.1}^{0.9}$ and $W_{1.4}^{0.8}$), have a Debye relaxation time longer than that of SPC water. The slowest model has a relaxation time of 24.7 ps ($W_{1.5}^{0.8}$). In the NPT series, the corresponding combinations of scaling factors s_b and s_q (the ones that led to feasible models) show the same slowed dynamics. In this case, the slowest model has a relaxation time of 15.2 ps ($Y_{0.8}^{1.2}$).

The same considerations made in the context of the rotational correlation times (previous section) are valid also for the Debye relaxation time. This is understandable as the latter relaxation is a collective variant of the former single-molecule one.

The analysis of the dynamic properties of the systems confirms what is suggested by the thermodynamic and structural analysis. In the NVT series, 41 models (large scaling factors s_b and s_q) are solid. Additionally, the dynamic properties show that the NVT models with scaling factors s_b and s_q smaller than 1.0 are faster than SPC water, in terms of both translational and rotational diffusion.

State of the systems

Considering the analysis of all the properties, it is clear that, when one of the two scaling factors is varied (modification of bond length or charges of the model), the balance between electrostatic interaction and Lennard-Jones repulsion that is responsible for the properties of water is disrupted.

A small reduction of either charges or bond lengths (s_q or s_b between 0.5 and 1.0) shift the equilibrium in favor of the Lennard-Jones repulsion, destructuring the systems. The effect is consistent for all properties. When these models are parametrized to have the same density as water in the NPT series, the Lennard-Jones repulsion has to be reduced ($s_c <$ 1.0), restoring the balance to a large extent. With a further reduction of the scaling factors (s_q or s_b smaller than 0.5), the system is completely dominated by the Lennard-Jones interactions and becomes a more structured. These Lennard-Jones liquid-like systems have high pressure, higher potential energy, a diffusion constant between $5 \cdot 10^{-9}$ m^{2}s^{-1} and $10 \cdot 10^{-9}$ m^{2}s^{-1} and very short rotational correlation times. When tentatively adjusted to the NPT series, these kind of models have a density higher than that of water, so that it is not possible to find an appropriate scaling factor s_c to tune the Lennard-Jones repulsion term for achieving the experimental density of liquid water.

An increase of the scaling factors s_b and s_q overstructures the systems, leading to a change of state from liquid to solid. In this case the electrostatic interactions become predominant. Increasing s_q generates a more homogeneous structure, whereas increasing of s_b creates a more disordered and inhomogeneous state. Also in this case, tuning the Lennard-Jones repulsive term is not successfully reproducing the density of a liquid water-like system.

2.3.5   Correlation between the properties

The linear correlation between the different properties analyzed in the previous sections is displayed in Figure 2.11a as a correlation matrix involving all the possible pairwise combinations. Dark red is indicative of a strong linear correlation, dark blue of a strong linear anti-correlation, while colors in the range yellow-green indicate no correlation between the properties. As linear correlation is only one possible type of correlation, correlation plots

for specific pairs of properties are also explicitly shown in Figure 2.11b.

Most of the properties turn out to be either highly linearly correlated or highly linearly anti-correlated with each others. The most remarkable features are the following:

- A strong correlation of pressure P and total potential energy U, along with strong correlation or anticorrelation between these two properties and all the others (except D).

- A strong correlation of the dielectric permittivity ε with the dipole moment μ and, to a lesser extent, the quadrupole moment Q_T.

- A strong correlation of the hydrogen-bonding capacity n_H with the dipole moment μ and, to a larger extent the square quadrupole moment Q_T and the square component of the octupole moment Ω_T.

- A relatively low correlation (0.74) between dielectric permittivity ε and the hydrogen-bonding capacity n_H.

- A relatively low correlation (below 0.65) of the dynamic properties (diffusion coefficient D, rotational correlation time τ_2^{OH} and the Debye relaxation time τ_D) with the other properties.

The correlation or anticorrelation of the thermodynamic properties P and U with the others is not entirely surprising. These properties define the state of the system and the interactions governing its behaviour. The correlation of the hydrogen-bonding capacity with the higher-order multipoles can explain some interesting characteristics of water and their dependence on the model as reported in literature, *e.g.* the dependence of the freezing point of different water models on their quadrupole moment [160].

The relatively low correlation between dielectric permittivity ε and the hydrogen-bonding capacity n_H is an interesting and useful observation concerning these water-like artificial
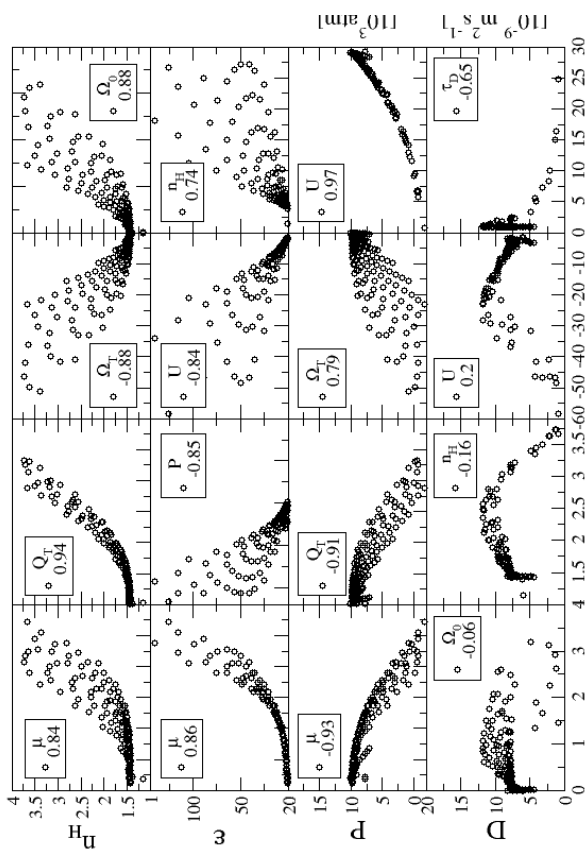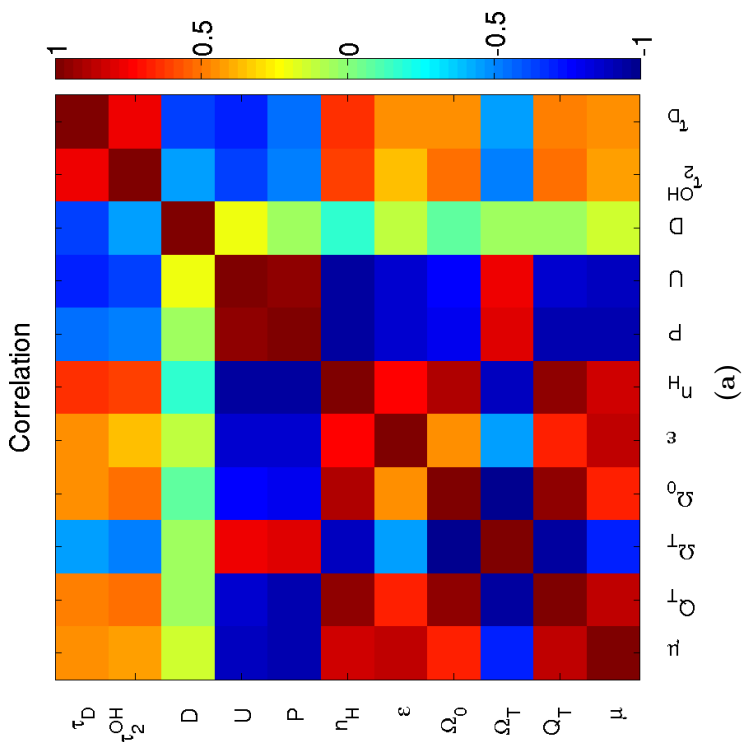
models. It offers the possibility to engineer models with almost uncorrelated values of the two properties, allowing for studies on the influence of the polarity of the solvent which can disentangle of the two aspects

The behaviour of the diffusion coefficient D is somewhat anomalous, seeming not to correlate with any of the other properties, except for a weak anticorrelation with the rotational correlation time τ_2^{OH} and the Debye relaxation time τ_D. A closer look at this property (Figure 2.11b) shows that the diffusion coefficient analyzed as a function *e.g.* of the hydrogen-bonding capacity n_H or the total potential energy U displays a double regime. This is easily explained based on the above considerations concerning the state of the systems. The first regime, characterized by an anticorrelation with n_H at high values of n_H and a correlation with U at low values of U, corresponds to systems for which either the charges or the bond lengths are slightly reduced. For these models the electrostatic interactions are less important than for normal SPC water. The second regime, characterized by opposite correlations between diffusion coefficient D and hydrogen-bonding capacity n_H or total potential energy U, includes models that behave like Lennard-Jones liquids. The turning point is given by the maximum in the diffusion coefficient (about 12

Figure 2.11: [Previous page] (a) Linear correlation matrix obtained considering all possible pairwise combinations of the calculated properties. The properties are the dipole μ, the square component of the quadrupole Q_T, the square component of the octupole Ω_T, the linear component of the octupole Ω_0, the dielectric permittivity ε, the hydrogen-bonding capacity n_H, the pressure P, the total potential energy per molecule U, the self-diffusion constant D, the rotational correlation time for the OH vector τ_2^{OH} and the Debye relaxation time τ_D. The correlation coefficients are calculated as Pearson's coefficient (see Section 2.2.6). For the NVT solvents series, only models that are in a liquid state are considered. A corresponding correlation matrix for the NPT models is given in Figure 2.12. (b) Correlation plots for specific pairs of properties considered. The hydrogen-bonding capacity n_H (first row from the top) is displayed as a function of all the multipole moments. The dielectric permittivity ε is displayed as a function of μ, P, U or n_H. The pressure P is displayed as a function of μ, Q_T, Ω_T or U. The diffusion coefficient D is displayed as a function of Ω_0, n_H, U or τ_D. For each plot, the value of the corresponding linear correlation coefficient is also indicated.

10^{-9} m$^2 \cdot$s^{-1}).
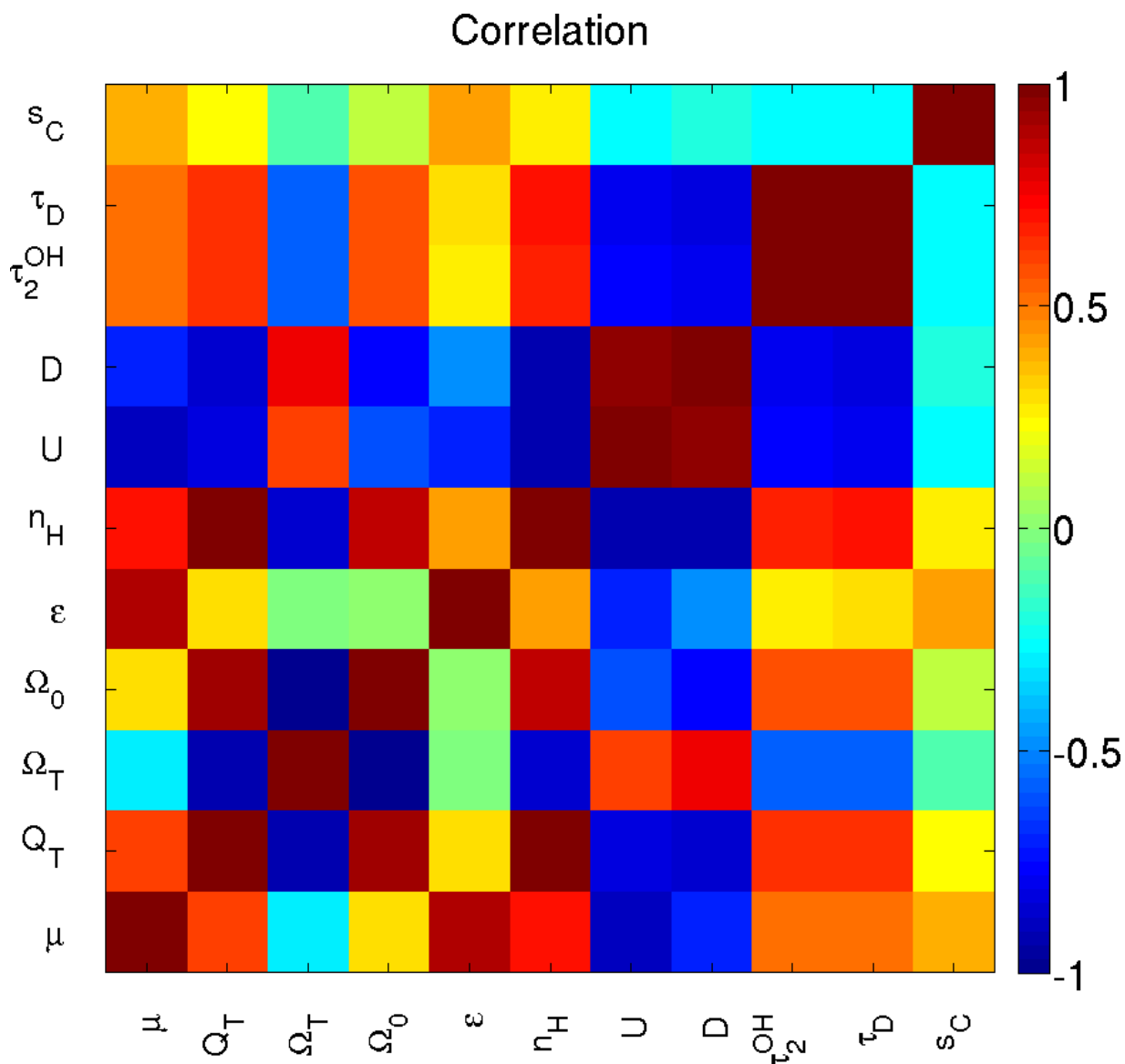
Figure 2.12: Linear correlation matrix obtained considering all possible pairwise combinations of the calculated properties for the NPT models. The properties are the dipole μ, the square component of the quadrupole Q_T, the square component of the octupole Ω_T, the linear component of the octupole Ω_0, the dielectric permittivity ε, the hydrogen-bonding capacity n_H, the total potential energy per molecule U, the self-diffusion constant D, the rotational correlation time for the OH vector τ_2^{OH}, the Debye relaxation time τ_D and the scaling factor s_c. The correlation coefficients are calculated as Pearson's coefficient (see Section 2.2.6).

2.3.6 Solvation free energy

In Figure 2.13, The solvation free energy of argon, water and a sodium ion is shown as a function of the solvents considered in the four series described in Section 2.2.7. In the figure, the main properties of the solvents, dielectric permittivity, hydrogen-bonding capacity and scaling factor s_c, are displayed in the first and the second panel of the figure from the top. The values are also reported in Table 2.1.

The results for the corresponding simulations in SPC water (horizontal dashed line in the panels) show that the sodium ion Na^+ (third panel from bottom) is highly soluble in water (highly negative free energy). For this species, the interaction with the solvent is entirely dominated by the electrostatic interactions. The variation in the free energy of solvation along the four series of solvents considered is strongly dependent on the variation of the dielectric permittivity of the solvent, even in the two series S_h^P and S_h^p, for which the hydrogen-bonding capacity is varied. The reduction of ε is increasing the ΔF, making the compound less soluble.

In the opposite, argon Ar (top panel) is poorly soluble in water (positive free energy). Here, the solvation free energy is completely correlated with the change in the scaling factor s_c that is adjusting the Lennard-Jones repulsion. The solvation free energy varies following the variation of the scaling factor s_c, in particular for the series S_p^h and S_h^p. A decrease in the Lennard-Jones repulsion (smaller s_c) decreases the free energy of solvation of the hydrophobic compound, making it less insoluble.

Water (middle panel between top and bottom) is a soluble molecule, even if the free energy value in this case is less negative than the one for the ionic species. A decrease of the permittivity of the solvent at hydrogen-bonding capacity similar to water is decreasing the solvation free energy of this molecule, as does a decrease in the hydrogen-bonding capacity at a permittivity similar to the one of water. When one of the two properties

is reduced, a variation of the other does not have a large influence on the solvation free energy of this species. Water solvation appears to be a cooperation between the specific H-bond network and the permittivity of the solvent.
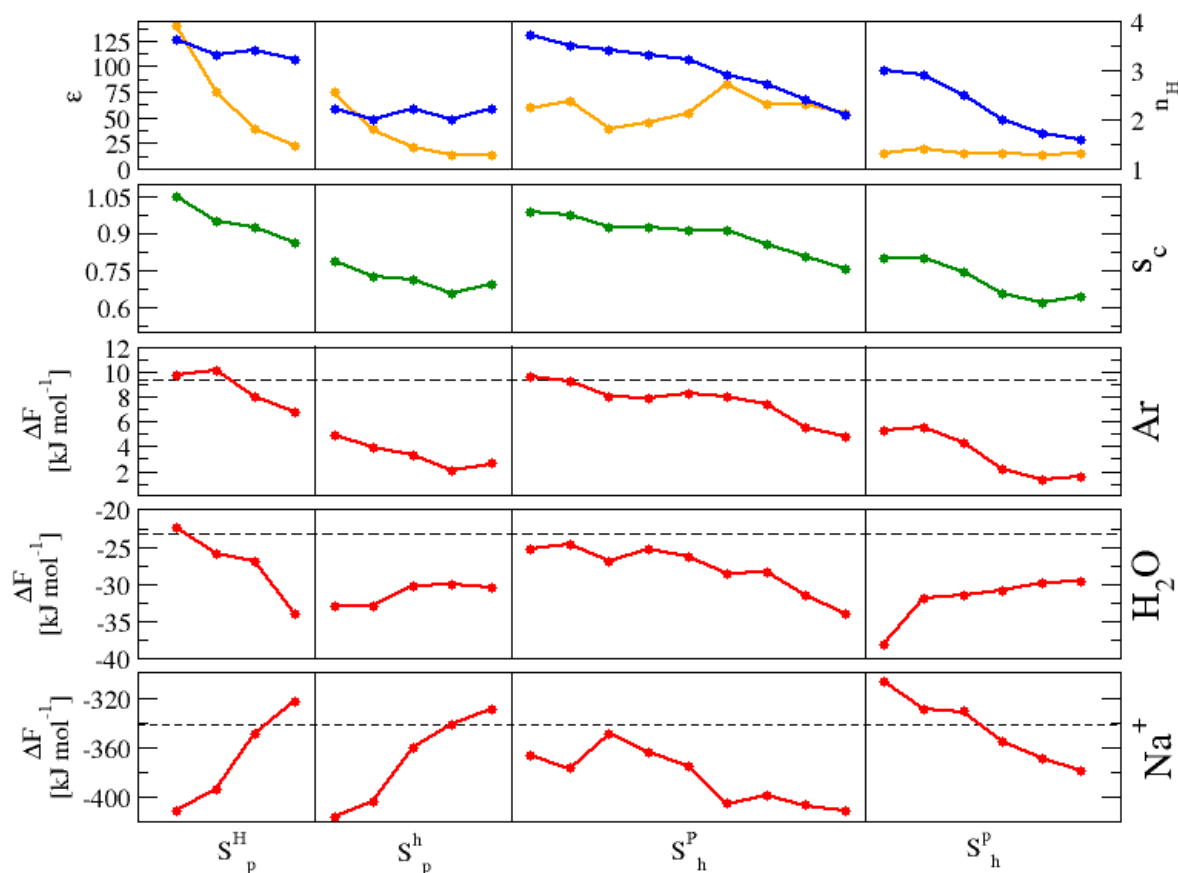
Figure 2.13: Solvation free energy of three species in artificial solvents of the NPT series at 298.15 K and 1 bar. The values are calculated from the progressive perturbation of the solute into a dummy atom using thermodynamic integration. The three compounds considered are (from top to bottom) argon Ar, SPC water H_2O and sodium ion Na^+. The points follow the four series of artificial solvents (Table 2.1) S_p^H, S_p^h, S_h^P and S_h^p. The values of the solvent permittivity ε (orange, scale on the left) and hydrogen-bonding capacity n_H (blue, scale on the right) are also displayed (top row), together with the scaling factor s_c controlling the Lennard-Jones repulsive term.

2.4   Conclusions

In the present work, we have considered a wide spectrum of water-like models derived from SPC water model[136] with altered geometries (OH bond distance) and dipole moments (O and H charges), but approximately the same molecular size (same or similar Lennard-Jones interaction parameters). Their simulated properties have been investigated in details in order to examine the influence of the altered molecular parameters on the collective properties of the liquid.

In the NVT series, the Lennard-Jones interaction parameters were kept identical to those of SPC water and the pressure in the system may significantly differ from the atmospheric pressure. In the NPT series, the Lennard-Jones repulsion was adjusted to reproduce densities similar to normal SPC water at room temperature and atmospheric pressure.

The analysis of the properties of these models illustrates the delicate balance between electrostatic and Lennard-Jones interactions responsible of the very particular properties of water. A decrease in the bond length or/and partial charges of the model reduces the importance of the electrostatic interaction, leading to models that behave like Lennard-Jones fluids. In the opposite, an increase in the bond length or/and partial charges overstructures the system, yielding models that ultimately adopt a glassy state.

The analysis of the correlations between the properties analyzed leads to interesting insight about the nature of certain characteristics of water. In particular, the dielectric permittivity seems to mainly depend on the dipole moment of the model, while the hydrogen-bonding capacity is more connected to higher order multipoles, in particular the quadrupole and the octupole. Additionally the correlation of the dynamic properties (diffusion coefficient D, rotational correlation time τ_2^{OH} and the Debye relaxation time τ_D) can be used to characterize the two different type of behaviors of the models described before.

The low correlation between dielectric permittivity and hydrogen-bonding capacity of these models allows for their use as liquid environment in studies concerning the effect of the polarity of the solvent on a solute, giving the possibility to analyze (almost) separately the influence of these two characteristics. A simple example of this is provided here with the calculation of the solvation free energy of three different chemical species as a function of the dielectric permittivity and of the hydrogen-bonding capacity of the solvent. It was possible to qualitatively determine the importance of the electrostatic interactions, and more specifically of the dielectric permittivity, for ionic solvation, and of the Lennard-Jones interactions for a hydrophobic compound. The interplay between these two aspect of the polarity of the solvent was highlighted in the determination of the solvation free energy of water.

More extensive studies on the conformational properties of saccharides (monomers and dimers) using these artificial water-like models are described in the following chapters of this thesis.

2.A Multipoles

In this appendix, we derive Eq. 2.1-2.5 concerning the multipole moments of a molecule with three atomic sites (SPC-like geometry). A discrete distribution of N point charges q_l at positions $\mathbf{r}_l$ in vacuum is considered. The electric potential $\Phi(\mathbf{R})$ generated by this distribution at position $\mathbf{R}$ from an origin O, chosen close to the center of mass of the distribution, is given by the Coulomb potential

$$\Phi(\mathbf{R}) = \frac{1}{4\pi\varepsilon_0} \sum_{l=1}^{N} \frac{q_l}{||\mathbf{R} - \mathbf{r}_l||}, \tag{A.2.1}$$

where ε_0 is the dielectric permittivity in vacuum.

This expression can be expanded as a Taylor series considering distances R from the origin O that are larger than the maximum distance R_0 between the charges, $i.e.$ when $R >> R_0$ with $R = ||\mathbf{R}||$ and $R_0 = \max(||\mathbf{r}_l||)$, as

$$\Phi(\mathbf{R}) = \frac{1}{4\pi\varepsilon_0} \left(\frac{1}{R}q + \frac{1}{R^2} \sum_{i=x,y,z} R_i\mu_i + \sum_{i,j=x,y,z} R_iR_jQ_{i,j} \right. \tag{A.2.2}$$

$$\left. + \frac{1}{6}\frac{1}{R^4} \sum_{i,j,k=x,y,z} R_iR_jR_kO_{i,j,k} + \cdots \right) \tag{A.2.3}$$

where

$$q = \sum_{l=1}^{N} q_l \tag{A.2.4}$$

$$\mu_i = \sum_{l=1}^{N} q_l r_{l,i} \tag{A.2.5}$$

$$Q_{i,j} = \frac{1}{2} \sum_{l=1}^{N} q_l (3 r_{l,i} r_{l,j} - r_l^2 \delta_{ij}) \tag{A.2.6}$$

$$O_{i,j,k} = \sum_{l=1}^{N} (15 r_{l,i} r_{l,j} r_{l,k} - 3 r_l^2 (r_{l,i} \delta_{jk} + r_{l,j} \delta_{ik} + r_{l,k} \delta_{i,j})) \tag{A.2.7}$$

are the total charge q of the system, the elements μ_i of the dipole vector, the elements $Q_{i,j}$ of the traceless quadrupole tensor Q_{ij} and the elements $O_{i,j,k}$ of the octupole tensor Ω_{ijk}. The index l refers to the individual charges, while the indices i, j, k refer to the x, y, z components of tensors. The whole formalism can also be generalized to a continuous distribution.

Let us consider a water molecule described by the SPC model[136] (see Figure 2.1). The charges are distributed on the oxygen $(-2q)$ and on the two hydrogens (q). The distances between them are given by the bond length b (d_{OH}) and by $2b \sin\alpha$ (d_{HH}), where $\alpha = \theta/2$ is half of the angle HOH. A coordinate system is defined with the origin O on the oxygen atom, the xz plane corresponding to the molecule plane, the z-axis on the bisector of the bond angle, and the y axis normal to the molecular plane. In this case the total charge of the system is zero, the molecule being neutral by construction.

The dipole vector μ_i, the traceless quadrupole tensor Q_{ij} and the traceless octupole tensor Ω_{ijk} can be written in their matrix form. The dipole vector μ_i has one non zero element

$$\mu_i = \begin{pmatrix} 0 & 0 & 2q r_{H,z} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 2qb \cos\alpha \end{pmatrix} \tag{A.2.8}$$

The non-zero elements of the quadrupole tensor Q_{ij} are the diagonal elements

$$Q_{ij} = \begin{pmatrix} Q_{xx} & 0 & 0 \\ 0 & Q_{yy} & 0 \\ 0 & 0 & Q_{zz} \end{pmatrix} \tag{A.2.9}$$

with $Q_{xx} = q(2r_{H,x}^2 - r_{H,z}^2)$, $Q_{yy} = q(-r_{H,x}^2 - r_{H,z}^2)$ and $Q_{zz} = q(-r_{H,x}^2 + 2r_{H,z}^2)$. The tensor is traceless, $Q_{xx} + Q_{yy} + Q_{zz} = 0$.

An equivalent alternative form of the quadrupole tensor

$$Q_{ij} = \begin{pmatrix} Q_T - Q_0 & 0 & 0 \\ 0 & -Q_T - Q_0 & 0 \\ 0 & 0 & 2Q_0 \end{pmatrix} \tag{A.2.10}$$

where $Q_T = (Q_{xx} - Q_{yy})/2$ and $Q_0 = Q_{zz}/2$, can be written using the linear component Q_0 and the square component Q_T of the quadrupole tensor. An origin O' for which the linear component of the quadrupole is equal to zero ($2Q_0 = Q_{zz} = 0$) can be chosen. In this way the quadrupole tensor is expressed just by one component, the square component Q_T, defined as

$$Q_T = \frac{3}{2}qr_{H,x}^2 = \frac{3}{2}qb^2\sin^2\alpha. \tag{A.2.11}$$

It is clear that the square component Q_T of the quadrupole is independent of the choice of the origin O of coordinate and it can be used univocally to describe the strength of the quadrupole interactions.

The non-zero elements of the octupole tensor Ω_{ijk} are

$$[\Omega]_{ijx} = \begin{pmatrix} 0 & 0 & O_{xzx} \\ 0 & 0 & 0 \\ O_{zxx} & 0 & 0 \end{pmatrix}, \tag{A.2.12}$$

$$[\Omega]_{ijy} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & O_{yzy} \\ 0 & O_{zyy} & 0 \end{pmatrix}, \tag{A.2.13}$$

$$[\Omega]_{ijz} = \begin{pmatrix} O_{xxz} & 0 & 0 \\ 0 & O_{yyz} & 0 \\ 0 & 0 & O_{zzz} \end{pmatrix}, \tag{A.2.14}$$

where $O_{xzx} = O_{zxx} = O_{xxz} = q(4r_{H,x}^2 r_{H,z} - r_{H,z}^3)$, $O_{yzy} = O_{zyy} = O_{yyz} = -q(r_{H,z}^3 + r_{H,x}^2 r_{H,z})$ and $O_{zzz} = q(2r_{H,z}^3 - 3r_{H,x}^2 r_{H,z})$.

As for the quadrupole, an alternative equivalent form can be written using the linear component Ω_0 and a square component Ω_T of the octupole tensor

$$[\Omega]_{ijx} = \begin{pmatrix} 0 & 0 & \Omega_T - \Omega_0/2 \\ 0 & 0 & 0 \\ \Omega_T - \Omega_0/2 & 0 & 0 \end{pmatrix}, \tag{A.2.15}$$

$$[\Omega]_{ijy} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\Omega_T - \Omega_0/2 \\ 0 & -\Omega_T - \Omega_0/2 & 0 \end{pmatrix}, \tag{A.2.16}$$

$$[\Omega]_{ijz} = \begin{pmatrix} \Omega_T - \Omega_0/2 & 0 & 0 \\ 0 & -\Omega_T - \Omega_0/2 & 0 \\ 0 & 0 & \Omega_0 \end{pmatrix}, \tag{A.2.17}$$

In this way only two components are necessary to describe the octupole interactions, the linear component of the octupole tensor Ω_0, defined as

$$\Omega_0 = q(2r_{H,z}^3 - 3r_{H,x}^2 r_{H,z}) = 2b^3\cos\alpha(2\sin^2\alpha - 3\cos^2\alpha) \tag{A.2.18}$$

and the square component Ω_T, defined as

$$\Omega_T = \frac{5}{2}qr_{H,x}^2 r_{H,z} = \frac{5}{2}qb^3\sin^2\alpha\cos\alpha. \tag{A.2.19}$$

2.S.1   Supplementary Material

Table 2.S.1: Simulated properies of the water-like models of the NVT series at 298.15 K and a density of 968 kg·m⁻³. For each model, the state of the system (S.) is indicated as L (liquid) or S (solid). The quantities reported are the scaling factor for the bond length s_b, the scaling factor for the partial charges s_q, the dipole moment μ of the model, the square component Q_T of the quadrupole, the linear component Ω_0 of the octupole, the square component Ω_T of the octupole, the pressure P, the total potential energy U, the self-diffusion constant D, the rotational correlation time τ_2^{HH} for the OH axis, the rotational correlation time τ_2^{OH} for the OH axis, the position r_{max}, the Debye relaxation time τ_D, the position r_{max} and the height g_{max} of the first peak of the radial distribution function $g(r)$, the position r_{max} and the height c_{max} of the first peak of the dipole-dipole orientation function $c(r)$, the static relative dielectric permittivity ε and the hydrogen-bonding capacity n_H. The quantities P, U, D, τ_D, ε and n_H are reported with a statistical error (between parenthesis) calculated as th root-mean-square deviation over five repeats of the simulations. The models that were not feasible are not reported, a blank line is denoting one or more missing combinations of scaling factors.

	S.	s_b	s_q	μ [D]	Q_T [D·Å]	Ω_0 [D·Å²]	Ω_T [D·Å²]	P [bar]	U [kJ·mol⁻¹]	D [10⁻⁹m²s⁻¹]	τ_2^{OH} [ps]	τ_2^{HH} [ps]	τ_D [ps]	r_{max} (g_{max}) [nm] (-)	r_{max} (c_{max}) [nm] (-)	ε	n_H
$W^{0.1}_{0.9}$	L	0.1	0.9	0.2	0.0	0.00	0.00	10183.7 (6.5)	-1.8 (0.0)	8.3 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.46)	0.31 (0.00)	1.1 (0.0)	1.4 (0.0)
$W^{0.1}_{1.0}$	L	0.1	1.0	0.2	0.0	0.00	0.00	7797.2 (3898.6)	-1.5 (0.8)	6.1 (3.0)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.31 (0.00)	1.2 (0.0)	1.4 (0.1)
$W^{0.1}_{1.1}$	L	0.1	1.1	0.2	0.0	0.00	0.00	9131.3 (15.4)	-2.1 (0.0)	6.8 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.54)	0.32 (0.00)	1.2 (0.0)	1.4 (0.0)
$W^{0.2}_{0.4}$	S	0.2	0.4	0.2	0.0	-0.00	0.01	7743.0 (3871.5)	-1.6 (0.8)	5.9 (2.9)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.00)	0.32 (2.50)	0.33 (0.0)	–	1.4 (0.0)
$W^{0.2}_{0.5}$	L	0.2	0.5	0.2	0.0	-0.00	0.01	9656.2 (3.9)	-2.0 (0.0)	7.6 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.50)	0.31 (0.00)	1.2 (0.0)	1.4 (0.0)
$W^{0.2}_{0.6}$	L	0.2	0.6	0.3	0.0	-0.00	0.01	9601.5 (2.6)	-2.0 (0.0)	7.5 (0.4)	<0.2	<0.2	1.0 (0.0)	0.32 (2.50)	0.32 (0.00)	1.3 (0.0)	1.4 (0.0)
$W^{0.2}_{0.7}$	L	0.2	0.7	0.3	0.1	-0.01	0.01	9514.0 (3.8)	-2.0 (0.0)	7.5 (0.2)	<0.2	<0.2	1.0 (0.0)	0.32 (2.50)	0.31 (0.00)	1.4 (0.0)	1.4 (0.0)
$W^{0.2}_{0.8}$	L	0.2	0.8	0.4	0.1	-0.01	0.01	9363.2 (4.4)	-2.1 (0.0)	7.2 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.52)	0.31 (0.00)	1.5 (0.0)	1.4 (0.0)
$W^{0.2}_{0.9}$	L	0.2	0.9	0.4	0.1	-0.01	0.01	9152.9 (2.6)	-2.2 (0.0)	6.9 (0.2)	<0.2	<0.2	1.0 (0.0)	0.32 (2.53)	0.31 (0.01)	1.6 (0.0)	1.4 (0.0)
$W^{0.2}_{1.0}$	L	0.2	1.0	0.5	0.1	-0.01	0.01	8865.3 (8.6)	-2.3 (0.0)	6.3 (0.2)	<0.2	<0.2	1.0 (0.0)	0.32 (2.55)	0.31 (0.01)	1.8 (0.0)	1.4 (0.0)
$W^{0.2}_{1.1}$	L	0.2	1.1	0.5	0.1	-0.01	0.02	8521.6 (6.2)	-2.5 (0.0)	6.0 (0.2)	<0.2	<0.2	1.0 (0.0)	0.32 (2.58)	0.31 (0.01)	2.0 (0.0)	1.4 (0.0)
$W^{0.2}_{1.2}$	L	0.2	1.2	0.5	0.1	-0.01	0.02	8140.7 (5.0)	-2.7 (0.0)	5.6 (0.2)	<0.2	<0.2	1.0 (0.0)	0.32 (2.61)	0.31 (0.01)	2.3 (0.0)	1.4 (0.0)
$W^{0.2}_{1.3}$	L	0.2	1.3	0.6	0.1	-0.01	0.02	7750.3 (6.7)	-2.9 (0.0)	5.2 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.65)	0.31 (0.01)	2.5 (0.0)	1.4 (0.0)
$W^{0.2}_{1.4}$	L	0.2	1.4	0.6	0.1	-0.01	0.02	7380.0 (1.8)	-3.1 (0.0)	4.9 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.68)	0.31 (0.02)	2.8 (0.0)	1.4 (0.0)
$W^{0.2}_{1.5}$	L	0.2	1.5	0.7	0.1	-0.01	0.02	7045.5 (2.9)	-3.3 (0.0)	4.5 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.72)	0.31 (0.02)	3.2 (0.0)	1.4 (0.0)
$W^{0.3}_{0.3}$	L	0.3	0.3	0.2	0.1	-0.01	0.01	9735.5 (5.6)	-1.9 (0.0)	7.5 (0.2)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.32 (0.00)	1.1 (0.0)	1.4 (0.0)
$W^{0.3}_{0.4}$	L	0.3	0.4	0.3	0.1	-0.01	0.02	9726.7 (4.5)	-2.0 (0.0)	7.5 (0.3)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.31 (0.00)	1.3 (0.0)	1.4 (0.0)
$W^{0.3}_{0.5}$	L	0.3	0.5	0.3	0.1	-0.02	0.02	9703.8 (2.4)	-2.0 (0.0)	7.7 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.31 (0.00)	1.4 (0.0)	1.4 (0.0)

Table 2.S.1 – Continued

	St.	s_b	s_q	μ [D]	Q_T [D·Å]	Ω_0 [D·Å²]	Ω_T [D·Å²]	P [bar]	U [kJ·mol⁻¹]	D [10⁻⁹ m²s⁻¹]	τ_2^{OH} [ps]	τ_2^{HH} [ps]	τ_D [ps]	r_{max} (g_{max}) [nm] (-)	r_{max} (c_{max}) [nm] (-)	ε	n_H
W0.3/0.6	L	0.3	0.6	0.4	0.1	-0.02	0.03	9652.1 (2.6)	-2.0 (0.0)	7.4 (0.1)	<0.2	<0.2	1.0 (0.0)	0.31 (2.49)	0.31 (0.01)	1.6 (0.0)	1.4 (0.0)
W0.3/0.7	L	0.3	0.7	0.5	0.1	-0.03	0.04	9573.4 (2.0)	-2.1 (0.0)	7.4 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.49)	0.31 (0.01)	1.9 (0.0)	1.4 (0.0)
W0.3/0.8	L	0.3	0.8	0.5	0.1	-0.03	0.04	9458.3 (1.3)	-2.3 (0.0)	7.6 (0.3)	<0.2	<0.2	1.0 (0.0)	0.32 (2.50)	0.31 (0.02)	2.3 (0.0)	1.4 (0.0)
W0.3/0.9	L	0.3	0.9	0.6	0.2	-0.03	0.04	9306.5 (2.0)	-2.5 (0.0)	7.5 (0.2)	<0.2	<0.2	1.0 (0.0)	0.32 (2.51)	0.31 (0.02)	2.7 (0.0)	1.4 (0.0)
W0.3/1.0	L	0.3	1.0	0.7	0.2	-0.04	0.05	9127.8 (1.5)	-2.7 (0.0)	7.3 (0.2)	<0.2	<0.2	1.0 (0.0)	0.32 (2.51)	0.31 (0.03)	3.2 (0.0)	1.4 (0.0)
W0.3/1.1	L	0.3	1.1	0.7	0.2	-0.04	0.06	8937.3 (4.3)	-3.0 (0.0)	6.9 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.52)	0.31 (0.04)	3.8 (0.0)	1.5 (0.0)
W0.3/1.2	L	0.3	1.2	0.8	0.2	-0.04	0.06	8729.6 (1.8)	-3.3 (0.0)	7.0 (0.2)	<0.2	<0.2	1.0 (0.0)	0.32 (2.52)	0.31 (0.05)	4.6 (0.0)	1.4 (0.0)
W0.3/1.3	L	0.3	1.3	0.9	0.2	-0.05	0.07	8513.7 (1.2)	-3.6 (0.0)	7.0 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.52)	0.31 (0.06)	5.5 (0.0)	1.5 (0.0)
W0.3/1.4	L	0.3	1.4	1.0	0.2	-0.06	0.07	8296.2 (3.1)	-4.0 (0.0)	7.1 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.53)	0.31 (0.07)	6.5 (0.0)	1.5 (0.0)
W0.3/1.5	L	0.3	1.5	1.0	0.3	-0.06	0.07	8074.1 (1.8)	-4.5 (0.0)	6.9 (0.1)	<0.2	<0.2	1.0 (0.0)	0.31 (2.52)	0.31 (0.09)	7.7 (0.1)	1.5 (0.0)
W0.4/0.2	L	0.4	0.2	0.2	0.1	-0.01	0.02	9765.1 (4.3)	-1.9 (0.0)	7.5 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.31 (0.00)	1.1 (0.0)	1.4 (0.0)
W0.4/0.3	L	0.4	0.3	0.3	0.1	-0.03	0.04	9759.1 (4.5)	-1.9 (0.0)	7.6 (0.3)	<0.2	<0.2	1.0 (0.0)	0.32 (2.48)	0.31 (0.00)	1.3 (0.0)	1.4 (0.0)
W0.4/0.4	L	0.4	0.4	0.4	0.1	-0.04	0.05	9735.1 (2.5)	-2.0 (0.0)	7.8 (0.2)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.31 (0.00)	1.5 (0.0)	1.4 (0.0)
W0.4/0.5	L	0.4	0.5	0.5	0.2	-0.04	0.06	9685.7 (3.0)	-2.1 (0.0)	7.6 (0.2)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.31 (0.01)	1.8 (0.0)	1.4 (0.0)
W0.4/0.6	L	0.4	0.6	0.5	0.2	-0.06	0.07	9603.7 (2.6)	-2.2 (0.0)	7.7 (0.2)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.31 (0.02)	2.3 (0.0)	1.4 (0.0)
W0.4/0.7	L	0.4	0.7	0.6	0.2	-0.07	0.08	9477.1 (2.3)	-2.5 (0.0)	7.7 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.31 (0.03)	2.9 (0.0)	1.4 (0.0)
W0.4/0.8	L	0.4	0.8	0.7	0.3	-0.07	0.10	9311.2 (1.7)	-2.8 (0.0)	7.8 (0.2)	<0.2	<0.2	1.0 (0.0)	0.32 (2.48)	0.31 (0.04)	3.7 (0.0)	1.5 (0.0)
W0.4/0.9	L	0.4	0.9	0.8	0.3	-0.09	0.11	9111.0 (1.9)	-3.2 (0.0)	7.8 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.31 (0.05)	4.7 (0.0)	1.5 (0.0)
W0.4/1.0	L	0.4	1.0	0.9	0.3	-0.10	0.12	8879.0 (0.5)	-3.7 (0.0)	7.8 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.31 (0.07)	5.9 (0.0)	1.5 (0.0)
W0.4/1.1	L	0.4	1.1	1.0	0.3	-0.10	0.13	8620.3 (0.9)	-4.3 (0.0)	7.9 (0.1)	<0.2	<0.2	1.0 (0.0)	0.31 (2.48)	0.31 (0.09)	7.5 (0.0)	1.5 (0.0)
W0.4/1.2	L	0.4	1.2	1.1	0.4	-0.11	0.14	8335.6 (2.1)	-5.0 (0.0)	7.8 (0.1)	<0.2	<0.2	1.0 (0.0)	0.31 (2.48)	0.30 (0.11)	9.5 (0.1)	1.5 (0.0)
W0.4/1.3	L	0.4	1.3	1.2	0.4	-0.12	0.16	8024.9 (0.5)	-5.8 (0.0)	7.7 (0.1)	<0.2	<0.2	1.0 (0.0)	0.31 (2.48)	0.30 (0.13)	11.8 (0.1)	1.5 (0.0)
W0.4/1.4	L	0.4	1.4	1.3	0.4	-0.13	0.17	7688.2 (0.8)	-6.7 (0.0)	8.2 (0.1)	<0.2	<0.2	1.0 (0.0)	0.31 (2.46)	0.30 (0.15)	15.3 (0.1)	1.5 (0.0)
W0.4/1.5	L	0.4	1.5	1.4	0.5	-0.14	0.18	7323.7 (0.2)	-7.7 (0.0)	8.3 (0.3)	<0.2	<0.2	1.0 (0.0)	0.31 (2.46)	0.30 (0.17)	19.3 (0.1)	1.5 (0.0)
W0.5/0.1	L	0.5	0.1	0.1	0.0	-0.02	0.02	9781.0 (2.7)	-1.9 (0.0)	7.7 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.30 (0.00)	1.0 (0.0)	1.4 (0.0)
W0.5/0.2	L	0.5	0.2	0.2	0.1	-0.04	0.05	9778.3 (4.1)	-1.9 (0.0)	7.9 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.31 (0.00)	1.2 (0.0)	1.4 (0.0)
W0.5/0.3	L	0.5	0.3	0.3	0.1	-0.06	0.07	9757.6 (3.7)	-2.0 (0.0)	7.8 (0.3)	<0.2	<0.2	1.0 (0.0)	0.31 (2.49)	0.30 (0.00)	1.4 (0.0)	1.4 (0.0)
W0.5/0.4	L	0.5	0.4	0.5	0.2	-0.07	0.09	9706.4 (1.2)	-2.1 (0.0)	7.6 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.31 (0.01)	1.8 (0.0)	1.4 (0.0)
W0.5/0.5	L	0.5	0.5	0.6	0.2	-0.09	0.12	9607.3 (1.9)	-2.3 (0.0)	7.6 (0.3)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.31 (0.02)	2.4 (0.0)	1.4 (0.0)
W0.5/0.6	L	0.5	0.6	0.7	0.3	-0.11	0.14	9454.2 (1.5)	-2.6 (0.0)	7.8 (0.3)	<0.2	<0.2	1.0 (0.0)	0.32 (2.48)	0.31 (0.03)	3.3 (0.0)	1.5 (0.0)
W0.5/0.7	L	0.5	0.7	0.8	0.3	-0.13	0.17	9239.5 (1.2)	-3.1 (0.0)	7.8 (0.2)	<0.2	<0.2	1.0 (0.0)	0.32 (2.48)	0.31 (0.05)	4.4 (0.0)	1.5 (0.0)
W0.5/0.8	L	0.5	0.8	0.9	0.4	-0.15	0.19	8968.9 (1.4)	-3.8 (0.0)	8.0 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.47)	0.31 (0.07)	6.0 (0.0)	1.5 (0.0)
W0.5/0.9	L	0.5	0.9	1.0	0.4	-0.17	0.21	8643.0 (1.4)	-4.5 (0.0)	8.0 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.47)	0.31 (0.09)	8.1 (0.0)	1.5 (0.0)
W0.5/1.0	L	0.5	1.0	1.1	0.5	-0.19	0.24	8264.5 (1.2)	-5.5 (0.0)	8.0 (0.0)	<0.2	<0.2	1.0 (0.0)	0.31 (2.47)	0.30 (0.12)	10.8 (0.0)	1.5 (0.0)
W0.5/1.1	L	0.5	1.1	1.2	0.5	-0.21	0.26	7837.4 (1.7)	-6.6 (0.0)	8.4 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.45)	0.30 (0.14)	14.5 (0.1)	1.6 (0.0)
W0.5/1.2	L	0.5	1.2	1.4	0.6	-0.23	0.28	7365.3 (0.9)	-8.0 (0.0)	8.7 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.44)	0.30 (0.17)	19.8 (0.2)	1.6 (0.0)
W0.5/1.3	L	0.5	1.3	1.5	0.6	-0.24	0.31	6847.2 (0.3)	-9.4 (0.0)	9.0 (0.3)	<0.2	<0.2	1.0 (0.0)	0.31 (2.41)	0.30 (0.19)	26.3 (0.1)	1.6 (0.0)
W0.5/1.4	L	0.5	1.4	1.6	0.7	-0.26	0.33	6288.4 (1.2)	-11.1 (0.0)	9.4 (0.3)	<0.2	<0.2	1.0 (0.0)	0.31 (2.40)	0.30 (0.22)	36.4 (0.1)	1.7 (0.0)

Continued on Next Page. . .

Table 2.S.1 – Continued

	St.	s_b	s_q	μ [D]	Q_T [D·Å]	Ω_0 [D·Å²]	Ω_T [D·Å²]	P [bar]	U [kJ·mol⁻¹]	D [10⁻⁹ m²s⁻¹]	τ_2^{OH} [ps]	τ_2^{HH} [ps]	τ_D [ps]	r_{max} (g_{max}) [nm] (-)	r_{max} (c_{max}) [nm] (-)	ε	n_H
$W^{0.5}_{1.5}$	L	0.5	1.5	1.7	0.7	-0.28	0.35	5695.1 (1.0)	-13.0 (0.0)	9.7 (0.1)	<0.2	<0.2	1.0 (0.0)	0.31 (2.38)	0.30 (0.25)	49.0 (0.3)	1.7 (0.0)
$W^{0.6}_{0.1}$	L	0.6	0.1	0.1	0.1	-0.03	0.04	9788.6 (1.0)	-1.9 (0.0)	7.6 (0.4)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.31 (0.00)	1.1 (0.0)	1.4 (0.0)
$W^{0.6}_{0.2}$	L	0.6	0.2	0.3	0.1	-0.06	0.08	9779.8 (1.5)	-1.9 (0.0)	7.8 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.31 (0.00)	1.3 (0.0)	1.4 (0.0)
$W^{0.6}_{0.3}$	L	0.6	0.3	0.4	0.2	-0.09	0.12	9735.9 (0.9)	-2.0 (0.0)	7.7 (0.2)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.31 (0.01)	1.6 (0.0)	1.4 (0.0)
$W^{0.6}_{0.4}$	L	0.6	0.4	0.5	0.3	-0.12	0.16	9637.3 (1.0)	-2.3 (0.0)	7.7 (0.2)	<0.2	<0.2	1.0 (0.0)	0.32 (2.48)	0.31 (0.02)	2.3 (0.0)	1.5 (0.0)
$W^{0.6}_{0.5}$	L	0.6	0.5	0.7	0.4	-0.16	0.20	9460.5 (1.1)	-2.7 (0.0)	7.9 (0.3)	<0.2	<0.2	1.0 (0.0)	0.32 (2.48)	0.31 (0.03)	3.3 (0.0)	1.5 (0.0)
$W^{0.6}_{0.6}$	L	0.6	0.6	0.8	0.4	-0.19	0.24	9194.2 (1.5)	-3.3 (0.0)	7.9 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.48)	0.31 (0.05)	4.7 (0.0)	1.5 (0.0)
$W^{0.6}_{0.7}$	L	0.6	0.7	1.0	0.5	-0.23	0.28	8839.0 (2.0)	-4.1 (0.0)	7.8 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.47)	0.31 (0.08)	6.8 (0.0)	1.5 (0.0)
$W^{0.6}_{0.8}$	L	0.6	0.8	1.1	0.6	-0.26	0.33	8399.1 (0.4)	-5.2 (0.0)	8.3 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.46)	0.30 (0.11)	9.6 (0.0)	1.5 (0.0)
$W^{0.6}_{0.9}$	L	0.6	0.9	1.2	0.6	-0.29	0.37	7879.3 (1.9)	-6.6 (0.0)	8.2 (0.1)	<0.2	<0.2	1.0 (0.0)	0.31 (2.45)	0.30 (0.14)	13.6 (0.1)	1.6 (0.0)
$W^{0.6}_{1.0}$	L	0.6	1.0	1.4	0.7	-0.33	0.41	7278.5 (1.1)	-8.2 (0.0)	9.3 (0.3)	<0.2	<0.2	1.0 (0.0)	0.31 (2.43)	0.30 (0.17)	19.4 (0.1)	1.7 (0.0)
$W^{0.6}_{1.1}$	L	0.6	1.1	1.5	0.8	-0.36	0.45	6609.3 (0.6)	-10.1 (0.0)	9.3 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.40)	0.30 (0.20)	27.3 (0.1)	1.7 (0.0)
$W^{0.6}_{1.2}$	L	0.6	1.2	1.6	0.8	-0.39	0.49	5876.1 (1.1)	-12.4 (0.0)	9.7 (0.3)	<0.2	<0.2	1.0 (0.0)	0.31 (2.38)	0.30 (0.23)	38.3 (0.4)	1.8 (0.0)
$W^{0.6}_{1.3}$	L	0.6	1.3	1.8	0.9	-0.42	0.53	5090.6 (1.2)	-14.9 (0.0)	10.4 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.35)	0.30 (0.26)	54.5 (0.6)	1.9 (0.0)
$W^{0.6}_{1.4}$	L	0.6	1.4	1.9	1.0	-0.45	0.57	4266.7 (1.7)	-17.8 (0.0)	11.0 (0.2)	<0.2	<0.2	1.0 (0.0)	0.30 (2.32)	0.30 (0.29)	75.6 (0.5)	2.0 (0.0)
$W^{0.6}_{1.5}$	L	0.6	1.5	2.0	1.1	-0.49	0.61	3419.0 (2.1)	-21.1 (0.0)	11.7 (0.3)	<0.2	<0.2	1.0 (0.0)	0.30 (2.29)	0.29 (0.32)	104.4 (0.6)	2.1 (0.0)
$W^{0.7}_{0.1}$	L	0.7	0.1	0.2	0.1	-0.05	0.06	9789.2 (1.6)	-1.9 (0.0)	7.7 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.48)	0.33 (0.00)	1.1 (0.0)	1.4 (0.0)
$W^{0.7}_{0.2}$	L	0.7	0.2	0.3	0.2	-0.10	0.13	9772.1 (1.8)	-2.0 (0.0)	7.7 (0.3)	<0.2	<0.2	1.0 (0.0)	0.32 (2.48)	0.31 (0.00)	1.5 (0.0)	1.4 (0.0)
$W^{0.7}_{0.3}$	L	0.7	0.3	0.5	0.3	-0.15	0.19	9695.0 (0.7)	-2.1 (0.0)	7.7 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.31 (0.01)	1.9 (0.0)	1.4 (0.0)
$W^{0.7}_{0.4}$	L	0.7	0.4	0.6	0.4	-0.20	0.26	9521.6 (1.2)	-2.5 (0.0)	7.8 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.48)	0.31 (0.03)	2.9 (0.0)	1.5 (0.0)
$W^{0.7}_{0.5}$	L	0.7	0.5	0.8	0.5	-0.26	0.32	9225.4 (1.6)	-3.2 (0.0)	7.6 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.47)	0.31 (0.05)	4.4 (0.0)	1.5 (0.0)
$W^{0.7}_{0.6}$	L	0.7	0.6	1.0	0.6	-0.31	0.39	8798.5 (1.5)	-4.2 (0.0)	7.9 (0.3)	<0.2	<0.2	1.0 (0.0)	0.31 (2.47)	0.31 (0.08)	6.8 (0.0)	1.5 (0.0)
$W^{0.7}_{0.7}$	L	0.7	0.7	1.1	0.7	-0.36	0.45	8243.0 (1.1)	-5.6 (0.0)	8.4 (0.1)	<0.2	<0.2	1.0 (0.0)	0.31 (2.46)	0.30 (0.11)	10.1 (0.1)	1.6 (0.0)
$W^{0.7}_{0.8}$	L	0.7	0.8	1.3	0.8	-0.41	0.52	7565.5 (0.9)	-7.4 (0.0)	8.6 (0.3)	<0.2	<0.2	1.0 (0.0)	0.31 (2.43)	0.30 (0.14)	15.4 (0.1)	1.7 (0.0)
$W^{0.7}_{0.9}$	L	0.7	0.9	1.4	0.9	-0.47	0.58	6774.2 (0.6)	-9.5 (0.0)	9.5 (0.1)	<0.2	<0.2	1.0 (0.0)	0.31 (2.40)	0.30 (0.18)	22.3 (0.2)	1.8 (0.0)
$W^{0.7}_{1.0}$	L	0.7	1.0	1.6	1.0	-0.52	0.65	5885.7 (1.3)	-12.1 (0.0)	9.8 (0.3)	<0.2	<0.2	1.0 (0.0)	0.31 (2.37)	0.30 (0.21)	32.3 (0.2)	1.9 (0.0)
$W^{0.7}_{1.1}$	L	0.7	1.1	1.7	1.1	-0.57	0.71	4914.6 (0.8)	-15.2 (0.0)	10.8 (0.1)	0.1	0.1	1.0 (0.0)	0.31 (2.34)	0.30 (0.24)	46.3 (0.2)	2.0 (0.0)
$W^{0.7}_{1.2}$	L	0.7	1.2	1.9	1.2	-0.62	0.78	3883.0 (1.3)	-18.9 (0.0)	11.5 (0.4)	<0.2	<0.2	1.0 (0.0)	0.30 (2.30)	0.29 (0.28)	64.6 (0.6)	2.2 (0.0)
$W^{0.7}_{1.3}$	L	0.7	1.3	2.1	1.3	-0.67	0.84	2827.1 (2.1)	-23.1 (0.0)	11.7 (0.1)	<0.2	<0.2	0.9 (0.1)	0.30 (2.28)	0.29 (0.32)	87.1 (0.5)	2.4 (0.0)
$W^{0.7}_{1.4}$	L	0.7	1.4	2.2	1.3	-0.72	0.91	1778.4 (1.3)	-28.2 (0.0)	11.7 (0.3)	0.2	0.2	1.2 (0.2)	0.29 (2.30)	0.29 (0.36)	115.1 (1.0)	2.6 (0.0)
$W^{0.7}_{1.5}$	L	0.7	1.5	2.4	1.4	-0.78	0.97	774.1 (1.3)	-34.1 (0.0)	9.8 (0.1)	0.2	0.2	2.0 (0.2)	0.29 (2.41)	0.28 (0.42)	140.4 (2.8)	2.8 (0.0)
$W^{0.8}_{0.1}$	L	0.8	0.1	0.2	0.1	-0.07	0.10	9790.6 (1.4)	-1.9 (0.0)	7.8 (0.2)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.31 (0.00)	1.1 (0.0)	1.4 (0.0)
$W^{0.8}_{0.2}$	L	0.8	0.2	0.4	0.2	-0.15	0.19	9756.1 (2.0)	-2.0 (0.0)	7.8 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.48)	0.31 (0.00)	1.5 (0.0)	1.4 (0.0)
$W^{0.8}_{0.3}$	L	0.8	0.3	0.5	0.4	-0.23	0.29	9625.5 (0.9)	-2.3 (0.0)	7.7 (0.0)	<0.2	<0.2	1.0 (0.0)	0.31 (2.48)	0.31 (0.02)	2.3 (0.0)	1.4 (0.0)
$W^{0.8}_{0.4}$	L	0.8	0.4	0.7	0.5	-0.31	0.39	9341.9 (0.8)	-2.9 (0.0)	8.2 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.48)	0.31 (0.04)	3.7 (0.0)	1.5 (0.0)
$W^{0.8}_{0.5}$	L	0.8	0.5	0.9	0.6	-0.38	0.48	8875.1 (0.7)	-4.0 (0.0)	8.0 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.47)	0.31 (0.07)	6.0 (0.0)	1.5 (0.0)
$W^{0.8}_{0.6}$	L	0.8	0.6	1.1	0.8	-0.46	0.58	8225.3 (1.4)	-5.6 (0.0)	8.3 (0.1)	<0.2	<0.2	1.0 (0.0)	0.31 (2.45)	0.30 (0.10)	9.5 (0.0)	1.6 (0.0)

Continued on Next Page. . .

Table 2.S.1 – Continued

	St.	s_b	s_q	μ [D]	Q_T [D·Å]	Ω_0 [D·Å²]	Ω_T [D·Å²]	P [bar]	U [kJ·mol⁻¹]	D [10⁻⁹ m²s⁻¹]	τ_2^{OH} [ps]	τ_2^{HH} [ps]	τ_D [ps]	r_{max} (g_{max}) [nm] (-)	r_{max} (c_{max}) [nm] (-)	ε	n_H
$W^{0.8}_{0.7}$	L	0.8	0.7	1.3	0.9	-0.54	0.68	7399.3 (0.3)	-7.7 (0.0)	8.6 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.42)	0.30 (0.14)	14.8 (0.1)	1.7 (0.0)
$W^{0.8}_{0.8}$	L	0.8	0.8	1.5	1.0	-0.62	0.78	6413.1 (1.2)	-10.4 (0.0)	9.5 (0.4)	<0.2	<0.2	1.0 (0.0)	0.31 (2.39)	0.30 (0.18)	22.2 (0.3)	1.9 (0.0)
$W^{0.8}_{0.9}$	L	0.8	0.9	1.6	1.1	-0.69	0.87	5291.1 (0.5)	-13.7 (0.0)	10.3 (0.5)	<0.2	<0.2	1.0 (0.0)	0.31 (2.34)	0.30 (0.21)	33.1 (0.4)	2.0 (0.0)
$W^{0.8}_{1.0}$	L	0.8	1.0	1.8	1.3	-0.77	0.97	4068.3 (0.9)	-17.8 (0.0)	11.3 (0.1)	<0.2	<0.2	1.0 (0.0)	0.30 (2.29)	0.29 (0.25)	47.4 (0.3)	2.3 (0.0)
$W^{0.8}_{1.1}$	L	0.8	1.1	2.0	1.4	-0.85	1.07	2798.9 (1.5)	-22.9 (0.0)	11.8 (0.1)	<0.2	<0.2	1.2 (0.3)	0.30 (2.27)	0.29 (0.30)	64.5 (0.3)	2.5 (0.0)
$W^{0.8}_{1.2}$	L	0.8	1.2	2.2	1.5	-0.93	1.16	1547.5 (0.4)	-29.1 (0.0)	10.7 (0.1)	0.3	0.3	1.5 (0.2)	0.29 (2.32)	0.28 (0.35)	82.2 (0.9)	2.8 (0.0)
$W^{0.8}_{1.3}$	L	0.8	1.3	2.4	1.6	-1.00	1.26	385.9 (2.3)	-36.8 (0.0)	7.8 (0.2)	0.4	0.4	2.4 (0.1)	0.28 (2.52)	0.28 (0.43)	99.1 (1.2)	3.1 (0.0)
$W^{0.8}_{1.4}$	L	0.8	1.4	2.5	1.8	-1.08	1.36	-667.2 (2.9)	-46.3 (0.0)	4.2 (0.1)	1.0	1.1	5.6 (0.2)	0.28 (2.93)	0.27 (0.52)	115.1 (1.1)	3.4 (0.0)
$W^{0.8}_{1.5}$	L	0.8	1.5	2.7	1.9	-1.16	1.45	-1691.2 (2.0)	-58.3 (0.0)	0.9 (0.0)	3.6	3.0	24.7 (0.9)	0.27 (3.66)	0.27 (0.66)	125.0 (6.0)	3.7 (0.0)
$W^{0.9}_{0.1}$	L	0.9	0.1	0.2	0.2	-0.10	0.14	9790.1 (0.7)	-1.9 (0.0)	7.7 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.48)	0.32 (0.00)	1.1 (0.0)	1.4 (0.0)
$W^{0.9}_{0.2}$	L	0.9	0.2	0.4	0.3	-0.21	0.28	9731.6 (2.1)	-2.0 (0.0)	7.9 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.48)	0.31 (0.01)	1.7 (0.0)	1.4 (0.0)
$W^{0.9}_{0.3}$	L	0.9	0.3	0.6	0.5	-0.33	0.41	9519.4 (1.3)	-2.5 (0.0)	7.9 (0.1)	<0.2	<0.2	1.0 (0.0)	0.31 (2.48)	0.31 (0.02)	2.7 (0.0)	1.5 (0.0)
$W^{0.9}_{0.4}$	L	0.9	0.4	0.8	0.6	-0.44	0.55	9076.9 (1.9)	-3.5 (0.0)	8.0 (0.2)	<0.2	<0.2	1.0 (0.0)	0.32 (2.47)	0.31 (0.05)	4.7 (0.0)	1.5 (0.0)
$W^{0.9}_{0.5}$	L	0.9	0.5	1.0	0.8	-0.55	0.69	8377.7 (1.4)	-5.1 (0.0)	8.3 (0.1)	<0.2	<0.2	1.0 (0.0)	0.31 (2.45)	0.31 (0.09)	7.9 (0.0)	1.6 (0.0)
$W^{0.9}_{0.6}$	L	0.9	0.6	1.2	1.0	-0.66	0.83	7431.1 (1.7)	-7.5 (0.0)	8.8 (0.1)	<0.2	<0.2	1.0 (0.0)	0.31 (2.43)	0.30 (0.13)	12.9 (0.1)	1.8 (0.0)
$W^{0.9}_{0.7}$	L	0.9	0.7	1.4	1.1	-0.77	0.96	6262.5 (1.7)	-10.6 (0.0)	9.6 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.38)	0.30 (0.17)	20.4 (0.2)	1.9 (0.0)
$W^{0.9}_{0.8}$	L	0.9	0.8	1.6	1.3	-0.88	1.10	4910.4 (0.8)	-14.7 (0.0)	10.3 (0.3)	<0.2	<0.2	1.0 (0.0)	0.31 (2.32)	0.30 (0.21)	30.7 (0.1)	2.2 (0.0)
$W^{0.9}_{0.9}$	L	0.9	0.9	1.8	1.4	-0.99	1.24	3439.4 (1.0)	-19.9 (0.0)	11.7 (0.2)	<0.2	<0.2	1.2 (0.5)	0.30 (2.25)	0.29 (0.25)	44.7 (0.2)	2.5 (0.0)
$W^{0.9}_{1.0}$	L	0.9	1.0	2.0	1.6	-1.10	1.38	1953.1 (2.5)	-26.7 (0.0)	11.0 (0.1)	0.2	0.3	1.4 (0.1)	0.29 (2.27)	0.28 (0.31)	58.3 (0.3)	2.8 (0.0)
$W^{0.9}_{1.1}$	L	0.9	1.1	2.2	1.8	-1.21	1.52	576.8 (1.7)	-35.4 (0.0)	7.6 (0.2)	0.4	0.4	2.5 (0.2)	0.28 (2.51)	0.28 (0.40)	74.5 (0.8)	3.1 (0.0)
$W^{0.9}_{1.2}$	L	0.9	1.2	2.5	1.9	-1.32	1.66	-632.4 (0.8)	-46.9 (0.0)	3.1 (0.1)	1.1	1.1	7.4 (0.3)	0.28 (3.07)	0.27 (0.51)	88.5 (1.0)	3.5 (0.0)
$W^{0.9}_{1.3}$	L	0.9	1.3	2.7	2.1	-1.43	1.79	-1835.3 (4.2)	-62.1 (0.0)	0.2 (0.0)	>21·10³	>21·10³	>21·10³ (0.0)	0.27 (4.03)	0.27 (0.66)	83.1 (4.8)	3.8 (0.0)
$W^{0.9}_{1.4}$	S	0.9	1.4	2.9	2.2	-1.54	1.93	-4428.1 (162.0)	-79.3 (0.0)	0.0 (0.0)	>21·10³	>21·10³	>21·10³ (0.0)	0.27 (5.13)	0.27 (0.83)	-	4.0 (0.0)
$W^{0.9}_{1.5}$	S	0.9	1.5	3.1	2.4	-1.65	2.07	-6449.2 (189.2)	-95.2 (0.0)	0.0 (0.0)	>21·10³	>21·10³	>21·10³ (0.0)	0.26 (5.72)	0.26 (0.87)	-	3.9 (0.1)
$W^{1.0}_{0.1}$	L	1.0	0.1	0.2	0.2	-0.15	0.19	9788.1 (0.6)	-1.9 (0.0)	7.7 (0.2)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.31 (0.00)	1.2 (0.0)	1.4 (0.0)
$W^{1.0}_{0.2}$	L	1.0	0.2	0.5	0.4	-0.30	0.38	9690.2 (1.8)	-2.1 (0.0)	7.7 (0.3)	<0.2	<0.2	1.0 (0.0)	0.31 (2.48)	0.31 (0.01)	1.8 (0.0)	1.4 (0.0)
$W^{1.0}_{0.3}$	L	1.0	0.3	0.7	0.6	-0.45	0.57	9359.1 (1.6)	-2.8 (0.0)	7.8 (0.1)	<0.2	<0.2	1.0 (0.0)	0.31 (2.48)	0.31 (0.03)	3.3 (0.0)	1.5 (0.0)
$W^{1.0}_{0.4}$	L	1.0	0.4	0.9	0.8	-0.61	0.76	8697.2 (1.1)	-4.3 (0.0)	8.1 (0.1)	<0.2	<0.2	1.0 (0.0)	0.31 (2.47)	0.31 (0.06)	5.9 (0.0)	1.6 (0.0)
$W^{1.0}_{0.5}$	L	1.0	0.5	1.1	1.0	-0.76	0.95	7693.4 (1.0)	-6.7 (0.0)	8.3 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.43)	0.30 (0.10)	10.1 (0.0)	1.7 (0.0)
$W^{1.0}_{0.6}$	L	1.0	0.6	1.4	1.2	-0.91	1.14	6380.1 (1.2)	-10.1 (0.0)	9.4 (0.3)	<0.2	<0.2	1.0 (0.0)	0.31 (2.38)	0.30 (0.15)	16.8 (0.2)	2.0 (0.0)
$W^{1.0}_{0.7}$	L	1.0	0.7	1.6	1.4	-1.06	1.33	4815.1 (0.3)	-14.7 (0.0)	10.5 (0.2)	<0.2	<0.2	1.0 (0.0)	0.30 (2.30)	0.29 (0.19)	26.6 (0.1)	2.3 (0.0)
$W^{1.0}_{0.8}$	L	1.0	0.8	1.8	1.6	-1.21	1.51	3096.8 (2.7)	-21.0 (0.0)	11.4 (0.2)	<0.2	0.2	0.9 (0.5)	0.30 (2.24)	0.29 (0.25)	38.2 (0.3)	2.6 (0.0)
$W^{1.0}_{0.9}$	L	1.0	0.9	2.0	1.8	-1.36	1.71	1411.7 (2.9)	-29.7 (0.0)	9.1 (0.2)	0.2	0.2	1.8 (0.1)	0.29 (2.34)	0.28 (0.33)	52.2 (0.7)	3.0 (0.0)
$W^{1.0}_{1.0}$	L	1.0	1.0	2.3	2.0	-1.51	1.89	-70.9 (0.9)	-41.5 (0.0)	4.2 (0.1)	1.4	1.4	5.0 (0.1)	0.28 (2.87)	0.27 (0.45)	64.0 (0.8)	3.4 (0.0)
$W^{1.0}_{1.1}$	L	1.0	1.1	2.5	2.2	-1.67	2.08	-1402.4 (4.4)	-58.1 (0.0)	0.4 (0.0)	12.5	14.1	>21·10³ (0.0)	0.27 (3.88)	0.27 (0.60)	74.4 (4.6)	3.8 (0.0)
$W^{1.0}_{1.2}$	S	1.0	1.2	2.7	2.4	-1.82	2.27	-4151.1 (57.7)	-78.2 (0.0)	0.0 (0.0)	>21·10³	>21·10³	>21·10³ (0.0)	0.27 (5.06)	0.27 (0.76)	-	4.0 (0.0)
$W^{1.0}_{1.3}$	S	1.0	1.3	3.0	2.6	-1.97	2.46	-6376.7 (125.9)	-96.8 (0.1)	0.0 (0.0)	>21·10³	>21·10³	>21·10³ (0.0)	0.26 (5.71)	0.26 (0.87)	-	4.0 (0.0)
$W^{1.0}_{1.4}$	S	1.0	1.4	3.2	2.8	-2.12	2.65	-8614.4 (381.9)	-118.0 (0.1)	0.0 (0.0)	>21·10³	>21·10³	>21·10³ (0.0)	0.26 (6.44)	0.26 (0.98)	-	4.0 (0.0)

Continued on Next Page...

Table 2.S.1 – Continued

	St.	s_b	s_q	μ [D]	Q_T [D.Å]	Ω_0 [D.Å²]	Ω_T [D.Å²]	P [bar]	U [kJ·mol⁻¹]	D [10⁻⁹m²s⁻¹]	τ_2^{OH} [ps]	τ_2^{HH} [ps]	τ_D [ps]	r_{max} (g_{max}) [nm] (-)	r_{max} (c_{max}) [nm] (-)	ε	n_H
W1.0/1.5	S	1.0	1.5	3.4	3.0	-2.27	2.84	-9007.8 (437.2)	-142.6 (0.1)	0.0 (0.0)	>21·10³	>21·10³	>21·10³ (0.0)	0.25 (7.37)	0.25 (1.10)	-	4.0 (0.0)
W1.1/0.1	L	1.1	0.1	0.2	0.2	-0.20	0.25	9783.1 (1.6)	-1.9 (0.0)	7.8 (0.3)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.30 (0.00)	1.2 (0.0)	1.4 (0.0)
W1.1/0.2	L	1.1	0.2	0.5	0.5	-0.40	0.50	9628.6 (2.1)	-2.3 (0.0)	7.8 (0.3)	<0.2	<0.2	1.0 (0.0)	0.32 (2.48)	0.30 (0.01)	2.0 (0.0)	1.5 (0.0)
W1.1/0.3	L	1.1	0.3	0.7	0.7	-0.60	0.76	9123.9 (1.1)	-3.3 (0.0)	7.9 (0.1)	<0.2	<0.2	1.0 (0.0)	0.31 (2.48)	0.30 (0.04)	3.9 (0.0)	1.5 (0.0)
W1.1/0.4	L	1.1	0.4	1.0	1.0	-0.80	1.01	8175.2 (1.3)	-5.4 (0.0)	8.3 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.44)	0.30 (0.08)	7.2 (0.0)	1.7 (0.0)
W1.1/0.5	L	1.1	0.5	1.2	1.2	-1.00	1.26	6797.2 (0.7)	-8.8 (0.0)	9.2 (0.1)	<0.2	<0.2	1.0 (0.0)	0.31 (2.39)	0.30 (0.12)	12.6 (0.1)	1.9 (0.0)
W1.1/0.6	L	1.1	0.6	1.5	1.4	-1.21	1.51	5058.2 (1.4)	-13.7 (0.0)	10.4 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.30)	0.29 (0.17)	21.3 (0.2)	2.3 (0.0)
W1.1/0.7	L	1.1	0.7	1.7	1.7	-1.41	1.76	3097.3 (1.0)	-20.8 (0.0)	10.8 (0.2)	0.2	0.2	1.3 (0.3)	0.30 (2.21)	0.29 (0.23)	31.6 (0.2)	2.7 (0.0)
W1.1/0.8	L	1.1	0.8	2.0	1.9	-1.61	2.02	1179.0 (2.3)	-31.1 (0.0)	7.8 (0.2)	0.3	0.3	2.1 (0.1)	0.28 (2.42)	0.28 (0.33)	44.4 (1.2)	3.2 (0.0)
W1.1/0.9	L	1.1	0.9	2.2	2.1	-1.81	2.27	-456.3 (2.9)	-46.4 (0.0)	2.1 (0.0)	1.4	1.6	10.1 (0.3)	0.27 (3.23)	0.27 (0.48)	56.5 (1.1)	3.6 (0.0)
W1.1/1.0	S	1.1	1.0	2.5	2.4	-2.02	2.52	-2662.5 (30.6)	-68.2 (0.0)	0.0 (0.0)	>21·10³	>21·10³	>21·10³ (0.0)	0.27 (4.62)	0.27 (0.67)	56.6 (6.4)	3.9 (0.0)
W1.1/1.1	S	1.1	1.1	2.8	2.6	-2.21	2.77	-6168.1 (163.5)	-89.9 (0.0)	0.0 (0.0)	>21·10³	>21·10³	>21·10³ (0.0)	0.26 (5.47)	0.26 (0.74)	-	3.9 (0.0)
W1.1/1.2	S	1.1	1.2	3.0	2.9	-2.42	3.02	-7981.0 (209.2)	-114.0 (0.0)	0.0 (0.0)	>21·10³	>21·10³	>21·10³ (0.0)	0.26 (6.39)	0.26 (0.85)	-	4.0 (0.0)
W1.1/1.3	S	1.1	1.3	3.2	3.1	-2.62	3.28	-9278.4 (407.0)	-141.8 (0.0)	0.0 (0.0)	>21·10³	>21·10³	>21·10³ (0.0)	0.25 (7.36)	0.25 (1.00)	-	4.0 (0.1)
W1.1/1.4	S	1.1	1.4	3.5	3.3	-2.82	3.53	-12780.9 (241.6)	-173.0 (0.1)	0.0 (0.0)	>21·10³	>21·10³	>21·10³ (0.0)	0.25 (8.20)	0.25 (1.24)	-	3.5 (0.1)
W1.1/1.5	S	1.1	1.5	3.8	3.6	-3.02	3.78	-15505.3 (519.9)	-209.3 (0.3)	0.0 (0.0)	>21·10³	>21·10³	>21·10³ (0.0)	0.25 (9.25)	0.25 (1.24)	-	3.9 (0.0)
W1.2/0.1	L	1.2	0.1	0.3	0.3	-0.26	0.33	9774.6 (1.2)	-2.0 (0.0)	7.7 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.31 (0.00)	1.3 (0.0)	1.4 (0.0)
W1.2/0.2	L	1.2	0.2	0.5	0.6	-0.52	0.65	9532.4 (0.9)	-2.4 (0.0)	7.7 (0.2)	<0.2	<0.2	1.0 (0.0)	0.32 (2.48)	0.31 (0.01)	2.3 (0.0)	1.5 (0.0)
W1.2/0.3	L	1.2	0.3	0.8	0.8	-0.78	0.98	8796.6 (1.1)	-3.9 (0.0)	7.9 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.47)	0.31 (0.04)	4.5 (0.0)	1.6 (0.0)
W1.2/0.4	L	1.2	0.4	1.1	1.1	-1.04	1.31	7490.1 (0.9)	-6.9 (0.0)	8.6 (0.3)	<0.2	<0.2	1.0 (0.0)	0.31 (2.42)	0.30 (0.09)	8.6 (0.0)	1.9 (0.0)
W1.2/0.5	L	1.2	0.5	1.4	1.4	-1.31	1.64	5667.9 (1.3)	-11.6 (0.0)	9.7 (0.4)	<0.2	<0.2	1.0 (0.0)	0.31 (2.33)	0.30 (0.14)	15.2 (0.1)	2.2 (0.0)
W1.2/0.6	L	1.2	0.6	1.6	1.7	-1.57	1.96	3484.0 (0.8)	-18.9 (0.0)	10.4 (0.2)	<0.2	<0.2	1.0 (0.3)	0.30 (2.20)	0.29 (0.20)	25.1 (0.2)	2.7 (0.0)
W1.2/0.7	L	1.2	0.7	1.9	2.0	-1.83	2.29	1264.3 (1.2)	-30.4 (0.0)	7.1 (0.1)	0.5	0.6	2.3 (0.2)	0.28 (2.43)	0.28 (0.32)	37.7 (0.3)	3.2 (0.0)
W1.2/0.8	L	1.2	0.8	2.2	2.3	-2.09	2.62	-647.1 (4.4)	-48.5 (0.0)	1.3 (0.0)	2.7	4.1	16.3 (0.3)	0.27 (3.46)	0.27 (0.49)	49.7 (1.4)	3.7 (0.0)
W1.2/0.9	S	1.2	0.9	2.5	2.5	-2.35	2.94	-4364.0 (194.8)	-74.1 (0.1)	0.0 (0.0)	>21·10³	>21·10³	>21·10³ (0.0)	0.27 (4.93)	0.27 (0.67)	-	4.0 (0.0)
W1.2/1.0	S	1.2	1.0	2.7	2.8	-2.62	3.27	-6427.1 (234.1)	-99.7 (0.1)	0.0 (0.0)	>21·10³	>21·10³	>21·10³ (0.0)	0.26 (6.00)	0.26 (0.84)	-	3.9 (0.0)
W1.2/1.1	S	1.2	1.1	3.0	3.1	-2.88	3.60	-8664.4 (160.2)	-129.0 (0.1)	0.0 (0.0)	>21·10³	>21·10³	>21·10³ (0.0)	0.25 (7.00)	0.25 (0.92)	-	4.0 (0.0)
W1.2/1.2	S	1.2	1.2	3.3	3.4	-3.14	3.93	-11047.9 (133.2)	-163.8 (0.1)	0.0 (0.0)	>21·10³	>21·10³	>21·10³ (0.0)	0.25 (8.19)	0.25 (1.12)	-	4.0 (0.1)
W1.2/1.3	S	1.2	1.3	3.5	3.7	-3.40	4.25	-12428.2 (331.7)	-205.0 (0.1)	0.0 (0.0)	>21·10³	>21·10³	>21·10³ (0.0)	0.24 (9.47)	0.24 (1.23)	-	3.7 (0.1)
W1.2/1.4	S	1.2	1.4	3.8	4.0	-3.66	4.58	-17228.1 (215.4)	-251.6 (0.1)	0.0 (0.0)	>21·10³	>21·10³	>21·10³ (0.0)	0.24 (10.54)	0.24 (1.28)	-	3.8 (0.0)
W1.2/1.5	S	1.2	1.5	4.1	4.3	-3.93	4.91	-19801.8 (282.5)	-305.3 (0.0)	0.0 (0.0)	>21·10³	>21·10³	>21·10³ (0.0)	0.24 (11.55)	0.24 (1.25)	-	3.8 (0.0)
W1.3/0.1	L	1.3	0.1	0.3	0.3	-0.33	0.41	9761.4 (2.1)	-2.0 (0.0)	7.8 (0.3)	<0.2	<0.2	1.0 (0.0)	0.31 (2.48)	0.31 (0.00)	1.3 (0.0)	1.4 (0.0)
W1.3/0.2	L	1.3	0.2	0.6	0.7	-0.66	0.83	9391.1 (1.1)	-2.7 (0.0)	7.8 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.48)	0.31 (0.02)	2.5 (0.0)	1.5 (0.0)
W1.3/0.3	L	1.3	0.3	0.9	1.0	-0.99	1.25	8357.0 (1.2)	-4.8 (0.0)	8.1 (0.1)	<0.2	<0.2	1.0 (0.0)	0.31 (2.45)	0.30 (0.05)	5.3 (0.0)	1.7 (0.0)
W1.3/0.4	L	1.3	0.4	1.2	1.3	-1.32	1.66	6620.9 (1.0)	-8.8 (0.0)	9.1 (0.1)	<0.2	<0.2	1.0 (0.0)	0.31 (2.37)	0.30 (0.10)	10.2 (0.0)	2.1 (0.0)
W1.3/0.5	L	1.3	0.5	1.5	1.7	-1.66	2.08	4304.0 (2.2)	-15.6 (0.0)	10.2 (0.4)	<0.2	<0.2	1.0 (0.0)	0.30 (2.22)	0.29 (0.16)	18.2 (0.1)	2.6 (0.0)
W1.3/0.6	L	1.3	0.6	1.8	2.0	-1.99	2.50	1736.3 (0.9)	-27.2 (0.0)	7.8 (0.1)	0.3	0.4	2.0 (0.1)	0.28 (2.31)	0.28 (0.27)	29.6 (0.2)	3.2 (0.0)

Continued on Next Page…

Table 2.S.1 – Continued

	St.	s_b	s_q	μ [D]	Q_T [D·Å]	Ω_0 [D·Å²]	Ω_T [D·Å²]	P [bar]	U [kJ·mol⁻¹]	D [10⁻⁹ m²s⁻¹]	τ_2^{OH} [ps]	τ_2^{HH} [ps]	τ_D [ps]	r_{max} (g_{max}) [nm] (-)	r_{max} (c_{max}) [nm] (-)	ε	n_H
$W^{1.3}_{0.7}$	L	1.3	0.7	2.1	2.3	-2.33	2.91	-595.5 (5.1)	-46.8 (0.0)	1.3 (0.0)	3.7	4.3	15.0 (0.5)	0.27 (3.43)	0.27 (0.46)	40.6 (1.3)	3.7 (0.0)
$W^{1.3}_{0.8}$	S	1.3	0.8	2.4	2.7	-2.65	3.33	-5049.9 (108.1)	-75.5 (0.1)	0.0 (0.0)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.0)	0.27 (5.02)	0.26 (0.63)	-	4.0 (0.0)
$W^{1.3}_{0.9}$	S	1.3	0.9	2.7	3.0	-2.99	3.75	-6595.3 (367.0)	-105.1 (0.0)	0.0 (0.0)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.0)	0.26 (6.29)	0.26 (0.88)	-	3.8 (0.1)
$W^{1.3}_{1.0}$	S	1.3	1.0	3.0	3.3	-3.32	4.16	-8545.7 (243.7)	-141.5 (0.1)	0.0 (0.0)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.0)	0.25 (7.65)	0.25 (0.94)	-	3.9 (0.1)
$W^{1.3}_{1.1}$	S	1.3	1.1	3.2	3.7	-3.66	4.58	-13050.4 (249.1)	-184.5 (0.1)	0.0 (0.0)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.0)	0.25 (8.91)	0.25 (0.99)	-	4.0 (0.0)
$W^{1.3}_{1.2}$	S	1.3	1.2	3.5	4.0	-3.99	5.00	-16793.2 (335.4)	-235.3 (0.1)	0.0 (0.0)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.0)	0.24 (10.09)	0.24 (1.00)	-	3.9 (0.0)
$W^{1.3}_{1.3}$	S	1.3	1.3	3.8	4.3	-4.32	5.41	-18230.7 (612.7)	-295.6 (0.4)	0.0 (0.0)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.0)	0.24 (11.93)	0.24 (1.09)	-	3.7 (0.0)
$W^{1.3}_{1.4}$	S	1.3	1.4	4.1	4.7	-4.66	5.83	-22987.6 (454.4)	-366.4 (0.0)	0.0 (0.0)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.0)	0.20 (1.62)	0.23 (1.17)	-	3.8 (0.0)
$W^{1.3}_{1.5}$	S	1.3	1.5	4.4	5.0	-4.99	6.24	-24555.5 (201.2)	-455.1 (0.4)	0.0 (0.0)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.0)	0.19 (4.86)	0.23 (1.09)	-	3.6 (0.0)
$W^{1.4}_{0.1}$	L	1.4	0.1	0.3	0.4	-0.41	0.52	9738.4 (1.7)	-2.0 (0.0)	7.8 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.31 (0.00)	1.4 (0.0)	1.5 (0.0)
$W^{1.4}_{0.2}$	L	1.4	0.2	0.6	0.8	-0.82	1.04	9189.2 (1.8)	-3.0 (0.0)	7.8 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.47)	0.31 (0.02)	2.8 (0.0)	1.6 (0.0)
$W^{1.4}_{0.3}$	L	1.4	0.3	1.0	1.2	-1.24	1.56	7793.4 (0.8)	-5.9 (0.0)	8.4 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.43)	0.30 (0.05)	6.1 (0.0)	1.9 (0.0)
$W^{1.4}_{0.4}$	L	1.4	0.4	1.3	1.5	-1.66	2.08	5557.3 (0.7)	-11.5 (0.0)	9.6 (0.2)	<0.2	<0.2	1.0 (0.0)	0.31 (2.30)	0.29 (0.11)	11.9 (0.1)	2.3 (0.0)
$W^{1.4}_{0.5}$	L	1.4	0.5	1.6	1.9	-2.07	2.60	2716.5 (1.4)	-21.7 (0.0)	9.2 (0.2)	0.4	0.4	1.3 (0.2)	0.29 (2.17)	0.28 (0.21)	21.3 (0.1)	3.0 (0.0)
$W^{1.4}_{0.6}$	L	1.4	0.6	1.9	2.3	-2.49	3.12	-144.2 (1.1)	-40.9 (0.0)	2.3 (0.0)	3.9	3.8	8.3 (0.2)	0.27 (3.10)	0.27 (0.39)	35.2 (0.4)	3.6 (0.0)
$W^{1.4}_{0.7}$	S	1.4	0.7	2.2	2.7	-2.90	3.64	-4399.2 (414.3)	-72.1 (0.1)	0.0 (0.0)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.0)	0.27 (4.94)	0.26 (0.57)	-	3.9 (0.0)
$W^{1.4}_{0.8}$	S	1.4	0.8	2.5	3.1	-3.32	4.16	-6004.6 (128.3)	-106.5 (0.1)	0.0 (0.0)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.0)	0.26 (6.55)	0.26 (0.76)	-	4.0 (0.0)
$W^{1.4}_{0.9}$	S	1.4	0.9	2.9	3.5	-3.73	4.68	-11148.7 (384.9)	-148.7 (0.1)	0.0 (0.0)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.0)	0.25 (7.85)	0.25 (0.86)	-	3.9 (0.0)
$W^{1.4}_{1.0}$	S	1.4	1.0	3.2	3.9	-4.16	5.20	-11407.4 (369.6)	-202.5 (0.1)	0.0 (0.0)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.0)	0.21 (2.99)	0.24 (0.97)	-	3.9 (0.0)
$W^{1.4}_{1.1}$	S	1.4	1.1	3.5	4.2	-4.57	5.72	-13555.8 (197.7)	-279.4 (0.2)	0.0 (0.0)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.0)	0.20 (5.16)	0.24 (0.92)	-	3.9 (0.0)
$W^{1.4}_{1.2}$	S	1.4	1.2	3.8	4.6	-4.98	6.24	-18550.8 (337.4)	-379.0 (0.4)	0.0 (0.0)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.0)	0.20 (6.86)	0.23 (0.96)	-	3.7 (0.0)
$W^{1.4}_{1.3}$	S	1.4	1.3	4.1	5.0	-5.40	6.76	-20910.1 (161.9)	-520.6 (0.2)	0.0 (0.0)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.0)	0.20 (10.63)	0.23 (1.10)	-	4.0 (0.0)
$W^{1.4}_{1.4}$	S	1.4	1.4	4.5	5.4	-5.82	7.28	-25656.4 (858.0)	-695.0 (0.5)	0.0 (0.0)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.0)	0.19 (12.56)	0.22 (1.22)	-	3.8 (0.0)
$W^{1.4}_{1.5}$	S	1.4	1.5	4.8	5.8	-6.24	7.80	-32735.8 (290.6)	-920.4 (1.0)	0.0 (0.0)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.0)	0.19 (15.17)	0.22 (1.16)	-	4.0 (0.0)
$W^{1.5}_{0.1}$	L	1.5	0.1	0.3	0.4	-0.50	0.64	9696.8 (1.2)	-2.1 (0.0)	7.7 (0.1)	<0.2	<0.2	1.0 (0.0)	0.32 (2.49)	0.31 (0.00)	1.4 (0.0)	1.5 (0.0)
$W^{1.5}_{0.2}$	L	1.5	0.2	0.7	0.9	-1.02	1.28	8917.9 (0.7)	-3.5 (0.0)	8.0 (0.2)	<0.2	<0.2	1.0 (0.0)	0.32 (2.46)	0.31 (0.02)	3.1 (0.0)	1.6 (0.0)
$W^{1.5}_{0.3}$	L	1.5	0.3	1.0	1.3	-1.53	1.92	7091.6 (0.9)	-7.3 (0.0)	8.6 (0.1)	<0.2	<0.2	1.0 (0.0)	0.31 (2.39)	0.30 (0.06)	6.8 (0.0)	2.0 (0.0)
$W^{1.5}_{0.4}$	L	1.5	0.4	1.4	1.8	-2.04	2.56	4275.4 (1.7)	-15.2 (0.0)	9.6 (0.3)	0.2	0.2	1.3 (0.4)	0.30 (2.18)	0.29 (0.13)	13.6 (0.1)	2.6 (0.0)
$W^{1.5}_{0.5}$	L	1.5	0.5	1.7	2.2	-2.55	3.19	904.6 (1.2)	-31.6 (0.0)	4.9 (0.2)	1.0	0.9	3.3 (0.2)	0.28 (2.57)	0.28 (0.29)	24.9 (0.3)	3.4 (0.0)
$W^{1.5}_{0.6}$	L	1.5	0.6	2.0	2.7	-3.06	3.84	-3213.0 (13.2)	-63.0 (0.0)	0.0 (0.0)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.0)	0.27 (4.46)	0.27 (0.51)	31.0 (4.0)	3.9 (0.0)
$W^{1.5}_{0.7}$	S	1.5	0.7	2.4	3.1	-3.58	4.47	-2628.6 (258.5)	-103.9 (0.1)	0.0 (0.0)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.0)	0.22 (3.09)	0.26 (0.68)	-	4.0 (0.0)
$W^{1.5}_{0.8}$	S	1.5	0.8	2.7	3.5	-4.09	5.12	-4979.0 (227.7)	-166.0 (0.2)	0.0 (0.0)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.0)	0.22 (5.25)	0.25 (0.72)	-	4.0 (0.0)
$W^{1.5}_{0.9}$	S	1.5	0.9	3.1	4.0	-4.59	5.75	-6595.6 (170.1)	-248.5 (0.2)	0.0 (0.0)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.0)	0.21 (5.97)	0.24 (0.80)	-	4.0 (0.0)
$W^{1.5}_{1.0}$	S	1.5	1.0	3.4	4.4	-5.11	6.39	-10607.3 (149.6)	-374.3 (0.8)	0.0 (0.0)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.0)	0.20 (8.50)	0.24 (0.84)	-	3.8 (0.3)
$W^{1.5}_{1.1}$	S	1.5	1.1	3.8	4.9	-5.62	7.03	-11938.9 (622.7)	-541.2 (1.3)	0.0 (0.0)	$>21\cdot10^3$	$>21\cdot10^3$	$>21\cdot10^3$ (0.0)	0.19 (10.67)	0.23 (0.87)	-	4.0 (0.0)

Table 2.S.2: Simulated properies of the water-like models of the NPT series at 298.15 K and 1 bar. For each model, the state of the system (St.) is indicated as L (liquid) or S (solid). The quantities reported are the scaling factor for the bond length s_b, the scaling factor for the partial charges s_q, the scaling factor for the repulsive Lennard-Jones parameter $C_{12}^{1/2}$ s_c, the density ρ of the system, the dipole moment μ of the model, the square component Q_T of the quadrupole, the linear component Ω_0 of the octupole, the square component Ω_T of the octupole, the total potential energy U, the self-diffusion constant D, the rotational correlation time τ_2^{OH} for the OH axis, the rotational correlation time τ_2^{HH} for the OH axis, the Debye relaxation time τ_D, the position r_{max} and the height g_{max} of the first peak of the radial distribution function $g(r)$, the position r_{max} and the height c_{max} of the first peak of the dipole-dipole orientation function $c(r)$, the static relative dielectric permittivity ε and the hydrogen-bonding capacity n_H. The quantities P, U, D, τ_D, ε and n_H are reported with a statistical error (between parenthesis) calculated as th root-mean-square deviation over five repeats of the simulations. The models that were not feasible are not reported, a blank line is denoting one or more missing combinations of scaling factors.

	St.	s_b	s_q	s_c	ρ [kg·m^{-3}]	U [kJ·mol^{-1}]	D [10^{-9}m^{2}s^{-1}]	τ_2^{OH} [ps]	τ_2^{HH} [ps]	τ_D [ps]	r_{max} (g_{max}) [nm] (-)	r_{max} (c_{max}) [nm] (-)	ε	n_H
$Y^{0.3}_{1.5}$	L	0.3	1.5	0.588235	1102.69 (0.02)	-12.1 (0.0)	13.2 (0.2)	<0.2	<0.2	1.0 (0.0)	0.28 (2.31)	0.28 (0.12)	9.1 (0.1)	1.8 (0.0)
$Y^{0.4}_{1.3}$	L	0.4	1.3	0.614666	1035.16 (0.07)	-12.7 (0.0)	15.9 (0.4)	<0.2	<0.2	1.0 (0.0)	0.28 (2.23)	0.28 (0.16)	13.0 (0.1)	1.7 (0.0)
$Y^{0.4}_{1.4}$	L	0.4	1.4	0.640779	978.01 (0.26)	-12.8 (0.0)	17.4 (0.3)	<0.2	<0.2	1.0 (0.0)	0.29 (2.19)	0.28 (0.18)	15.3 (0.1)	1.6 (0.0)
$Y^{0.4}_{1.5}$	L	0.4	1.5	0.650026	1009.58 (0.13)	-14.2 (0.0)	16.3 (0.3)	<0.2	<0.2	1.0 (0.0)	0.29 (2.22)	0.28 (0.21)	20.4 (0.1)	1.7 (0.0)
$Y^{0.5}_{1.2}$	L	0.5	1.2	0.653125	1007.07 (0.09)	-14.4 (0.0)	16.6 (0.4)	<0.2	<0.2	1.0 (0.0)	0.28 (2.22)	0.28 (0.20)	20.6 (0.2)	1.8 (0.0)
$Y^{0.5}_{1.3}$	L	0.5	1.3	0.677782	998.52 (0.07)	-15.7 (0.0)	17.1 (0.4)	<0.2	<0.2	1.0 (0.0)	0.29 (2.22)	0.28 (0.23)	26.5 (0.1)	1.8 (0.0)
$Y^{0.5}_{1.4}$	L	0.5	1.4	0.702395	996.74 (0.06)	-17.2 (0.0)	16.7 (0.6)	<0.2	<0.2	1.0 (0.0)	0.29 (2.24)	0.28 (0.26)	36.5 (0.2)	1.9 (0.0)
$Y^{0.5}_{1.5}$	L	0.5	1.5	0.727061	998.82 (0.06)	-18.9 (0.0)	16.3 (0.5)	<0.2	<0.2	1.0 (0.0)	0.29 (2.24)	0.28 (0.29)	48.6 (0.3)	1.9 (0.0)
$Y^{0.6}_{0.9}$	L	0.6	0.9	0.624766	1045.75 (0.05)	-13.8 (0.0)	15.9 (0.2)	<0.2	<0.2	1.0 (0.0)	0.28 (2.25)	0.28 (0.17)	15.0 (0.1)	1.9 (0.0)
$Y^{0.6}_{1.0}$	L	0.6	1.0	0.659283	1006.16 (0.07)	-14.7 (0.0)	16.7 (0.7)	<0.2	<0.2	1.0 (0.0)	0.29 (2.23)	0.28 (0.20)	20.0 (0.2)	1.9 (0.0)
$Y^{0.6}_{1.1}$	L	0.6	1.1	0.690084	999.83 (0.04)	-16.3 (0.0)	16.7 (0.5)	<0.2	<0.2	1.0 (0.0)	0.29 (2.23)	0.28 (0.23)	27.6 (0.3)	1.9 (0.0)
$Y^{0.6}_{1.2}$	L	0.6	1.2	0.721501	1000.14 (0.13)	-18.4 (0.0)	16.3 (0.4)	<0.2	<0.2	1.0 (0.0)	0.29 (2.24)	0.28 (0.27)	38.2 (0.3)	2.0 (0.0)
$Y^{0.6}_{1.3}$	L	0.6	1.3	0.75171	1008.23 (0.05)	-20.8 (0.0)	15.9 (0.3)	<0.2	<0.2	1.0 (0.0)	0.29 (2.27)	0.28 (0.30)	54.1 (0.4)	2.1 (0.0)
$Y^{0.6}_{1.4}$	L	0.6	1.4	0.788644	1001.40 (0.03)	-23.3 (0.0)	15.0 (0.5)	<0.2	<0.2	1.0 (0.0)	0.28 (2.26)	0.28 (0.34)	74.4 (0.8)	2.2 (0.0)
$Y^{0.6}_{1.5}$	L	0.6	1.5	0.828706	991.74 (0.05)	-25.9 (0.0)	14.3 (0.3)	<0.2	<0.2	1.5 (0.4)	0.28 (2.29)	0.28 (0.37)	98.8 (0.6)	2.3 (0.0)

Continued on Next Page…

Table 2.S.2 – Continued

	St.	s_b	s_q	s_c	ρ [kg·m^{-3}]	U [kJ·mol^{-1}]	D [10^{-9}m^{2}s^{-1}]	τ_2^{OH} [ps]	τ_2^{HH} [ps]	τ_D [ps]	r_{max} (g_{max}) [nm] (-)	r_{max} (c_{max}) [nm] (-)	ε	n_H
$Y^{0.8}_{0.7}$	L	0.7	0.8	0.650026	997.11 (0.11)	-13.8 (0.0)	17.6 (0.6)	<0.2	<0.2	1.0 (0.0)	0.28 (2.21)	0.28 (0.17)	15.3 (0.2)	1.9 (0.0)
$Y^{0.9}_{0.7}$	L	0.7	0.9	0.683901	1002.33 (0.07)	-15.8 (0.0)	16.8 (0.2)	<0.2	<0.2	1.0 (0.0)	0.28 (2.23)	0.28 (0.21)	22.6 (0.1)	2.0 (0.0)
$Y^{1.0}_{0.7}$	L	0.7	1.0	0.727061	987.82 (0.06)	-17.9 (0.0)	17.0 (0.7)	<0.2	<0.2	1.0 (0.0)	0.29 (2.23)	0.28 (0.25)	31.3 (0.2)	2.1 (0.0)
$Y^{1.1}_{0.7}$	L	0.7	1.1	0.769231	987.45 (0.03)	-20.7 (0.0)	16.3 (0.3)	<0.2	<0.2	1.0 (0.0)	0.29 (2.24)	0.28 (0.29)	44.3 (0.5)	2.2 (0.0)
$Y^{1.2}_{0.7}$	L	0.7	1.2	0.807168	1000.78 (0.03)	-24.2 (0.0)	15.0 (0.2)	<0.2	<0.2	1.2 (0.6)	0.28 (2.27)	0.28 (0.33)	63.3 (0.4)	2.4 (0.0)
$Y^{1.3}_{0.7}$	L	0.7	1.3	0.850268	1002.43 (0.05)	-27.9 (0.0)	13.1 (0.3)	<0.2	<0.2	1.3 (0.4)	0.28 (2.31)	0.28 (0.37)	84.7 (1.1)	2.6 (0.0)
$Y^{1.4}_{0.7}$	L	0.7	1.4	0.899604	994.08 (0.05)	-31.9 (0.0)	11.8 (0.3)	<0.2	0.4	1.8 (0.2)	0.28 (2.38)	0.28 (0.41)	111.8 (1.3)	2.7 (0.0)
$Y^{1.5}_{0.7}$	L	0.7	1.5	0.936505	1005.11 (0.06)	-37.1 (0.0)	9.0 (0.2)	0.4	0.4	2.4 (0.1)	0.28 (2.49)	0.28 (0.46)	147.3 (1.8)	2.9 (0.0)
$Y^{0.7}_{0.8}$	L	0.8	0.7	0.656211	1008.52 (0.06)	-14.3 (0.0)	16.6 (0.4)	<0.2	<0.2	1.0 (0.0)	0.29 (2.22)	0.28 (0.17)	15.1 (0.2)	2.0 (0.0)
$Y^{0.8}_{0.8}$	L	0.8	0.8	0.702395	998.24 (0.05)	-16.6 (0.0)	16.5 (0.6)	<0.2	<0.2	1.0 (0.0)	0.29 (2.23)	0.28 (0.21)	22.2 (0.1)	2.1 (0.0)
$Y^{0.9}_{0.8}$	L	0.8	0.9	0.745545	1013.05 (0.04)	-20.0 (0.0)	15.5 (0.4)	<0.2	<0.2	1.0 (0.0)	0.28 (2.24)	0.28 (0.25)	33.8 (0.2)	2.3 (0.0)
$Y^{1.0}_{0.8}$	L	0.8	1.0	0.800961	1001.07 (0.05)	-23.4 (0.0)	15.0 (0.3)	<0.2	<0.2	1.2 (0.6)	0.28 (2.25)	0.28 (0.30)	46.9 (0.5)	2.5 (0.0)
$Y^{1.1}_{0.8}$	L	0.8	1.1	0.856458	995.47 (0.02)	-27.6 (0.0)	13.0 (0.1)	<0.2	<0.2	1.6 (0.5)	0.28 (2.30)	0.28 (0.35)	62.8 (0.9)	2.7 (0.0)
$Y^{1.2}_{0.8}$	L	0.8	1.2	0.909091	995.85 (0.04)	-32.7 (0.0)	10.4 (0.3)	<0.2	0.4	1.8 (0.1)	0.28 (2.41)	0.28 (0.40)	82.4 (1.5)	2.9 (0.0)
$Y^{1.3}_{0.8}$	L	0.8	1.3	0.955019	1002.86 (0.04)	-39.1 (0.0)	7.3 (0.2)	0.6	0.8	3.1 (0.1)	0.28 (2.60)	0.28 (0.46)	104.2 (1.4)	3.2 (0.0)
$Y^{1.4}_{0.8}$	L	0.8	1.4	1.00614	998.48 (0.08)	-46.2 (0.0)	4.3 (0.1)	1.5	1.7	5.5 (0.2)	0.28 (2.88)	0.27 (0.53)	121.1 (3.8)	3.4 (0.0)
$Y^{1.5}_{0.8}$	L	0.8	1.5	1.04745	999.16 (0.10)	-55.0 (0.0)	1.7 (0.0)	3.8	3.9	13.6 (0.5)	0.28 (3.31)	0.27 (0.63)	140.1 (8.8)	3.6 (0.0)
$Y^{0.6}_{0.9}$	L	0.9	0.6	0.656211	1006.37 (0.10)	-14.1 (0.0)	16.6 (0.5)	<0.2	<0.2	1.0 (0.0)	0.29 (2.22)	0.28 (0.15)	13.1 (0.2)	2.0 (0.0)
$Y^{0.7}_{0.9}$	L	0.9	0.7	0.708567	1003.81 (0.03)	-16.9 (0.0)	16.1 (0.5)	<0.2	<0.2	1.0 (0.0)	0.29 (2.21)	0.28 (0.20)	20.4 (0.2)	2.2 (0.0)
$Y^{0.8}_{0.9}$	L	0.9	0.8	0.769231	995.88 (0.06)	-20.5 (0.0)	15.8 (0.3)	<0.2	<0.2	1.0 (0.0)	0.28 (2.22)	0.28 (0.25)	30.4 (0.5)	2.5 (0.0)
$Y^{0.9}_{0.9}$	L	0.9	0.9	0.828706	999.88 (0.03)	-25.2 (0.0)	13.5 (0.3)	<0.2	<0.2	1.6 (0.7)	0.28 (2.26)	0.28 (0.30)	44.0 (0.7)	2.7 (0.0)
$Y^{1.0}_{0.9}$	L	0.9	1.0	0.887233	1004.30 (0.05)	-31.2 (0.0)	10.5 (0.3)	<0.2	0.4	1.7 (0.2)	0.28 (2.38)	0.28 (0.37)	58.5 (0.5)	3.0 (0.0)
$Y^{1.1}_{0.9}$	L	0.9	1.1	0.948857	999.37 (0.02)	-38.1 (0.0)	6.8 (0.3)	1.2	1.0	2.9 (0.1)	0.28 (2.61)	0.27 (0.44)	74.5 (0.7)	3.3 (0.0)
$Y^{0.6}_{1.0}$	L	1.0	0.6	0.702395	1010.77 (0.05)	-16.6 (0.0)	16.1 (0.3)	<0.2	<0.2	1.0 (0.0)	0.28 (2.20)	0.28 (0.18)	17.3 (0.1)	2.3 (0.0)
$Y^{0.7}_{1.0}$	L	1.0	0.7	0.770179	1006.55 (0.03)	-20.9 (0.0)	14.9 (0.4)	<0.2	<0.2	1.0 (0.0)	0.28 (2.22)	0.28 (0.24)	27.0 (0.3)	2.6 (0.0)
$Y^{0.8}_{1.0}$	L	1.0	0.8	0.844096	998.82 (0.04)	-26.3 (0.0)	12.2 (0.3)	<0.2	<0.2	1.3 (0.4)	0.28 (2.28)	0.28 (0.30)	38.4 (0.2)	2.9 (0.0)
$Y^{0.9}_{1.0}$	L	1.0	0.9	0.909091	1005.86 (0.03)	-33.9 (0.0)	7.9 (0.0)	0.9	0.8	2.3 (0.2)	0.28 (2.50)	0.27 (0.38)	53.4 (0.8)	3.2 (0.0)
$Y^{1.0}_{1.0}$	L	1.0	1.0	0.97352	1002.91 (0.04)	-43.4 (0.0)	3.9 (0.1)	1.8	1.9	5.6 (0.3)	0.27 (2.93)	0.27 (0.48)	66.7 (1.7)	3.5 (0.0)
$Y^{0.5}_{1.1}$	L	1.1	0.5	0.690084	990.52 (0.07)	-15.0 (0.0)	16.7 (0.2)	<0.2	<0.2	1.0 (0.0)	0.29 (2.19)	0.28 (0.15)	12.7 (0.1)	2.2 (0.0)
$Y^{0.6}_{1.1}$	L	1.1	0.6	0.764	998.24 (0.03)	-19.8 (0.0)	15.3 (0.2)	<0.2	<0.2	1.0 (0.0)	0.28 (2.19)	0.28 (0.21)	21.4 (0.2)	2.6 (0.0)
$Y^{0.7}_{1.1}$	L	1.1	0.7	0.844096	1000.47 (0.05)	-26.3 (0.0)	11.4 (0.3)	<0.2	<0.2	1.4 (0.2)	0.28 (2.28)	0.28 (0.29)	32.2 (0.1)	3.0 (0.0)
$Y^{0.8}_{1.1}$	L	1.1	0.8	0.924214	998.33 (0.04)	-35.0 (0.0)	6.5 (0.1)	0.8	1.0	3.0 (0.2)	0.28 (2.60)	0.27 (0.38)	45.7 (0.5)	3.3 (0.0)
$Y^{0.9}_{1.1}$	L	1.1	0.9	0.985804	1005.25 (0.08)	-47.6 (0.0)	1.9 (0.0)	3.6	3.8	10.4 (0.3)	0.27 (3.25)	0.27 (0.51)	59.3 (1.0)	3.7 (0.0)
$Y^{0.5}_{1.2}$	L	1.2	0.5	0.742446	986.26 (0.06)	-17.7 (0.0)	15.6 (0.4)	<0.2	<0.2	1.0 (0.0)	0.28 (2.16)	0.28 (0.17)	15.4 (0.1)	2.5 (0.0)
$Y^{0.6}_{1.2}$	L	1.2	0.6	0.828706	1001.78 (0.03)	-24.8 (0.0)	11.8 (0.2)	<0.2	<0.2	1.4 (0.3)	0.28 (2.25)	0.27 (0.26)	26.0 (0.2)	3.0 (0.0)

Continued on Next Page. . .

Table 2.S.2 – Continued

	St.	s_b	s_q	s_c	ρ [kg·m^{-3}]	U [kJ·mol^{-1}]	D [10^{-9}m^{2}s^{-1}]	τ_2^{OH} [ps]	τ_2^{HH} [ps]	τ_D [ps]	$r_{max}(g_{max})$ [nm] (-)	$r_{max}(c_{max})$ [nm] (-)	ε	n_H
$Y^{1.2}_{0.7}$	L	1.2	0.7	0.924214	994.02 (0.03)	-34.3 (0.0)	6.1 (0.2)	1.1	1.1	2.9 (0.1)	0.28 (2.62)	0.27 (0.37)	38.8 (0.5)	3.4 (0.0)
$Y^{1.2}_{0.8}$	L	1.2	0.8	1	996.47 (0.03)	-48.6 (0.0)	1.3 (0.0)	5.1	5.3	15.2 (0.8)	0.27 (3.41)	0.27 (0.51)	52.7 (1.6)	3.7 (0.0)
$Y^{1.3}_{0.4}$	L	1.3	0.4	0.690084	1025.86 (0.04)	-15.8 (0.0)	15.3 (0.4)	<0.2	<0.2	1.0 (0.0)	0.28 (2.20)	0.27 (0.12)	10.8 (0.0)	2.5 (0.0)
$Y^{1.3}_{0.5}$	L	1.3	0.5	0.796813	999.88 (0.04)	-21.9 (0.0)	12.7 (0.3)	<0.2	<0.2	1.0 (0.3)	0.28 (2.19)	0.27 (0.21)	19.0 (0.2)	2.9 (0.0)
$Y^{1.3}_{0.6}$	L	1.3	0.6	0.899604	1003.83 (0.03)	-32.1 (0.0)	6.5 (0.1)	0.8	1.1	2.6 (0.2)	0.28 (2.54)	0.27 (0.33)	31.4 (0.5)	3.4 (0.0)
$Y^{1.3}_{0.7}$	L	1.3	0.7	1	994.94 (0.06)	-46.9 (0.0)	1.4 (0.0)	5.0	5.3	14.5 (0.5)	0.27 (3.38)	0.27 (0.48)	42.4 (1.1)	3.7 (0.0)
$Y^{1.4}_{0.4}$	L	1.4	0.4	0.745545	996.21 (0.03)	-17.9 (0.0)	14.7 (0.3)	<0.2	<0.2	1.0 (0.0)	0.28 (2.14)	0.27 (0.14)	12.4 (0.1)	2.7 (0.0)
$Y^{1.4}_{0.5}$	L	1.4	0.5	0.862589	1001.67 (0.05)	-27.5 (0.0)	8.3 (0.1)	0.7	0.6	1.7 (0.3)	0.28 (2.34)	0.27 (0.26)	22.6 (0.3)	3.2 (0.0)
$Y^{1.4}_{0.6}$	L	1.4	0.6	0.985804	993.20 (0.04)	-42.0 (0.0)	2.1 (0.0)	3.3	3.3	9.1 (0.2)	0.27 (3.12)	0.27 (0.41)	35.9 (0.9)	3.7 (0.0)
$Y^{1.5}_{0.3}$	L	1.5	0.3	0.666667	1042.25 (0.07)	-14.6 (0.0)	14.4 (0.3)	<0.2	<0.2	1.0 (0.0)	0.28 (2.18)	0.27 (0.07)	7.3 (0.0)	2.5 (0.0)
$Y^{1.5}_{0.4}$	L	1.5	0.4	0.800961	998.09 (0.05)	-21.7 (0.0)	11.4 (0.1)	<0.2	0.4	1.4 (0.2)	0.28 (2.14)	0.27 (0.18)	14.3 (0.2)	3.0 (0.0)
$Y^{1.5}_{0.5}$	L	1.5	0.5	0.936505	1005.34 (0.03)	-35.6 (0.0)	3.8 (0.1)	1.7	1.8	4.6 (0.2)	0.27 (2.78)	0.27 (0.33)	26.4 (0.3)	3.5 (0.0)

Solvent-modulated influence of intramolecular hydrogen-bonding on the conformational properties of the hydroxymethyl group in glucose and galactose: A molecular dynamics simulation study.

Abstract

Intramolecular hydrogen-bonding (H-bonding) is commonly regarded as a major determinant of the conformation of (bio)molecules. However, in an aqueous environment, solvent-exposed H-bonds are likely to represent only a marginal (possibly adverse) conformational driving as well as steering force. For example, the hydroxymethyl rotamers of glucose (Glc) and galactose (Gal) permitting the formation of an H-bond with the adjacent hydroxyl group are not favored in water but, in the opposite, least populated. This is because in water, solvent-exposed H-bonds are efficiently screened as well as subject to intense H-bonding competition by the solvent molecules. In the present study, the effect of a decrease in the solvent polarity on this rotameric equilibrium is probed using molecular dynamics (MD) simulation. This is done by considering 6 physical solvents (H_2O, DMSO, CH_3OH, $CHCl_3$, CCl_4 and vacuum) along with 19 artificial water-like solvent models, where the dielectric permittivity and H-bonding capacity can be varied independently *via* a scaling of the oxygen-hydrogen distance and of the atomic partial charges. In the high polarity solvents, the intramolecular H-bond is observed, but arises as an opportunistic consequence of the proximity of the H-bonding partners in a given rotameric state. Only when the polarity of the solvent is decreased does the intramolecular H-bond start to induce a significant conformational pressure on the rotameric equilibrium. The artificial solvent series also reveals that the effects of the solvent permittivity and of its H-bonding capacity mutually enhance each other, with a slightly larger influence of the permittivity. The hydroxymethyl conformation in hexopyranoses also appears to be particularly sensitive to solvent polarity effects, as the H-bond involving the hydroxymethyl group is only one out of up to five H-bonds forming a network around the ring, each of which is also affected by the solvent polarity.

3.1 Introduction

It is often assumed that intramolecular hydrogen-bonding (H-bonding) exerts a significant influence on the conformational properties of aqueous (bio)polymers[126–128]. To discuss this statement, one should however distinguish[132,169–171] between solvent-exposed and buried H-bonds, and between their respective roles in promoting stability (*i.e.* as driving forces) and specificity (*i.e.* as steering forces; see the Appendix of Ref.[132] for a definition of the concept).

When a specific conformational change is accompanied by the formation of a buried H-bond, the cost associated with the removal of the H-bonding partners from the solution environment (desolvation) offsets the electrostatic gain upon formation of the interaction itself. However, if the desolvation of two potentially H-bonding groups is not accompanied by H-bond formation upon burial, the corresponding conformation will be penalized by the desolvation term. Considering the high polarity of water, the desolvation and electrostatic terms will typically be of comparable magnitudes. As a result, the formation of a buried H-bond can be thought of as representing a minor (possibly negligible or even, in some cases, adverse) conformational driving force, but an important conformational steering force[132].

When a specific conformational change is accompanied by the formation of a solvent-exposed H-bond, the H-bonding partners remain highly solvated, and their interaction is screened by the solvent dielectric response as well as subject to H-bonding competition by the solvent molecules. In the case of an aqueous environment, H-bonding interactions within the solute and between solute and solvent will typically be of comparable magnitudes. As a result, the formation of a solvent-exposed H-bond can be thought of as representing a minor (possibly negligible or even, in some cases, adverse) conformational driving as well as steering force[132,169–175]. According to this interpretation, solvent-exposed

intramolecular H-bonding in a specific conformation of an aqueous solute should be viewed as an opportunistic consequence of the close proximity of two H-bonding groups in this conformation[132,176,177], and not as a factor contributing to the stability of the conformation.

A prototypical example for this second situation is provided by the conformational properties of the hydroxymethyl group in aqueous hexopyranoses. As illustrated in Figure 3.1, the three staggered rotameric states of this group are labelled gg, gt and tg, the two successive letters referring to the values of the dihedral angles $\tilde{\omega}$ (O_5-C_5-C_6-O_6) and ω (C_4-C_5-C_6-O_6), namely $gauche$ (g) or $trans$ (t). For β-D-Glucopyranose (Glc), the hydroxyl group at the adjacent carbon atom C_4 is equatorial in the most stable $^4C_1$ chair conforma-
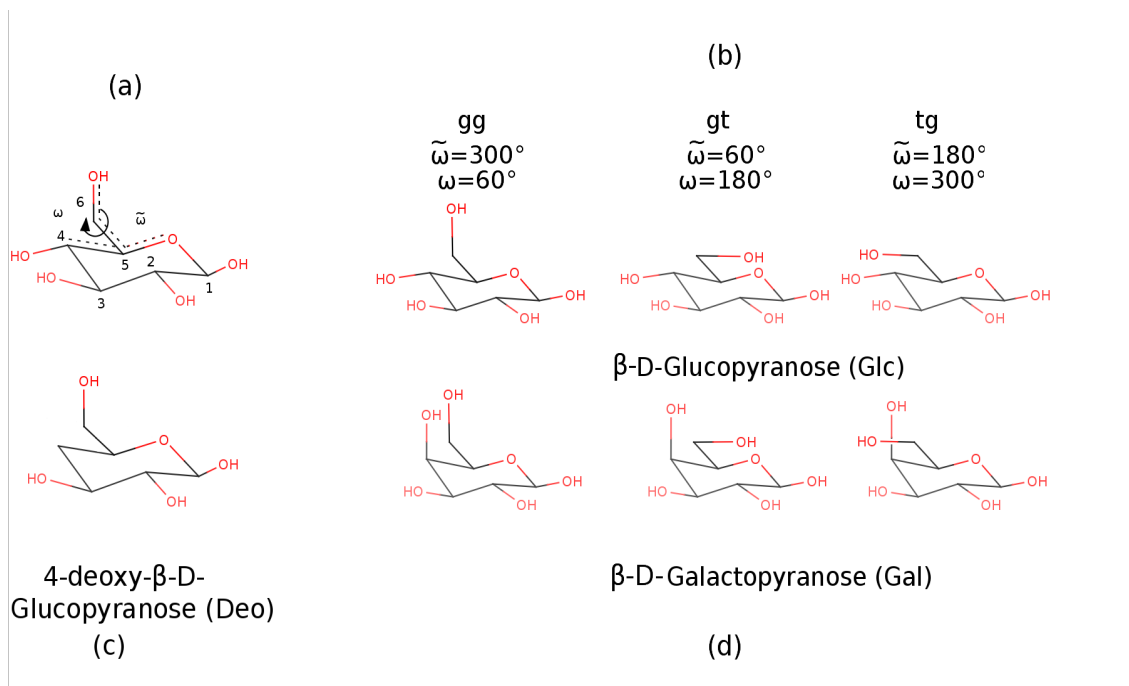
Figure 3.1: Model compounds considered in the present study, dihedral-angle definitions and canonical conformations of the hydroxymethyl group. The hexopyranoses considered are β-D-Glucopyranose (Glc; a, b), β-D-Galactopyranose (Gal; d) and 4-deoxy-β-D-Glucopyranose (Deo; c). They are drawn in their most stable $^4C_1$ chair conformation. The atom numbering and the definition of the dihedral angles $\tilde{\omega}$ (O_5-C_5-C_6-O_6) and ω (C_4-C_5-C_6-O_6) are shown (a), as well as the canonical rotameric conformations of the hydroxymethyl group (b,d).

tion. As a result, the tg rotamer enables the formation of a solvent-exposed intramolecular H-bond between these two groups ($H_4 \rightarrow O_6$ or $H_6 \rightarrow O_4$). For β-D-Galactopyranose (Gal), the hydroxyl group at C_4 is axial in the $^4C_1$ chair conformation. As a result, it is the gg rotamer that enables the formation of this H-bond.

For the two compounds, the rotameric distributions in dilute aqueous solution inferred from the most recent nuclear magnetic resonance (NMR) experiments [184,185,191] are reported in Table 3.1 (see also Suppl. Mat. Tables 3.S.1 and 3.S.2 for a more extensive compilation of estimated populations and source J-coupling constants, respectively). They suggest

Glc			
Year/Ref.	gg [%]	gt [%]	tg [%]
1975[178]([179])	49	49	2
1984[180]	53	45	2
1988[181]	53	45	2
1994[182]([181])	49	49	2
2000[183]	45	62	-7
2004[184]	31	61	8
2006[185]	31	59	10

Gal			
Year/Ref.	gg [%]	gt [%]	tg [%]
1983[186]	27	56	17
1984[180]([181])	22	53	25
1987[187]	22	53	25
1987[188]	14	65	21
1988[181]	18	62	20
1994[189]	12	56	32
2001[190]	15	71	14
2002[191]	3	67	30
2004[184]	3	72	25
2004[184]	0	71	29

Table 3.1: Experimentally inferred population estimates for the three staggered rotamers of the hydroxymethyl group of Glc and Gal. The data pertains to aqueous solution and to the β-anomer of the unmethylated compound. Secondary references between parentheses indicate that the raw data of the primary reference is quoted as analyzed and reported in the secondary reference. A more extensive compilation of population estimates along with source J-coupling constants (also including α-anomers, O_1-methylated derivatives and non-aqueous solvents) is provided in Suppl. Mat. Tables 3.S.1 and 3.S.2.

relative gg:gt:tg populations (in percent) on the order of 30:60:10 for Glc and 5:70:25 for Gal. In other words, the rotamers permitting the formation of a solvent-exposed H-bond between the hydroxymethyl group and the hydroxyl group at C_4 are not favored but, in the opposite, least populated.

This observation is compatible with a weak (here, adverse) influence of solvent-exposed H-bonding on the conformational preferences of the hydroxymethyl group in water. However, it could also be argued that the experimentally observed preferences actually result from other stronger effects of steric, stereoelectronic or/and electrostatic nature, and thus do not permit to draw any conclusion on the specific contribution of intramolecular H-bonding. For example, the stereoelectronic $gauche$-effect[182,188,192–206] will favor the gg and gt conformations for both compounds, and intramolecular H-bonds between the hydroxymethyl group and the ring oxygen atom O_5 are also in principle possible in these two conformations. Steric and bond-dipole 1,3-syn repulsions between the C_6-O_6 and C_4-O_4 bonds may also be invoked to explain the destabilization of the tg rotamer for Glc and of the gg rotamer for Gal, if not compensated for by sufficiently strong H-bonding.

Experimentally, different approaches have been employed to tentatively disentangle the specific role of intramolecular H-bonding from these other effects. The most common ones involve the consideration of model compounds with altered (non-H-bonding) substitutions[179,182,188–190,200,202,204,207–213] or of non-aqueous solvents with lower polarities[175,179,182,188,200,202,204,207,211,214]. Unfortunately, both types of changes do not solely affect intramolecular H-bonding. For example, the deletion of a hydroxyl group converts a polar to a non-polar site (aliphatic group), significantly affecting the local solvation properties, while its replacement by a fluoride atom, although presumably more conservative in terms of solvation, still affects the steric and stereoelectronic properties of the molecule. Similarly, the use of non-aqueous solvents changes not only the dielectric permittivity of the medium, but simultaneously many other properties of the solvent molecules such as their

H-bond donor and acceptor capacities, dispersive interactions, size and shape. As a result, the contributions of the different effects still remain difficult to disentangle.

Computationally, numerous studies have investigated the problem of the hydroxymethyl group rotation in hexopyranoses, at the quantum-mechanical[190,215–232] (QM) and classical[185,211,233–238] levels. Although they introduce approximations inherent to the physical model employed, computer-simulation approaches offer three key advantages compared to experiment: (*i*) rotameric populations can be monitored directly rather than indirectly, *e.g.* by interpretation of experimental NMR J-coupling constants *via* a specific Karplus relationship[191,239–242] and an assumed functional form for the dihedral-angle probability distribution[189,191,204] (*e.g.* simple three-state population model); (*ii*) energetic contributions can be monitored individually, permitting to some extent an analysis of conformational trends in terms of underlying energy components; (*iii*) artificial (unphysical) situations can also be considered.

Because the consideration of unphysical situations in simulations may help disentangling effects relevant to physical ones, we have recently developed (see Chapter 2) artificial solvent models derived from the simple point charges (SPC) water model[136] by systematic variation of the oxygen-hydrogen bond length and of the atomic partial charges. These variations allow for a separate modulation of the dielectric permittivity and H-bonding capacity of the solvent, while preserving water-like dispersive interactions as well as molecular size and shape.

In the present work, these artificial solvent models are used to investigate the specific role of solvent-exposed intramolecular H-bonding on the rotameric properties of the hydroxymethyl group in Glc and Gal, by monitoring the sensitivity of these properties to the solvent permittivity and H-bonding capacity. To disentangle the influence of intramolecular H-bonding from that of the *gauche*-effect, which is also solvent-

dependent[170,200,204,206,243], a third compound lacking the hydroxyl group at C_4 is also considered, namely 4-deoxy-β-D-Glucopyranose (Deo), also shown in Figure 3.1. Molecular dynamics (MD) simulations with the local-elevation umbrella-sampling[14] (LEUS) method to enhance the sampling efficiency along the dihedral angle ω are performed using the GROMOS 53A6 force field[244] and 56A6$_{CARBO}$ force field[245,246], considering different physical and artificial solvent models. The physical solvents considered are water (H_2O), dimethyl-sulfoxide (DMSO), methanol (CH_3OH), chloroform (CH_3Cl) and carbon-tetrachloride (CCl_4), along with vacuum (VAC). The artificial solvents considered are a subset of 19 among the above-mentioned water-like models with tunable permittivity and H-bonding capacity (see Chapter 2).

3.2 Computational details

3.2.1 Simulated systems

All MD simulations were carried out using the GROMOS MD++ simulation program[81–84] along with two different versions of the GROMOS force field, namely the 53A6 parameter set[244] and the 56A6$_{CARBO}$ parameter set[245]. In the context of carbohydrates, the 53A6 parameter set[244] is equivalent to the most recent 54A7 set[50] and 54A8 set[247] of the GROMOS biomolecular force field, and essentially identical to the 45A4 set developed for hexopyranose-based carbohydrates by Lins & Hünenberger[248] in 2005 (see Ref.[245] for a description of the marginal differences). The 56A6$_{CARBO}$ parameter set of Hansen & Hünenberger[245] results from a complete reoptimization of the 53A6 set in the context of hexopyranose-based carbohydrates, carried out in 2011. Note that this set was recently revised[246] to a new set 56A6$_{CARBO\_R}$, but this change does not affect the compounds considered here.

Solvent	ρ [kg·m^{-3}]				μ [D]				ε				δH	
	Exp.	Ref.	Sim.	Ref.	Exp.	Ref.	Sim.	Ref.	Exp.	Ref.	Sim.	Ref.	Exp.	Ref.
H$_2$O	997	147	973	138	1.85	249	2.27	136	78.5	138	66.0	138	42.3	250
DMSO	1095	251	1096	149	3.96	166	5.25	149	46.0	251	38.0	149	10.2	250
CH$_3$OH	791	251	785	252	1.70	253	2.29	252	32.6	253	19.8	252	22.3	250
CHCl$_3$	1489	251	1520	254	4.81	251	2.30	255	4.8	253	2.4	254	5.7	250
CCl$_4$	1595	253	1601	256	0.00		0.00		2.2	166	1.0	256	-	
VAC	-		-		-		-		1.0		1.0		-	

Table 3.2: Experimental and simulated properties of the physical solvents considered in the present study at 298.15 K and 1 bar. The quantities reported are the density ρ, the gas-phase molecular dipole moment μ, the static relative dielectric permittivity ε, and the Hansen parameter[250] δH as a measure of the H-bonding propensity.

The simulations involved computational boxes containing one solute molecule and N_s solvent molecules. The three solutes considered are β-D-Glucopyranose (Glc), β-D-Galacto-pyranose (Gal) and 4-deoxy-β-D-Glucopyranose (Deo), see Figure 3.1. The first two compounds differ in the orientation of the hydroxyl group at C_4, equatorial for Glc and axial for Gal in the $^4C_1$ chair conformation, the most stable one in aqueous solution for these compounds[215,245,246,257–263]. The last compound lacks the hydroxyl group. In the 53A6 force field, the selection of the hydroxymethyl torsional potentials[248] depends on the orientation of the hydroxyl group at C_4. For this reason, the definition of a 53A6 topology for Deo would be ambiguous, and this compound was only simulated with the 56A6$_{CARBO}$ parameter set. The three solutes were simulated in various physical and artificial solvents.

For the physical solvents, the models considered are the SPC water (H_2O) model of Berendsen $et\ al.$[136] ($N_s = 1200$), the dimethyl-sulfoxide (DMSO) model of Geerke $et\ al.$[149] ($N_s = 310$), the methanol (CH_3OH) model of Walser $et\ al.$[252] (N_s=1200), the chloroform ($CHCl_3$) model of Tironi & van Gunsteren[254] ($N_s = 275$), and the carbon-tetrachloride (CCl_4) model of Tironi $et\ al.$[256] ($N_s = 230$). The three solutes were also simulated in vacuum (VAC), using stochastic dynamics[100,101] (SD) instead of MD in this specific case. Basic properties characterising the polarity and H-bonding capacity of the five solvents, either experimental or based on previous simulations with the indicated models, are summarized in Table 3.2.

The artificial solvents considered are a subset of the tunable solvent models introduced in Chapter 2 for use under constant-volume (NVT) conditions. These models were generated starting from the SPC water model[136], by changing systematically the oxygen-hydrogen bond length (scaling factor s_b) and the atomic partial charges (scaling factor s_q), without any change in the Lennard-Jones interaction parameters. The scaling factors s_b and s_q were both varied systematically by increments of 0.1 in the range 0.1 to 1.5, leading to the definition of 195 water-like models with different dielectric permittivities and H-bonding

Series	Solvent	s_b	s_q	ε	n_H
S_p^H	$W_{1.4}^{0.8}$	0.8	1.4	115	3.4
	$W_{1.2}^{0.9}$	0.9	1.2	89	3.5
	$W_{1.0}^{1.0}$ (SPC)	1.0	1.0	64	3.4
	$W_{0.5}^{1.5}$	1.5	0.5	25	3.4
S_p^h	$W_{1.4}^{0.5}$	0.5	1.4	36	1.7
	$W_{0.9}^{0.8}$	0.8	0.9	33	2.0
	$W_{0.6}^{0.7}$	0.7	0.6	7	1.5
	$W_{0.3}^{0.5}$	0.5	0.3	1	1.4
S_h^P	$W_{0.9}^{1.1}$	1.1	0.9	56	3.6
	$W_{1.0}^{1.0}$ (SPC)	1.0	1.0	64	3.4
	$W_{1.1}^{0.8}$	0.8	1.1	64	2.5
	$W_{1.2}^{0.7}$	0.7	1.2	65	2.2
	$W_{1.4}^{0.6}$	0.6	1.4	76	2.0
	$W_{1.5}^{0.5}$	0.5	1.5	49	1.7
S_h^p	$W_{0.4}^{1.5}$	1.5	0.4	14	2.6
	$W_{0.5}^{1.2}$	1.2	0.5	15	2.2
	$W_{0.5}^{1.1}$	1.1	0.5	13	1.9
	$W_{0.7}^{0.8}$	0.8	0.7	15	1.7
	$W_{0.9}^{0.6}$	0.6	0.9	14	1.6
	$W_{1.3}^{0.4}$	0.4	1.3	12	1.5

Table 3.3: Definition and simulated properties of the 19 artificial solvent models considered in the present study at 298.15 K and 968 kg·m^{-3}. The reported quantities are the scaling factors s_b and s_q applied to the oxygen-hydrogen bond length and to the atomic partial charges, respectively, relative to the SPC water model[136], the static relative dielectric permittivity ε of the liquid, and the average number n_H of H-bonds per molecule in the liquid as a measure of its H-bonding capacity. The models are grouped into four series as described in section 2.1. Note that the SPC water model $W_{1.0}^{1.0}$ belongs to both series S_p^H and S_h^P. These 19 models are a subset of the 195 models developed in Chapter 2, their main thermodynamic, dynamic, dielectric and H-bonding properties being reported in Suppl. Mat. Table 2.S.1 of Chapter 2.

capacities, but identical dispersive interactions as well as molecular shape and size. Note that 30 extreme combinations of s_b and s_q were disregarded, as the corresponding models could only be simulated with very short timesteps. The 195 artificial models are labelled $W_{s_q}^{s_b}$ according to the values of the two scaling factors, so that in particular $W_{1.0}^{1.0}$ is the SPC water model. The main thermodynamic, dynamical, dielectric and H-bonding properties of these solvents, as calculated in Chapter 2 based on pure-liquid simulations, are reported

in Suppl. Mat. Table 2.S.1 therein. Their permittivities ε and H-bonding capacities n_H, the latter value representing the average number of H-bonds per molecule in the pure liquid, are also illustrated graphically in Figure 3.2.

In this set of 195 solvents, a subset of 19 is considered here, corresponding to four series S_p^H, S_p^h, S_h^P and S_h^p. These series permit to investigate specifically the effect of the following trends: (i) in series S_p^H, the progressive decrease of the permittivity (p subscript) at water-like H-bonding capacity (H superscript); (ii) in series S_p^h, the progressive decrease of the permittivity (p subscript) at lower H-bonding capacity (h superscript); (iii) in series S_h^P, the progressive decrease of the H-bonding capacity (h subscript) at water-like permittivity (P superscript); (iv) in series S_h^p, the progressive decrease of the H-bonding capacity (h

Figure 3.2: Simulated properties of the 195 artificial solvent models developed in Chapter 2 at 298.15 K and 968 kg·m^{-3}. The static relative dielectric permittivity ε of the liquid (a) and the average number n_H of H-bonds per molecule in the liquid (b) are shown as a function of the scaling factors s_b and s_q applied to the oxygen-hydrogen bond length and to the atomic partial charges, respectively, relative to the SPC water model[136]. The squares identify the subset of 19 models considered in the present study, colored according to the four series (S_p^H, S_p^h, S_h^P, and S_h^p; Table 3.3). The SPC model $W_{1.0}^{1.0}$ is shown in violet and belongs to both S_p^H and S_h^P. The areas in black and gray correspond to models that are either not in the set (30 s_b and s_q combinations were disregarded as the corresponding models could only be simulated with very short timesteps) or not liquid (glassy state), respectively. The main thermodynamic, dynamic, dielectric and H-bonding properties of the 195 models are reported numerically in Suppl. Mat. Table 2.S.1 in Chapter 2.

subscript) at lower permittivity (p superscript). The main properties of the 19 artificial solvents considered, along with the definition of the four series, are summarized in Table 3.3 and illustrated graphically in Figure 3.2. Note that the SPC water model $W_{1.0}^{1.0}$ belongs to both series S_p^H and S_h^P. For the artificial solvents, the number of solvent molecules N_s in the computational box was always set to 1200.

3.2.2   Simulations

The simulations include explicit-solvent MD simulations of the three solutes (Glc, Gal, Deo) in the two different sets of solvents (5 physical and 19 artificial) where the sampling along the dihedral angle ω was enhanced by the LEUS approach[14], along with plain SD simulations[100,101] of the three compounds in vacuum.

   The MD simulations were performed under periodic boundary conditions based on cubic computational boxes containing one solute and N_s solvent molecules (values of N_s given in Section 3.2.1). For the physical solvents, the simulations were carried out in the isothermal-isobaric (NPT) ensemble at 298.15 K and 1 bar. For the artificial solvents, the simulations were carried out in the canonical (NVT) ensemble at 298.15 K and an effective solvent density of 998.3 kg·m^{-3}. Here, the effective solvent density is estimated as $V^{-1}M(N_s+\tilde{N}_s)$, where V is the box volume, M the molecular mass of water, and $\tilde{N}_s$ an effective number of water molecules accounting for the solute volume (set here to $\tilde{N}_s = 12$).

   The simulations involving the artificial solvent models must be carried out under NVT conditions because the Lennard-Jones interaction parameters of these models were not adjusted to reproduce the equilibrium density of water at 298.15 K and 1 bar. Accordingly, under NVT conditions at 298.15 K and 968 kg·m^{-3} (as used in Chapter 2), they are characterized by equilibrium pressures that range from -0.6 to 8.9 kbar for the 19 models considered here (see Suppl. Mat. Table 2.S.1 in Chapter 2). However, incorrect pressures

are expected to have only a limited influence on the rotameric equilibrium of the hydroxy-methyl group. This was verified explicitly by calculating the relative molar volumes of the three rotamers at different pressures. Numerical integration of the $P\Delta V$ terms suggests relative free energy shifts on the order of 0.4-0.9 kJ·mol^{-1} for the maximal pressure of 10 kbar. The details of these calculations can be found in Suppl. Mat. Table 3.S.3.

In both the NPT and NVT simulations, the temperature was maintained close to its reference value of 298.15 K by weakly coupling[115] solute and solvent degrees of freedom jointly (to avoid solute damping[264]) to an external bath using a relaxation time of 0.1 ps. In the NPT simulations, the pressure was maintained close to its reference value of 1 bar by weakly coupling[115] the atomic coordinates and box dimensions (isotropic coordinate scaling, group-based virial) to an external bath using a relaxation time of 0.5 ps and an isothermal compressibility of 4.575·10^{-4} kJ^{-1}·mol·nm^3 as appropriate for aqueous biomolecular systems[101]. The compressibility was not adjusted for the different solvents, because it is combined with the arbitrary choice of a pressure relaxation time and does not affect the average thermodynamic properties of the system. The center of mass translation of the box was removed every timestep.

The leap-frog algorithm[103] was used to integrate Newton's equations of motion with a timestep of 2 fs. Solute bond-length constraints as well as the full rigidity of the solvent molecules were enforced by application of the SHAKE procedure[53] with a relative geo-metric tolerance of 10^{-4}. The non-bonded interactions were calculated using a twin-range scheme[69], with short- and long-range cutoff distances set to 0.8 and 1.4 nm, respectively, and an update frequency of 5 timesteps for the short-range pairlist and intermediate-range interactions. A reaction-field correction[60,65] was applied to account for the mean effect of electrostatic interactions beyond the long-range cutoff distance, using the relative dielectric permittivity appropriate for the solvent model considered (Tables 3.2 and 3.3).

The LEUS method[14] was applied to improve the conformational sampling around the dihedral-angle ω (C_4-C_5-C_6-O_6, see Figure 3.1) characterizing the orientation of the hydroxymethyl group, which can present slow relaxation[170,265–269]. The LEUS calculations involved two steps: (*i*) a local elevation[90] (LE) build-up phase of duration t_{LE} = 4 ns, to progressively optimize a memory-based biasing potential along ω; (*ii*) an umbrella sampling[105] (US) phase of duration t_{US} = 40 ns, using this preoptimized (now time-independent) biasing potential to enhance the sampling. The biasing potential was represented by means of 36 truncated-polynomial basis functions[270] with a spacing of 10°. The polynomial widths were set equal to the grid spacing. The LE build-up phase relied on a fixed force-constant increment per visit set to k_{LE} = 10^{-4} kJ·mol^{-1}.

The SD simulations in vacuum were performed by integrating the Langevin equation of motion[100,101]. They relied on a reference temperature of 298.15 K and a friction coefficient of 91 ps^{-1}. The choice of the latter value, appropriate for water[100], has no effect on the average thermodynamic properties of the system. The SD simulations were carried out for a duration t_{SD} = 1 μs. The LEUS procedure was not applied in this case.

For both the MD+LEUS and the SD simulations, the initial structure of the solute was in the $^4C_1$ chair conformation. After filling the computational box with molecules of the solvent considered, the equilibration consisted of a steepest-descent energy minimization, followed by a 0.5 ps thermalization MD (progressively increased temperature), and by a 1 ns plain MD simulation (constant temperature). From this point, configurations (along with the value of the biasing potential in the US sampling phases of the LEUS simulations) were written to file every 0.5 ps for subsequent analysis.

3.2.3   Analysis

The analysis of the simulations was performed in terms of: (i) ring conformations; (ii) probability distribution profiles around ω; (iii) relative populations of the three canonical hydroxymethyl rotamers; (iv) corresponding relative free energies; (v) occurrence of intramolecular and solute-solvent H-bonds; (vi) correlated probability distributions of the exocyclic dihedral angles; (vii) calculated NMR J-coupling constants. For the MD+LEUS simulations, all the thermodynamic and structural quantities analyzed (except the exocyclic dihedral-angle correlations; see below) were calculated based on the configurations generated during the US sampling phase, with a reweighting factor depending on the value of the biasing potential associated with each configuration, as detailed elsewhere[14].

The ring conformations were assigned based on three out-of-plane dihedral angles α_1, α_2 and α_3 defined according to Pickett & Strauss[271] as the values of the improper 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 $180°$. The hexopyranose was considered to be in the $^4C_1$ chair conformation when the three α-angles were in a range between $-17°$ and $-50°$. For more details about the assignment procedure, see Ref.[14] and Table 2 therein.

The probability distribution profiles $p(\omega)$ around ω (C_4-C_5-C_6-O_6, see Figure 3.1) were calculated as normalized histograms using a bin width $\Delta\omega = 2°$. The canonical conformations of the hydroxymethyl group were defined based on the same dihedral angle. Values in the ranges $[0°; 120°[$, $[120°; 240°[$ and $[240°; 360°[$ were integrated to define the relative populations of the gg, tg and gt rotameric states, respectively. Relative free energies factoring out the influence of the solvent on Deo and on the gt conformer of Glc and Gal (see Eq. 3.6 below) were also monitored.

For the intramolecular H-bonding analysis, only the H-bonds involving the hydroxymethyl group were considered, namely $H_4 \rightarrow O_6$, $H_6 \rightarrow O_4$ and $H_6 \rightarrow O_5$. The occurrence

of these three H-bonds was analyzed separately for each rotameric state of the hydroxymethyl group. An intramolecular H-bond was assumed to be present when the distance between the hydrogen and acceptor atoms is below 0.25 nm, and the angle between the donor, hydrogen and acceptor atoms is above 100°. The use of a somewhat relaxed angular criterion[170] (a minimum angle of 135° is typically used instead[101]) is necessary to encompass H-bonds between the hydroxymethyl group and the ring oxygen atom. The average numbers of solute-solvent H-bonds involving the hydroxymethyl group, the hydroxyl group at C_4 and the ring oxygen atom O_5 were also analyzed, using in this case the normal angular criterion (minimum angle of 135°). For the artificial solvent models involving a scaling of the oxygen-hydrogen bond length (scaling factor $s_b \neq 1.0$), the solute-solvent H-bond analysis of each trajectory configuration was preceded by a rescaling of the solvent oxygen-hydrogen distances as described in Chapter 2. This procedure involves a displacement of the hydrogen atoms of each solvent molecule along the oxygen-hydrogen bond vector, while keeping the oxygen position fixed, so that the value of 0.1 nm corresponding to the SPC water model is recovered.

The analysis of correlations between the exocyclic dihedral angles considered the five dihedral angles ω (C_4-C_5-C_6-O_6), ϕ (O_5-C_1-O_1-H_1), χ_2 (C_1-C_2-O_2-H_2), χ_3 (C_2-C_3-O_3-H_3) and χ_4 (C_3-C_4-O_4-H_4). The dihedral angle χ_6 (C_5-C_6-O_6-H_6) was omitted from this analysis, as it is either strongly correlated[191] with ω in the presence of an intramolecular H-bond involving O_6, or essentially uncorrelated[170,184] in the absence of such an H-bond. For the above set of five dihedral angles, the correlations were monitored for all possible pairwise combinations in the form of scatter plots involving successive points sampled at 0.5 ps intervals along the MD+LEUS simulations. No reweighting was applied here, which corresponds to an artificial situation where the probability distribution $p(\omega)$ is essentially flat. For the discussion, values of χ in the ranges [0°, 120°[, [120°, 240°[and [240°, 360°[are labelled g^+, t and g^-, respectively. This analysis was only performed considering the

physical solvents of highest (H_2O) and lowest (CCl_4) polarities, as well as the two artificial solvents $W_{0.3}^{0.5}$ and $W_{1.3}^{0.4}$ presenting the lowest H-bonding capacity ($n_H = 1.4 - 1.5$) along with the lowest or an intermediate permittivity ($\varepsilon = 1$ or 12, respectively).

Finally, for the simulations involving the physical solvents H_2O, DMSO and CH_3OH, a direct comparison of the hydroxymethyl rotameric populations with primary experimental data was performed, namely in terms of the vicinal proton-proton NMR J-coupling constants ($^3J_{HH}$) between the proton at the carbon atom C_5 and the R or S protons of the methylene group C_6. The two values, which will be noted J_R for $^3J_{H5,H6R}$ and J_S for $^3J_{HR,H6S}$, were calculated based on the simulations considering three alternative Karplus equations. The Karplus equation of Haasnoot *et al.* (original version from Ref.[241] without the modification of Ref.[272]) is given by

$$^3J_{HH}/j_0 = 13.86 \cos^2\phi - 0.81 \cos\phi + \sum_{i=1}^{I} \Delta\chi_i[0.56 - 2.32 \cos^2(\xi_i\phi + 17.9|\Delta\chi_i|)], \quad (3.1)$$

where $j_0 = 1$ s^{-1}, ϕ is the dihedral angle between the coupled protons (in degrees, to be consistent with the value of 17.9° in the last term), I the number of non-hydrogen substituents S_i of the H-C-C-H fragment, $\Delta\chi_i = \chi_i - \chi_H$, χ_i being the Huggins (unitless) electronegativity[273] of an atom i, and ξ_i is minus the sign of the H-C-C-S_i dihedral angle in the conformation where the coupled protons are eclipsed. The angle ϕ corresponds to $\omega + \delta$, with $\delta = 0°$ for J_R and $\delta = -120°$ for J_S. For the compounds considered here, $I = 3$ and Eq. (3.1) is applied with $\xi_i = +1$ and -1 for the substituents C_4 and O_5, respectively, and with $\xi_i = $ -1 and +1 for the substituent O_6 when evaluating J_R and J_S, respectively. The Karplus equation of Tafazzoli & Ghiasi[242] is given by

$$J_R/j_0 = 5.06 + 0.45 \cos(\tilde{\omega}) - 0.90 \cos(2\tilde{\omega}) + 0.80 \sin(\tilde{\omega}) + 4.65 \sin(2\tilde{\omega}) \qquad (3.2)$$

and

$$J_S/j_0 = 4.86 - 1.22\,\cos(\tilde{\omega}) + 4.32\,\cos(2\tilde{\omega}) + 0.04\,\sin(\tilde{\omega}) + 0.07\,\sin(2\tilde{\omega}), \qquad (3.3)$$

where $\tilde{\omega}$ is the dihedral angle O_5-C_5-C_6-O_6 (see Figure 3.1). Finally, the Karplus equation of Stenutz *et al.*[191] is given by

$$J_R/j_0 = 5.08 + 0.47\,\cos(\tilde{\omega}) - 0.12\,\cos(2\tilde{\omega}) + 0.90\,\sin(\tilde{\omega}) + 4.86\,\sin(2\tilde{\omega}) \qquad (3.4)$$

and

$$J_S/j_0 = 4.92 - 1.29\,\cos(\tilde{\omega}) + 4.58\,\cos(2\tilde{\omega}) + 0.05\,\sin(\tilde{\omega}) + 0.07\,\sin(2\tilde{\omega}). \qquad (3.5)$$

3.3 Results

3.3.1 Ring conformation

The results concerning the analysis of the ring conformation in the physical solvents are reported in Table 3.4. Corresponding results for the 19 artificial solvents can be found in Suppl. Mat. Table 3.S.4. This analysis serves to establish that the leading ring conformation sampled in the simulations is as expected the $^4C_1$ chair conformation. Otherwise, any subsequent discussion concerning the correlation between the rotameric states and H-bonding properties of the hydroxymethyl group would be complicated by the presence of alternative ring conformations, *e.g.* inverted $^1C_4$ chair, boats or skew-boats. Such conformations, which present only marginal populations experimentally for unfunctionalized hexopyranoses[215,245,246,257–263] (with the exception of idose), have been seen to occur in previous simulations[14,170,246,248,274–276] owing to force-field inaccuracies. The 53A6 force

field[244] is *a priori* likely to be more prone to ring distortions considering that, unlike the 56A6$_{\mathrm{CARBO}}$ force field[245] (see also Ref.[246] for the revised version CARBO_R), its parameterization did not involve the consideration of ring-conformational free energies[248].

From the results reported in Tables 3.4 and 3.S.4, it appears that the $^4C_1$ chair conformation accounts in all cases for at least 87% of the sampled configurations. The $^4C_1$ populations are tendentially lower for the 56A6$_{\mathrm{CARBO}}$ force field compared to the 53A6 force field (contrary to the above expectations) and for Glc compared to Gal, Deo being

		53A6				56A6$_{\mathrm{CARBO}}$			
	Solvent	α_1 [°]	α_2 [°]	α_3 [°]	$^4C_1$ Pop. [%]	α_1 [°]	α_2 [°]	α_3 [°]	$^4C_1$ Pop. [%]
Glc	H$_2$O	-35 (7)	-31 (8)	-35 (7)	91	-38 (7)	-36 (8)	-34 (7)	89
	DMSO	-35 (7)	-29 (8)	-35 (7)	89	-37 (8)	-35 (8)	-35 (8)	87
	CH$_3$OH	-35 (7)	-30 (8)	-35 (7)	91	-37 (8)	-36 (8)	-34 (7)	88
	CHCL$_3$	-34 (7)	-35 (7)	-34 (7)	96	-36 (7)	-39 (7)	-34 (7)	89
	CCL$_4$	-34 (7)	-35 (7)	-34 (7)	96	-37 (7)	-40 (7)	-34 (7)	88
	VAC	-34 (7)	-35 (7)	-34 (7)	96	-37 (7)	-40 (7)	-34 (7)	88
Deo	solvent	α_1 [°]	α_2 [°]	α_3 [°]	$^4C_1$ Pop. [%]	α_1 [°]	α_2 [°]	α_3 [°]	$^4C_1$ Pop. [%]
	H$_2$O					-36 (7)	-37 (7)	-37 (7)	90
	DMSO					-35 (8)	-36 (8)	-38 (7)	87
	CH$_3$OH					-36 (7)	-36 (8)	-37 (7)	89
	CHCL$_3$					-37 (7)	-38 (7)	-35 (7)	89
	CCL$_4$					-37 (7)	-38 (7)	-35 (7)	90
	VAC					-37 (7)	-38 (7)	-35 (7)	91
Gal	solvent	α_1 [°]	α_2 [°]	α_3 [°]	$^4C_1$ Pop. [%]	α_1 [°]	α_2 [°]	α_3 [°]	$^4C_1$ Pop. [%]
	H$_2$O	-31 (7)	-38 (6)	-36 (6)	94	-36 (7)	-38 (7)	-36 (7)	92
	DMSO	-30 (7)	-38 (6)	-36 (6)	93	-35 (7)	-39 (7)	-36 (7)	90
	CH$_3$OH	-31 (7)	-38 (6)	-36 (6)	94	-35 (7)	-39 (7)	-36 (7)	91
	CHCL$_3$	-33 (6)	-36 (6)	-35 (6)	97	-38 (7)	-37 (7)	-33 (7)	92
	CCL$_4$	-32 (6)	-36 (6)	-35 (6)	97	-38 (7)	-37 (6)	-32 (7)	93
	VAC	-34 (7)	-35 (7)	-34 (7)	96	-37 (7)	-40 (7)	-34 (7)	88

Table 3.4: Ring-conformational properties of the three compound considered in the physical solvents at 298.15 K and 1 bar. The reported quantities are the Pickett & Stauss dihedral angles[271] α_1, α_2 and α_3, and the population of the $^4C_1$ chair conformation (defined by α_1, α_2 and α_3 simultaneously in the range between -17° and -50°). The values are calculated as reweighted averages over the 40 ns US phase of the different LEUS simulations (plain average over the 1 μs SD simulations for VAC). The corresponding standard deviations are reported between parentheses. The compounds considered are Glc, Deo and Gal (Figure 3.1). The solvents considered are physical solvents of decreasing polarity (Table 3.2). The force fields considered are the 53A6 force field[244] and the 56A6$_{\mathrm{CARBO}}$ force field[245]. The simulations of Deo with 53A6 are omitted due to ambiguous torsional potential definitions. See Suppl. Mat. Table 3.S.4 for corresponding results with the artificial solvents.

intermediate. The values of the Pickett & Strauss dihedral angles[271] α_1, α_2 and α_3, averaged over the entire simulation, are in all cases close to the canonical value of $-35°$. The nature of the solvent has only a limited effect on the $^4C_1$ populations, although DMSO tends to induce slightly more pronounced ring distortions compared to the other physical solvents. In all cases, most of the sampled configurations that do not correspond to a $^4C_1$ chair are distorted $^4C_1$ chairs. Only four simulations present brief flips (at most 0.1 ns duration) to boat conformers, while the $^1C_4$ conformation is never observed.

3.3.2 Hydroxymethyl-group orientation

The normalized probability distributions around the dihedral angle ω defining the orientation of the hydroxymethyl group are displayed in Figure 3.3 for the physical solvents. Corresponding graphs for the 19 artificial solvents can be found in Suppl. Mat. Figures 3.S.1, 3.S.2 and 3.S.3 (series $S_p^H + S_p^h$, S_h^P and S_h^p, respectively). Given the selected LE build-up and US sampling times $t_{LE} = 4$ and $t_{US} = 40$ ns, respectively, the LEUS simulations provide smooth and well converged probability distributions. The same applies to the 1 μs SD simulation in vacuum. In all cases, the distributions present relatively narrow peaks (root-mean-square fluctuations on the order of 10-15°) centered close to the three canonical conformations gg ($\omega = 60°$), gt ($\omega = 180°$) and tg ($\omega = 300°$), with deviations of the peak position relative to the canonical values on the order of 5° at most (10° for Gal with the 53A6 force field in some of the artificial solvents). Note that the peak positions are also slightly shifted when going from the 53A6 to the 56A6$_{CARBO}$ force field, by about -10°, -5° and -10° for gg, gt and tg, respectively. Because the populations can be unambiguously ascribed to the three canonical rotamers, further discussion of the solvent-dependent trends is formulated in terms of integrated probability distributions over the three wells.
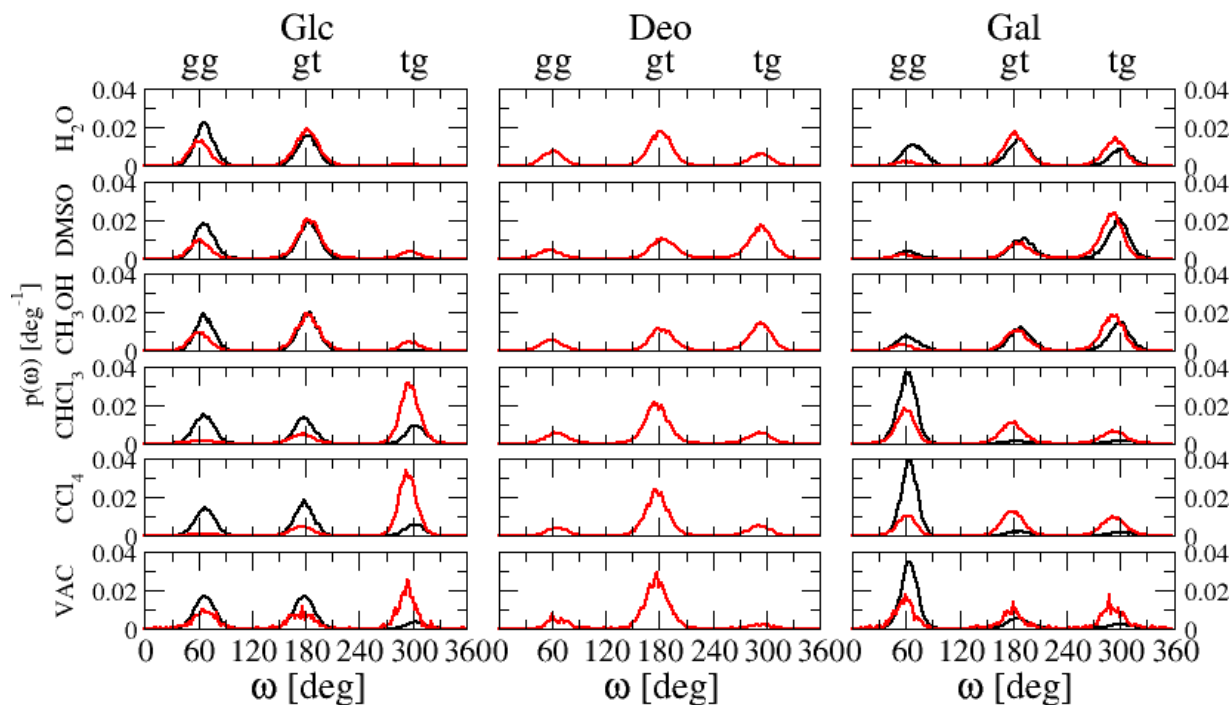
Figure 3.3: Normalized probability distributions around the dihedral angle ω for the three compounds considered in the physical solvents at 298.15 K and 1 bar. The values are calculated as reweighted averages over the 40 ns US phase of the different LEUS simulations (or 1 μs plain SD simulation for VAC) using a bin size of $2°$. The different panels from left to right correspond to Glc, Deo and Gal (Figure 3.1), and from top to bottom to the series of physical solvents (Table 3.2) in order of decreasing polarity. The black curves correspond to the 53A6 force field[244] and the red curves to the 56A6$_{CARBO}$ force field[245]. The simulations of Deo with 53A6 are omitted due to ambiguous torsional potential definitions. See Suppl. Mat. Figures 3.S.1-3.S.3 for corresponding results with the artificial solvents.

3.3.3 Relative populations of the rotamers

The populations of the three hydroxymethyl rotamers are displayed in Figure 3.4 for the physical solvents and in Figures 3.5-3.7 for the artificial solvents (series $S_p^H + S_p^h$, S_h^P and S_h^p, respectively). The data is also reported numerically in Suppl. Mat. Tables 3.S.5 (physical solvents) and 3.S.6-3.S.8 (artificial solvent).

The results obtained using the 53A6 and 56A6$_{\text{CARBO}}$ force fields are clearly different. However, the trends observed along the five series of solvents are qualitatively similar for the two force fields, the difference arising mainly from an offset in the relative populations of the three rotamers. Taking water as a reference, the relative *gg*:*gt*:*tg* populations in the 53A6 force field are found to be 56:44:0 for Glc and 34:41:25 for Gal. These results are consistent with those of previous studies using this force field or the nearly identical 45A4 version (see Suppl. Mat. Table 3.S.1), *e.g.* 55:45:0 and 34:41:25 in Ref.[248] or 57:43:0 and 33:41:26 in Ref.[170] (β-anomer, free lactol group). The corresponding relative populations in the 56A6$_{\text{CARBO}}$ force field are found to be 36:59:4 for Glc, 24:56:20 for Deo and 7:53:40 for Gal. For Glc and Gal, this is again consistent with the results of other studies using this force field (see Suppl. Mat. Table 3.S.1), *e.g.* 35:60:5 and 6:52:42 in Ref.[246] (β-anomer, free lactol group). The results are also similar to the estimates of 37:60:3 and 7:67:26 in Ref.[245] for the corresponding O$_1$-methylated compounds (β-anomer). Note, however, that if O$_1$-methylation has essentially no influence on the relative populations of the hydroxymethyl rotamers for Glc, it leads to changes of about 15% in the *gt* and *tg* populations for Gal. This effect was also observed within the 56A6$_{\text{CARBO\_R}}$ force field[246] (see Table 7 therein).

The difference between the rotamer populations in the two parameter sets is not entirely surprising, considering that the 53A6 (45A4) force field[248] was calibrated against older NMR-derived populations[181,182,190], whereas the 56A6$_{\text{CARBO}}$ force field[245] (see also Ref.[246])

is based on more recent estimates[184,185,191], suggesting a slightly lower gg:gt ratio along with a non-negligible tg population for Glc, and a significantly lower gg population for Gal (see Table 3.1). Two other differences between the force-field versions should be noted: (i) 53A6 relies on distinct hydroxymethyl torsional potentials for Glc (equatorial hydroxyl group at C_4 in the $^4C_1$ chair; potential on ω only) and Gal (axial hydroxyl group at C_4 in the $^4C_1$ chair; other potential types on both ω and $\tilde{\omega}$), whereas 56A6$_{CARBO}$ uses a unique pair of torsional potentials (one on ω and one on $\tilde{\omega}$) for the two compounds; (ii) 56A6$_{CARBO}$ was validated against primary NMR data in aqueous solution, namely J-coupling constants, whereas 53A6 was only compared to secondary data, namely NMR-derived populations. Note, finally, that although the present work considers β-anomers with a free lactol group, experimental (see Suppl. Mat. Tables 3.S.1 and 3.S.2) and simulation[246] results suggest that α-anomers or/and O_1-methylated compounds present similar J-values and rotamer population ratios in water, with population differences on the order of 10% or less.

Most of the population discussion below refers to the 56A6$_{CARBO}$ results, considered to be more accurate. Although the solvent-dependent trends in 53A6 are qualitatively similar, the main differences are: (i) an underrepresentation of the tg rotamer of Glc, due to the selection of torsional potentials entirely precluding the occurrence of this rotamer for aqueous Glc; (ii) an overrepresentation of the gg relative to the gt rotamer of Glc; (iii) an overrepresentation of the gg relative to the gt and tg rotamers of Gal, the two latter conformers presenting about the same population ratios in the two force fields. The following paragraphs summarize the trends observed along five series of solvents, considering the 56A6$_{CARBO}$ population estimates.

Along the series of physical solvents with decreasing polarity (Figure 3.4; numerical values in Suppl. Mat. Table 3.S.5), the following trends are observed. For Glc, the population of tg systematically increases from 4% in H_2O to 80% in VAC. The popula-
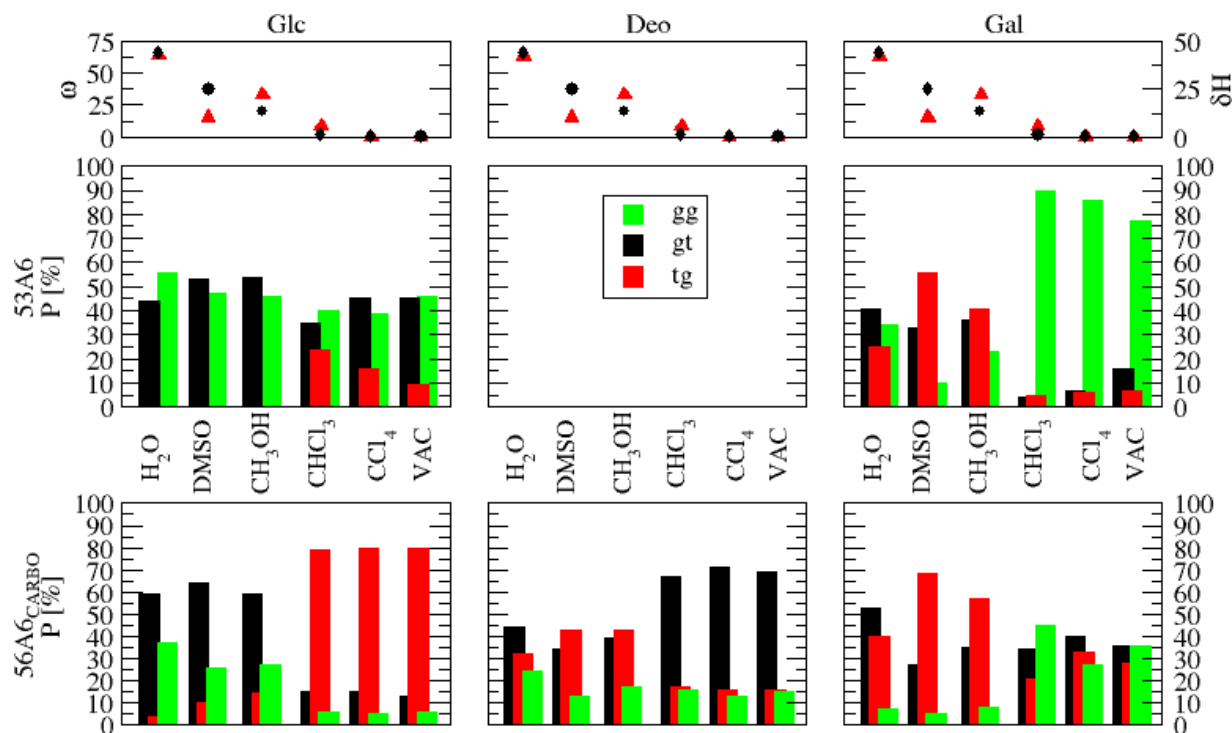
Figure 3.4: Populations of the three staggered hydroxymethyl rotamers for the three compounds considered in the physical solvents at 298.15 K and 1 bar. The values are calculated as reweighted averages over the 40 ns phase of the different LEUS simulations (or 1 μs plain SD simulation for VAC). The columns from left to right correspond to Glc, Deo, and Gal (Figure 3.1) using the 53A6 force field[244] (middle row) or the 56A6$_{\text{CARBO}}$ force field[245] (bottom row). The bars follow the series of physical solvents (Table 3.2) in order of decreasing polarity. The values of the solvent permittivity ε (black circles, scale on the left) and Hansen parameter δH (red triangles, scale on the right) are also displayed (top row). The conformer populations refer to the canonical rotamers gg, gt and tg (Figure 3.1). The simulations of Deo with 53A6 are omitted due to ambiguous torsional potential definitions. The data is reported numerically in Suppl. Mat. Table 3.S.5.

tions of gg and gt decrease accordingly, at approximately constant relative ratio. This is consistent with an increasingly important role of the $O_4 \leftrightarrow O_6$ intramolecular H-bond, compatible with the tg rotamer only (Figure 3.1). The sharpest change along the series is from CH_3OH to $CHCl_3$, also associated with the most pronounced decrease in the solvent polarity. For Gal, the population of gg systematically increases from 7% in H_2O to 36% in VAC, except for $CHCl_3$ with an anomalously high value of 45%. The populations of gt and tg decrease accordingly, at approximately constant relative ratio except in DMSO and CH_3OH, where the gt:tg ratio is inverted. The general trend is again consistent with

an increasingly important role of the $O_4 \leftrightarrow O_6$ intramolecular H-bond compatible with the gg rotamer only (Figure 3.1), from destabilizing in H_2O to stabilizing in VAC. Finally, for Deo, one observes an increase of the gt:tg ratio, at approximately constant gg population. Here also, DMSO and CH_3OH represent exceptions, with an inversion of the gt:tg ratio. As the formation of an $O_4 \leftrightarrow O_6$ H-bond is impossible in Deo (no hydroxyl group at C_4), one may think that the dominant remaining driving forces are the solvent-dependence of the *gauche*-effect and the possible formation of an $H_6 \rightarrow O_5$ H-bond. The former effect would be expected to increasingly favor the tg rotamer upon decreasing the solvent polarity, as this conformer has the lowest dipole for the O_5-C_5-C_6-O_6 segment. However, the observed trends are incompatible with this expectation. The formation of a $H_6 \rightarrow O_5$ H-bond would be expected to increasingly favor the gg and gt conformers, where this H-bond is possible. The observed trends are not incompatible with this second hypothesis, although only the gt population increases. Still, it should be stressed that the physical solvent molecules also vary in shape, size and dispersive interactions, which may affect the rotamer distribution in a complex and indirect way *via* specific solvation and packing effects involving the entire ring.

Along the series S_p^H of artificial solvents with decreasing permittivity (ε from 115 down to 25) at water-like H-bonding capacity ($n_H \approx 3.4$-3.5), see Figure 3.5 (left parts of the graphs; numerical values in Suppl. Mat. Table 3.S.6), the following trends are observed. For Glc and Deo, the decreasing permittivity has essentially no influence on the rotameric equilibrium (except possibly for the last solvent $W_{0.5}^{1.5}$), *i.e.* the relative rotamer populations are essentially the same as in water. For Gal, the gg and gt populations slightly increase at the expense of the tg population (the effect is also more pronounced for the last solvent $W_{0.5}^{1.5}$).

Along the series S_p^h of artificial solvents with decreasing permittivity (ε from 36 down to 1) at low H-bonding capacity ($n_H \approx 1.4$-2.0), see Figure 3.5 (right parts of the graphs;
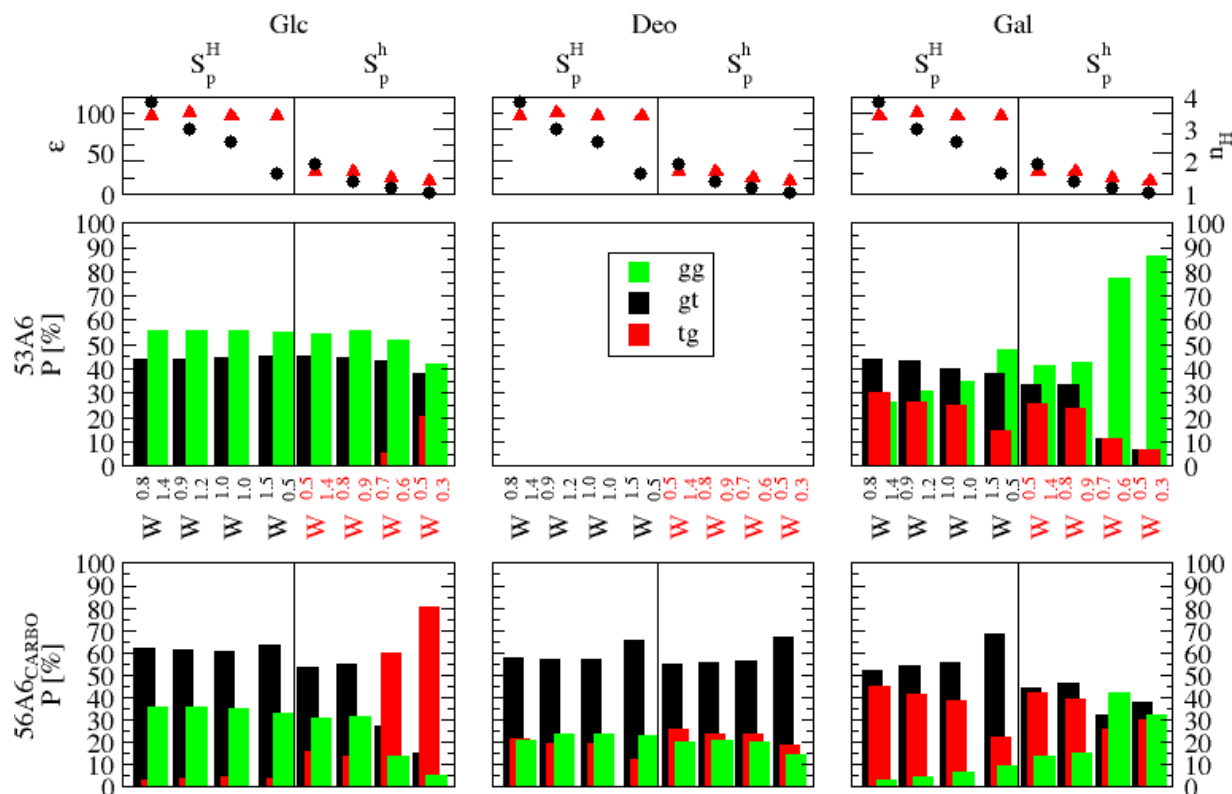
Figure 3.5: Populations of the three staggered hydroxymethyl rotamers for the three compounds considered in the series S_p^H and S_p^h of artificial solvents at 298.15 K and 998.3 kg·m^{-3}. The values are calculated as reweighted averages over the 40 ns phase of the different LEUS simulations. The columns from left to right correspond to Glc, Deo and Gal (Figure 3.1) using the 53A6 force field[244] (middle row) or the 56A6$_{\text{CARBO}}$ force field[245] (bottom row). The bars follow the series S_p^H and S_p^h of artificial solvents (Table 3.3) in order of decreasing dielectric permittivity at water-like or lower H-bonding capacity, respectively. The values of the solvent permittivity ε (black circles, scale on the left) and number of H-bonds per molecule n_H (red triangles, scale on the right) are also displayed (top row). The conformer populations refer to the canonical rotamers gg, gt and tg (Figure 3.1). The simulations of Deo with 53A6 are omitted due to ambiguous torsional potential definitions. The data is reported numerically in Suppl. Mat. Table 3.S.6.

numerical values in Suppl. Mat. Table 3.S.6), the following trends are observed. For Glc, the population of tg systematically increases from 16 to 82%. The populations of gg and gt decrease accordingly, at approximately constant relative ratio. For Gal, the population of gg increases from 14 to 33%. The populations of gt and tg decrease accordingly, at approximately constant relative ratio. For Deo, decreasing permittivity has little influence on the rotameric equilibrium, as was the case for series S_p^H (except possibly for the last

solvent $W_{0.3}^{0.5}$).

For Glc and Gal, these trends are qualitatively similar to those observed for the physical solvents, a decrease in the solvent permittivity favoring the potentially H-bonding conformer (tg for Glc and gg for Gal). The fact that the corresponding trends are absent (Glc) or less pronounced (Gal) in series S_p^H suggests that the effect of the solvent permittivity only becomes significant when the solvent H-bonding capacity is low. For Deo, the absence of significant trends along both series S_p^H and S_p^h provides a hint that the effects observed along the physical solvent series are not caused by a change in the solvent permittivity.

Along the series S_h^P of artificial solvents with decreasing H-bonding capacity (n_H from 3.6 down to 1.7) at water-like permittivity ($\varepsilon \approx 49 - 76$), see Figure 3.6 (numerical values in Suppl. Mat. Table 3.S.7), the following trends are observed. For Glc, the decreasing H-bonding capacity has a very weak influence on the rotameric equilibrium, with a slight increase of tg at the expense of gg. For Gal, the influence is also limited, with a small increase of tg at the expense of gt. For Deo, the population of tg increases slightly, at the expense of both gg and gt.

Finally, along the series S_h^p of artificial solvents with decreasing H-bonding capacity (n_H from 2.6 down to 1.5) at low permittivity ($\varepsilon \approx 12 - 15$), see Figure 3.7 (numerical values in Suppl. Mat. Table 3.S.8), the following trends are observed. For Glc, the population of tg increases from 11 to 45%. The populations of gg and gt decrease accordingly, at approximately constant relative ratio. For Gal, the populations of gg and tg increase from 15 to 34% and from 25 to 32%, respectively. The population of gt decreases accordingly. For Deo, the population of tg increases slightly, at the expense of gg and gt.

For Glc and Gal, the trends along series S_h^p are similar but more pronounced compared to those observed along the series S_h^P, *i.e.* the effect of a decreasing solvent H-bonding capacity becomes more significant when the solvent permittivity is low. As was observed
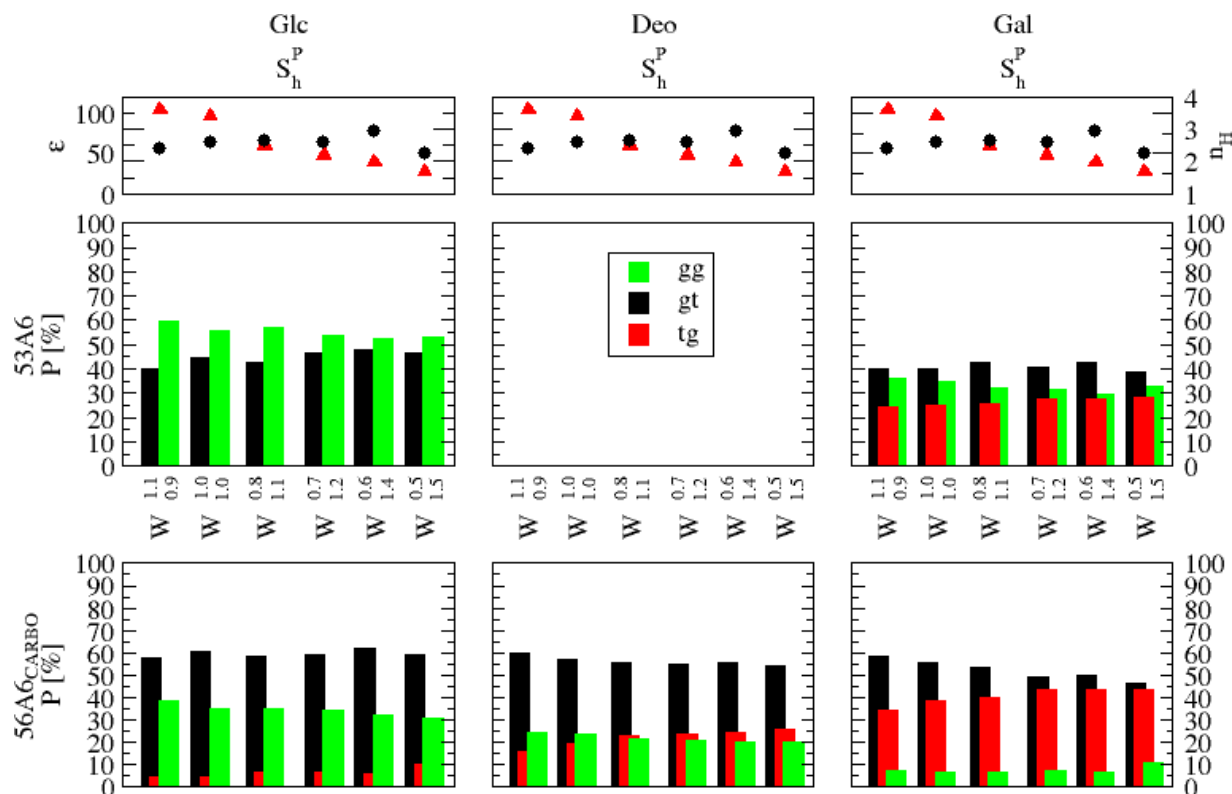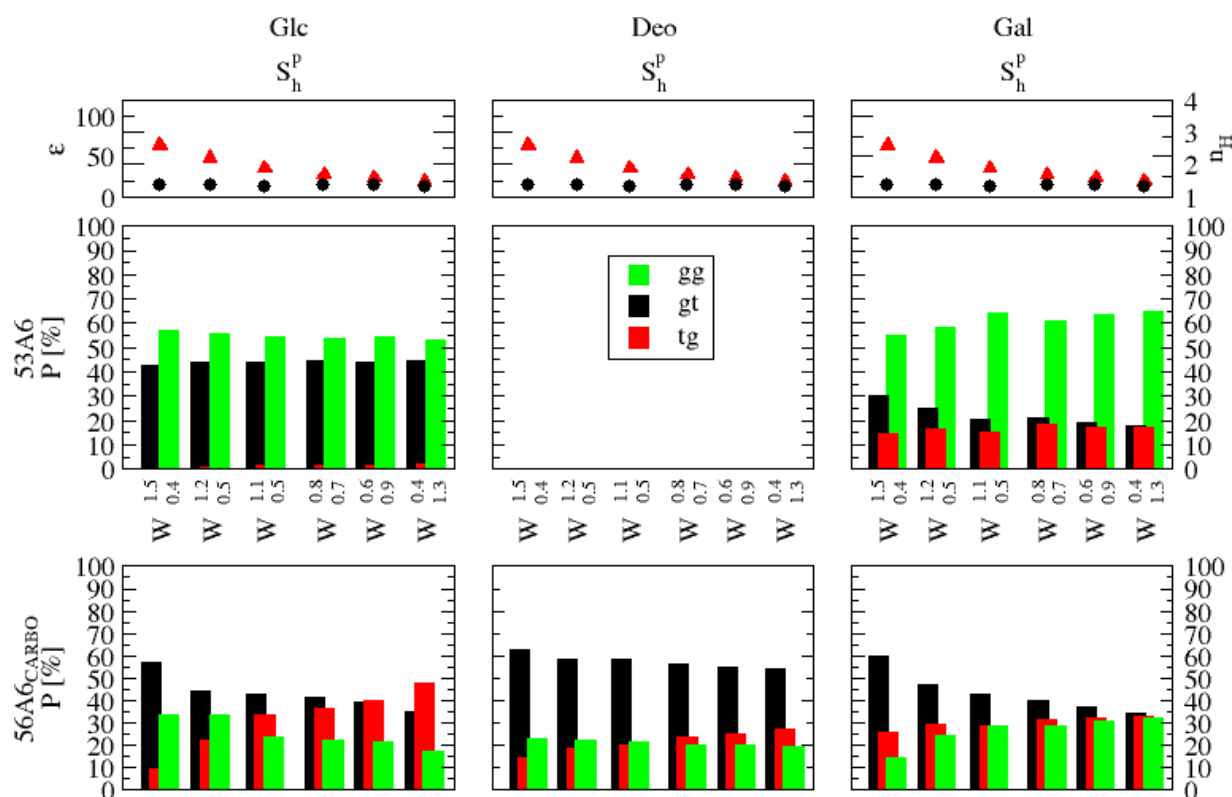
Figure 3.6: Populations of the three staggered hydroxymethyl rotamers for the three compounds considered in the series S_h^P of artificial solvents at 298.15 K and 998.3 kg·m^{-3}. The values are calculated as reweighted averages over the 40 ns phase of the different LEUS simulations. The columns from left to right correspond to Glc, Deo and Gal (Figure 3.1) using the 53A6 force field[244] (middle row) or the 56A6$_{CARBO}$ force field[245] (bottom row). The bars follow the series S_h^P of artificial solvents (Table 3.3) in order of decreasing H-bonding capacity at water-like permittivity. The values of the solvent permittivity ε (black circles, scale on the left) and number of H-bonds per molecule n_H (red triangles, scale on the right) are also displayed (top row). The conformer populations refer to the canonical rotamers gg, gt and tg (Figure 3.1). The simulations of Deo with 53A6 are omitted due to ambiguous torsional potential definitions. The data is reported numerically in Suppl. Mat. Table 3.S.7.

for the physical solvents, a decrease in the solvent H-bonding capacity favors the potentially H-bonding conformer (tg for Glc and gg for Gal). One main difference, however, is the simultaneous tg population increase for Gal along the series S_h^P and S_h^p, which is opposite to the trend observed along the series of physical solvents and along the series S_p^H and S_p^h. For Deo, the effect of the solvent H-bonding capacity is comparable in series S_h^P and S_h^p, and remains relatively weak, as was the case for series S_p^H and S_p^h. This provides a hint

Figure 3.7: Populations of the three staggered hydroxymethyl rotamers for the three compounds considered in the series S_h^p of artificial solvents at 298.15 K and 998.3 kg·m^{-3}. The values are calculated as reweighted averages over the 40 ns phase of the different LEUS simulations. The columns from left to right correspond to Glc, Deo and Gal (Figure 3.1) using the 53A6 force field[244] (middle row) or the 56A6$_{\text{CARBO}}$ force field[245] (bottom row). The bars follow the series S_h^p of artificial solvents (Table 3.3) in order of decreasing H-bonding capacity at low permittivity. The values of the solvent permittivity ε (black circles, scale on the left) and number of H-bonds per molecule n_H (red triangles, scale on the right) are also displayed (top row). The conformer populations refer to the canonical rotamers gg, gt and tg (Figure 3.1). The simulations of Deo with 53A6 are omitted due to ambiguous torsional potential definitions. The data is reported numerically in Suppl. Mat. Table 3.S.8.

that the comparatively much larger changes observed along the physical solvent series are not caused by a change in the solvent H-bonding capacity.

In summary, the study of the four artificial solvent series suggests that, for Glc and Gal, decreasing solvent permittivity or H-bonding capacity both promote an increase in the population of the potentially H-bonding conformer (tg for Glc and gg for Gal). However, these effects appear to be cooperative rather than additive, $i.e.$ the influence of either of

the two parameters only becomes significant when the other parameter is already low. In contrast, for Deo, the two types of changes have a relatively limited effect and the trends are not those expected on the basis of the solvent-dependence of the *gauche*-effect. This suggests that the more pronounced trends observed for this compound along the physical solvent series are primary related to other properties of the solvent molecules such as shape and size, possibly along with the influence of a $H_6 \rightarrow O_5$ H-bond.

3.3.4   Relative free energies of the rotamers

The trends noted in the previous section along the five series of solvents are clearly dominated by the influence of the solvent on the population of the potentially H-bonding conformer. However, they are somewhat shadowed by the fact that the populations of the two non-H-bonding conformers also vary, as the populations must add up to 100%. In addition, population variations for Deo, in which the $O_4 \leftrightarrow O_6$ H-bond is precluded, reveal that other influences may be active for the three compounds, such as the solvent-dependence of the *gauche* effect, the possible presence of a weak $H_6 \rightarrow O_5$ H-bond and, for the physical solvent series, the variation of the shape and size of the solvent molecule.

One may attempt to factor out these secondary effects by monitoring for each solvent the relative free energy

$$
\begin{aligned}
\Delta\Delta G(\text{R}, \text{C}) &= [G(\text{R}, \text{C}) - G(gt, \text{C})] - [G(\text{R}, \text{Deo}) - G(gt, \text{Deo})] \\
&= -k_B T \ln \frac{p(\text{R}, \text{C})p(gt, \text{Deo})}{p(gt, \text{C})p(\text{R}, \text{Deo})},
\end{aligned}
\tag{3.6}
$$

where R and C denote a rotamer and a compound, respectively, k_B is the Boltzmann constant and T the absolute temperature. These relative free energies, simply noted $\Delta\Delta G$ below, remove as a baseline possible solvent-induced shifts in the free energies of all conformers of Deo, and of the non-H-bonding conformer gt of Glc and Gal. As a result,

$\Delta\Delta G$ is only non-zero for the *gg* and *tg* conformers of Glc and Gal, and should mainly characterize the solvent-induced shift in the relative free energies of these two rotameric states caused by the $O_4 \leftrightarrow O_6$ H-bond.

The results of this analysis are displayed in Figure 3.8 in terms of $\Delta\Delta G$ values along the five solvent series. The curves strikingly illustrate that when the above secondary effects are disregarded, the solvent influences exclusively the relative free energy of the potentially H-bonding conformer (*tg* for Glc and *gg* for Gal), with nearly no effect on that
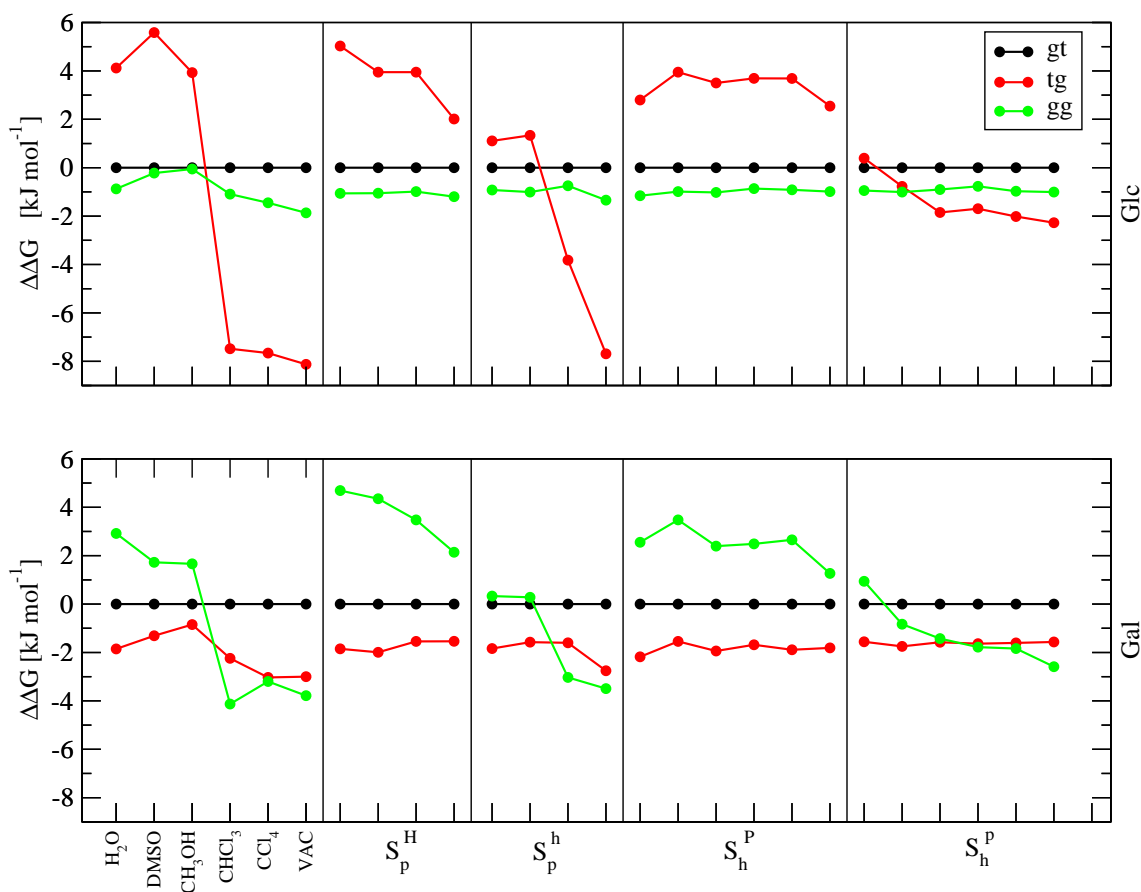
Figure 3.8: Relative free energies $\Delta\Delta G$ of 3.6 for Glc and Gal in the series of physical solvents at 298.15 K and 1 bar, and the four series of artificial solvents at 298.15 K and 998.3 kg·m^{-3}. The values are displayed for Glc (top) and Gal (bottom) for the three canonical rotamers *gg*, *gt* and *tg* (Figure 3.1) along the series of physical solvents (Table 3.2) and the four series of artificial solvents (Table 3.3), as calculated using the 56A6$_{CARBO}$ force field[245].

of the other conformer.

Considering the artificial solvent series, the $\Delta\Delta G$ curves for tg in Glc and gg in Gal present similar features, see for example the large drop between the second and third solvents in series S_p^h and the somewhat higher values for the second and fifth solvents in series S_h^P. These irregularities arise because the selected solvents span corresponding ranges of permittivity ε and H-bonding capacity n_H in a somewhat inhomogeneous rather than smoothly continuous fashion (Table 3.3). The observed variations show that reducing either of the two parameters ε or n_H only has a significant influence on the relative free energy of the H-bonding rotamer when the other parameter is already low. In series S_p^H and S_h^P, one of the two parameters is decreased while the other remains water-like, and the $\Delta\Delta G$ value for the H-bonding conformer decreases only by about 1-3 kJ mol^{-1}. In series S_p^h and S_h^p, one parameter is decreased while the other is already much lower than the value appropriate for water, and the $\Delta\Delta G$ value for the H-bonding conformer decreases much more pronouncedly, by about 2-9 kJ mol^{-1}. The influence of ε (over the considered range 115 to 1) also appears to be intrinsically larger than that of n_H (over the considered range 3.6 to 1.5), and is more pronounced for Glc compared to Gal.

Along the series of physical solvents, the variations of $\Delta\Delta G$ for the non-H-bonding conformer are slightly more important, but remain limited (range of about 2 kJ mol^{-1}). The variations for the H-bonding conformer clearly split the series into two groups, namely polar (H$_2$O, DMSO and CH$_3$OH) and non-polar (CHCl$_3$, CCl$_4$ and VAC) solvents. The difference in $\Delta\Delta G$ between the two groups is more pronounced for Glc compared to Gal (about 13 $vs.$ 7 kJ mol^{-1}, respectively). It is also larger than the corresponding variation along the artificial solvent series S_p^h, the one presenting the most pronounced trends (about 9 and 4 kJ mol^{-1}, respectively, for Glc and Gal).

3.3.5   Hydrogen-bonding

The occurrence of intramolecular and solute-solvent H-bonds during the simulations was monitored considering the hydroxyl groups at C_4 (except Deo) and C_6, as well as the ring oxygen atom O_5. The corresponding numbers of H-bonds in a given instantaneous configuration were averaged (including reweighting) either over the entire trajectory (average number N), or separately over subsets of configurations presenting each of the three hydroxymethyl rotameric states (average numbers N_{gg}, N_{gt} and N_{tg}). Note that N is not equal to the sum of N_{gg}, N_{gt} and N_{tg}, but to the corresponding population-weighted sum. The detailed results of this analysis considering the two force fields can be found in Suppl. Mat., graphically in Figures 3.S.4-3.S.11 and numerically in Tables 3.S.9-3.S.12. To facilitate the discussion, a subset of these results is presented synoptically in Figure 3.9, restricting the discussion to the $56A6_{CARBO}$ force field. The results for 53A6 differ quantitatively, but the trends are qualitatively similar.

The occurrence of intramolecular H-bonds is illustrated in the left column of Figure 3.9. The first row in each graph shows the conformer populations P_{gg}, P_{gt} and P_{tg} (reported from Figures 3.4-3.7). The second and third rows show the average numbers N_{gg}, N_{gt}, N_{tg} and N of $O_4 \leftrightarrow O_6$ and $H_6 \rightarrow O_5$ H-bonds, respectively, where $O_4 \leftrightarrow O_6$ indicates either of the flip-flop variants $H_4 \rightarrow O_6$ and $H_6 \rightarrow O_4$. Since at most one H-bond $O_4 \leftrightarrow O_6$ and one H-bond $H_6 \rightarrow O_5$ can be formed in a given instantaneous configuration, the values can be interpreted as fractional occurrences.

As expected based on geometric considerations, the formation of a $O_4 \leftrightarrow O_6$ H-bond is nearly exclusively compatible with the *tg* conformer of Glc and the *gg* conformer of Gal, and impossible for Deo. For the two former compounds, the *gt* conformer expectedly presents rigorously no occurrence of this H-bond. The same approximately holds for the third conformer, *gg* for Glc or *tg* for Gal, although marginal occurrences can be observed
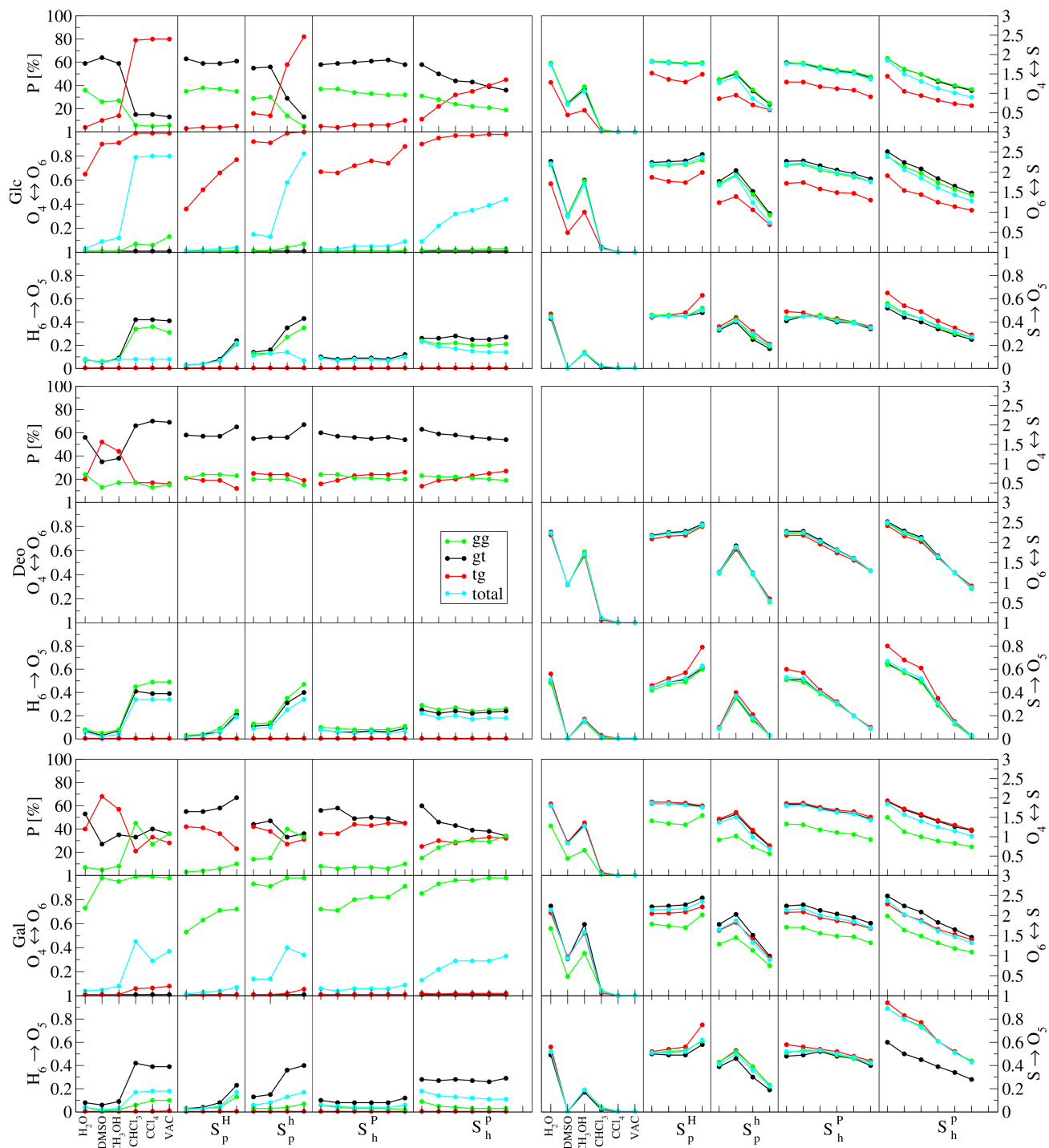
Figure 3.9: [Previous page] Average numbers of intramolecular and solute-solvent H-bonds for the three compounds considered in the series of physical solvents at 298.15 K and 1 bar, and the four series of artificial solvents at 298.15 K and 998.3 kg·m^{-3}. The values are calculated as reweighted averages over the 40 ns phase of the different LEUS simulations (plain average over 1μs SD simulations for VAC). All results pertain to the 56A6$_{\text{CARBO}}$ force field[245]. The panels from top to bottom correspond to Glc, Deo and Gal (Figure 3.1). The left and right columns correspond to intramolecular and solute-solvent H-bonds, respectively. In each panel of the left column, the rows from top to the bottom refer to the populations P of the three staggered hydroxymethyl rotamers, to the average numbers of O$_4$↔O$_6$ H-bonds, and to the average numbers of H$_6$→O$_5$ H-bonds. In each panel of the right column, the rows from top to the bottom refer to the the average numbers of O$_4$↔S, O$_6$↔S and S→O$_5$ H-bonds, where S is a solvent donor or acceptor site. The average numbers are calculated over the entire simulation (N) or over the three canonical rotamers separately (N_{gg}, N_{gt}, N_{tg}). A more detailed analysis of the H-bonds is provided in Suppl. Mat. Figures 3.S.4-3.S.7 (53A6 force field[244]) and Figures 3.S.8-3.S.11 (56A6$_{\text{CARBO}}$ force field[245]). The corresponding numerical values can be found in Suppl. Mat. Tables 3.S.9-3.S.12.

in some of the solvents with the lowest polarities. In this case, the H-bond has a distorted geometry that barely fits the cutoff criteria employed in the H-bond assignment.

The occurrence of the O$_4$↔O$_6$ H-bond within the potentially H-bonding conformation (N_{tg} for Glc or N_{gg} for Gal) is generally high and increases along the five solvent series. For the physical solvent series and the artificial solvent series S$_p^h$ and S_h^p, values very close to one are reached for the least polar solvents, *i.e.* the intramolecular H-bond is always formed when the hydroxymethyl rotameric state is compatible with its formation. Simultaneously, the formation of the H-bond starts to drive the rotameric equilibrium towards this specific conformer, as visible in the evolution of the populations P, but also in the increase of the occurrence N of this H-bond over the entire conformational ensemble. In addition, distorted H-bonds start to present marginal occurrences in the alternative rotamers gg for Glc and tg for Gal (see above), also indicative of a high conformational pressure towards intramolecular H-bonding in these solvents. In other words, for the least polar solvents in these three series, H-bond formation represents a strong driving force directing the conformational equilibrium of the hydroxymethyl group.

For the artificial solvent series S$_p^H$ and S_h^P, the occurrence of the O$_4$↔O$_6$ H-bond within

the potentially H-bonding conformer also increases, but remains below 0.9 even for the least polar solvent of the series. In addition, the frequent formation of this H-bond does not promote any shift in the rotameric populations P, and its occurrence N over the entire conformational ensemble remains below 0.1. In other words, for these two series as well as for the more polar solvents of the three other series (see above), H-bond formation represents an opportunistic consequence of the proximity of the H-bonding partners in a given rotameric state, but not a driving force directing the conformational equilibrium towards this state.

The $O_4 \leftrightarrow O_6$ H-bond is a flip-flop H-bond. As can be seen in Suppl. Mat. Figures 3.S.8-3.S.11, the ratio between the two orientations $H_4 \rightarrow O_6$:$H_6 \rightarrow O_4$ varies in the approximate range 90:10 to 10:90 for Glc and Gal, with a tendency to increase upon decreasing the solvent permittivity.

Based on geometric considerations, the formation of a $H_6 \rightarrow O_5$ H-bond is only compatible with the gg and gt conformers for the three compounds considered and indeed, the tg conformer presents zero occurrence of this H-bond. This H-bond is also expected to be intrinsically weaker than the $O_4 \leftrightarrow O_6$ H-bond, as it involves a five-atom rather than a six-atom H-bonding ring. It is found with approximately equal occurrences ranging from 0 to about 0.5, depending on the solvent, in the gt conformation of the three compounds and in the gg conformation of Glc and Deo. Its occurrence is significantly smaller for the gg conformation of Gal, due to competition with the stronger $O_4 \leftrightarrow O_6$ H-bond in this conformation.

The trends concerning the $H_6 \rightarrow O_5$ H-bond along the different solvent series are most visible considering Deo, which does not involve a competing $O_4 \leftrightarrow O_6$ H-bond. Along the physical solvent series and the artificial solvent series S_p^H and S_p^h, the occurrence of $H_6 \rightarrow O_5$ in the gg and gt conformations increases. For the least polar solvents in these series, it

represents a driving force promoting an increase in the corresponding populations. This suggests that the $H_6 \rightarrow O_5$ H-bond has a more important conformational influence than the solvent-dependence of the *gauche*-effect, which would be expected to stabilize the *tg* rotamer upon decreasing the solvent polarity. No significant changes are observed along the artificial solvent series S_h^P and S_h^p, where the occurrences are essentially constant at about 0.1 and 0.2, respectively, and the population shifts are marginal. When considering Glc and Gal, the trends in the occurrence of $H_6 \rightarrow O_5$ are similar to Deo (except for *gg* in Gal, see above), but the contribution of this H-bond to the conformational equilibrium is masked by the much larger effect of the $O_4 \leftrightarrow O_6$ H-bond.

The occurrence of solute-solvent H-bonds is illustrated in the right column of Figure 3.9. The three rows in each graph show the average numbers of $O_4 \leftrightarrow S$, $O_6 \leftrightarrow S$ and $S \rightarrow O_5$ H-bonds, where S denotes a solvent donor or acceptor site, and $O_4 \leftrightarrow S$ and $O_6 \leftrightarrow S$ indicate that two flip-flop variants are possible. Because solute oxygen atoms can form H-bonds with multiple solvent molecules in a given instantaneous configuration, the values N_{gg}, N_{gt}, N_{tg} and N should be interpreted here as average numbers of H-bonded solvent molecules.

In most cases, these numbers are highest for O_6 (primary hydroxyl group, most exposed to the solvent), slightly lower for O_4 (secondary hydroxyl group, slightly less exposed), and significantly lower for O_5 (least exposed and only an H-bond acceptor), typically ranging from 0.1 to 2.6 for O_6, from 0.1 to 1.9 for O_4 and from 0.0 to 0.9 for O_5. Expectedly, the numbers are also lower for the physical solvents DMSO (exclusively H-bond acceptor), $CHCl_3$ (exclusively and weak H-bond donor), CCl_4 (non-H-bonding) and VAC.

Along the artificial solvent series S_p^h and S_h^p, the numbers of solute-solvent H-bonds decrease markedly for the three compounds and the three oxygen atoms considered. For Glc and Gal, this is in line with the trends observed in the left column of Figure 3.9, *i.e.* as solute-solvent H-bonding is reduced, the solute compensates by increasingly populating the

intramolecular $O_4 \leftrightarrow O_6$ H-bond and the rotamer compatible with this bond. In contrast, along the artificial solvent series S_p^H and S_h^P, the numbers vary much less, with a slight tendential increase for S_p^H and decrease for S_h^P. Here also, for Glc and Gal, this is in line with the trends observed in the left column of Figure 3.9, *i.e.* since solute-solvent H-bonding is preserved, there is less need for intramolecular compensation by the $O_4 \leftrightarrow O_6$ H-bond.

The competition (or compensation) between intramolecular and solute-solvent H-bonds is clearly visible when the numbers of solute-solvent H-bonds are resolved in terms of contributing rotamers. For Deo, solute-solvent H-bonding is essentially the same for the three rotamers, with the exception of O_5 in the *tg* rotamer. In this rotamer, enhanced solute-solvent H-bonding is correlated with the impossibility of a weak $H_6 \rightarrow O_5$ H-bond. For Glc and Gal, solute-solvent H-bonding is noticeably weaker for O_4 and O_6 in the conformation enabling the $O_4 \leftrightarrow O_6$ H-bond (*tg* for Glc and *gg* for Gal). It is also slightly higher for O_5 in the *tg* conformation which cannot form a weak $H_6 \rightarrow O_5$ H-bond, as was observed for Deo.

3.3.6 Correlation between the exocyclic groups

The results of the analysis of all pairwise correlations between the exocyclic dihedral angles ω, ϕ, χ_2, χ_3 and χ_4 based on the simulations performed using the 56A6$_{\text{CARBO}}$ force field are displayed in Figure 3.10. Corresponding graphs for the 53A6 force field are provided in Suppl. Mat. Figure 3.S.12. Note that these results involve no reweighting, *i.e.* they correspond to an artificial situation where the probability distribution $p(\omega)$ is essentially flat and the hydroxymethyl group samples its three canonical conformations with approximately equal probabilities, irrespective of the compound and solvent. This eliminates the influence of the solvent on $p(\omega)$, discussed in details in the previous sections, permitting to

focus on the intramolecular orientational correlations. For this reason, one would expect the results for the 53A6 and 56A6$_{\text{CARBO}}$ parameter sets to be similar. In practice, however, important differences are observed, because the description of the rotameric preferences of the lactol group (dihedral angle ϕ) is not very accurate in 53A6 (see discussions in Refs.[170,245]). In the following, only the results for 56A6$_{\text{CARBO}}$ are discussed.

The preferential exocyclic-group orientations and intramolecular H-bonding patterns of hexopyranoses in vacuum have been extensively studied in the past by means of QM calculations[190,215,217,218,220–223,225,226,228,277–280], including for Glc[215,217,218,221,223,225,226,228,277–279] and Gal[228,280]. These studies have revealed low-energy conformations involving either clockwise (CW) or counterclockwise (CC) H-bonded networks[217,218,220,221,225,226,228,280], referring to the orientation of the network as viewed from the β-face of the ring. Most QM studies suggest that the lowest-energy pattern in vacuum is of the CC type for both Glc and Gal[215,217,218,221,223,225,226,228,277–279]. Illustrative structures for H-bonding networks encountered in the simulations are shown in Figure 3.10 for the three compounds. The values of the exocyclic dihedral angles corresponding to these structures are also marked on the correlation plots.

Figure 3.10: [Next page] Pairwise correlations between the exocyclic dihedral angles of Glc and Gal in the physical solvents H_2O and CCl_4 at 298.15 K and 1 bar, and in the artificial solvents $W_{1.3}^{0.4}$ and $W_{0.3}^{0.5}$ at 298.15 K and 998.3 kg·m^{-3}. The individual points correspond to structures sampled at 0.5 ps intervals along the 40 ns US phase of the different LEUS simulations using the 56A6$_{\text{CARBO}}$ force field[245]. The successive rows from top to bottom correspond to Glc, Deo and Gal (Figure 3.1). The solvents considered (Tables 3.2 and 3.3) are H_2O (black), CCl_4 (blue), $W_{1.3}^{0.4}$ (red) and $W_{0.3}^{0.5}$ (orange). The dihedral angles are defined as C_4-C_5-C_6-O_6 (ω), O_5-C_1-O_1-H_1 (ϕ), C_1-C_2-O_2-H_2 (χ_2), C_2-C_3-O_3-H_3 (χ_3) and C_3-C_4-O_4-H_4 (χ_4). On the right, three representative structures (A, B, C) for H-bonding networks (dotted lines) encountered in the simulations are displayed for each of the three compounds. Representative points for these specific structures are also shown on the correlation plots (green triangles).
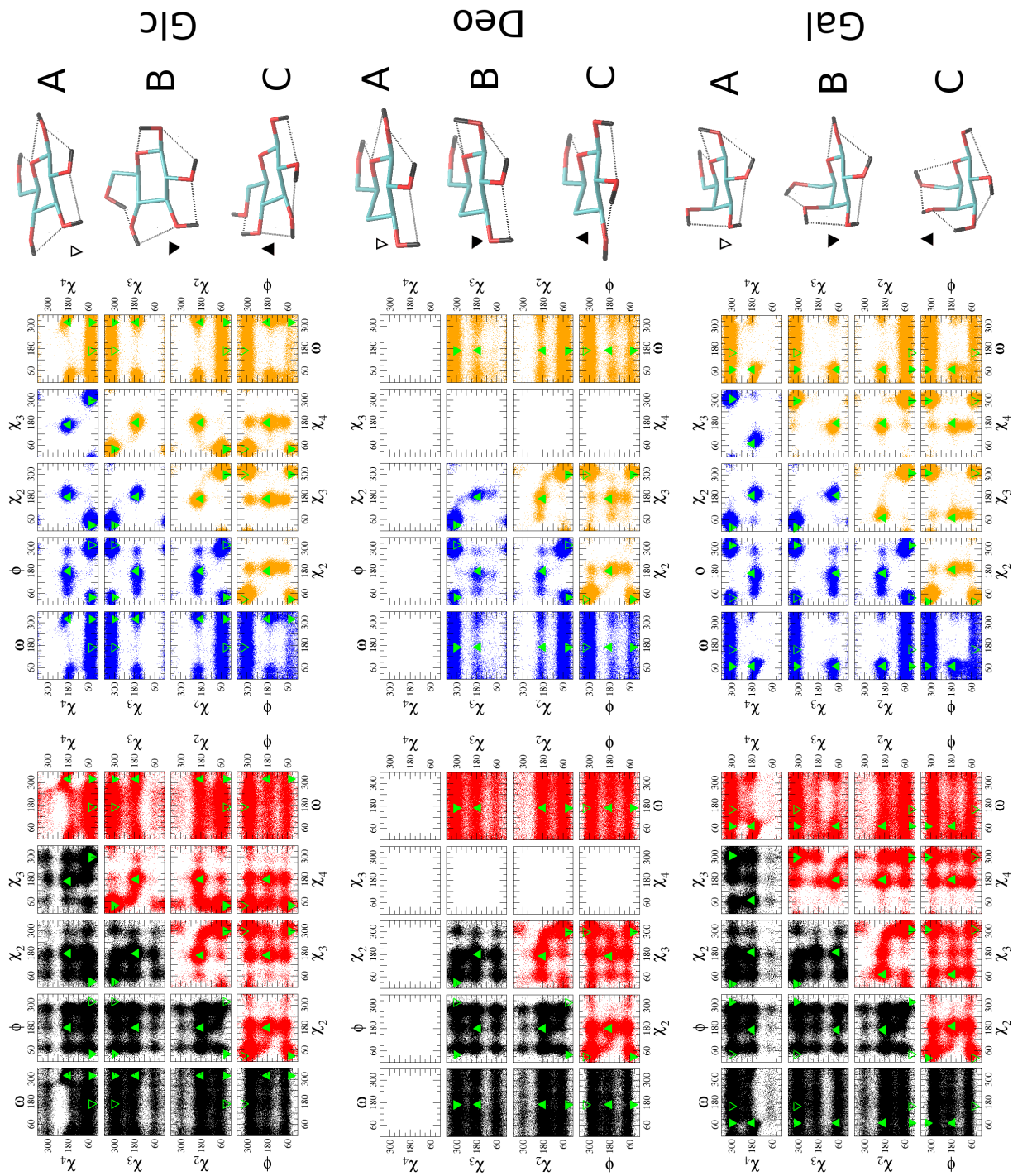
The first structure A involves dihedral angles $(\phi, \chi_2, \chi_3, \chi_4, \omega)$ in $(g^-, g^+, g^-, g^+, gg/gt)$ for Glc, in $(g^-, g^+, g^-, -, gg/gt)$ for Deo or in $(g^+, g^+, g^-, g^-, gg/gt)$ for Gal (the variant with ω in gt is represented in the figure). This structure is representative of a CC network involving an $O_4{\rightarrow}O_3{\rightarrow}O_2{\rightarrow}O_1{\rightarrow}O_5$ H-bond series for Glc and Gal ($O_3{\rightarrow}O_2{\rightarrow}O_1{\rightarrow}O_5$ for Deo) with an additional $H_6{\rightarrow}O_5$ H-bond. The second structure B involves dihedral angles $(\phi, \chi_2, \chi_3, \chi_4, \omega)$ in (g^+, g^+, g^-, g^+, tg) for Glc, in $(g^+, g^+, g^-, -, gt)$ for Deo or in (g^-, g^+, g^-, g^-, gg) for Gal.

This structure is also representative of a CC network. For Glc and Gal, it enables the formation of an extended $O_6{\rightarrow}O_4{\rightarrow}O_3{\rightarrow}O_2{\rightarrow}O_1{\rightarrow}O_5$ H-bond series. For Deo it present the same H-bonding pattern as A, except that ϕ is in g^+ instead of g^-. The third structure C involves dihedral angles $(\phi, \chi_2, \chi_3, \chi_4, \omega)$ in (t, t, t, t, tg) for Glc, in $(t, t, t, -, gt)$ for Deo or in (t, t, g^+, t, gg) for Gal. This structure is representative of a CW network. For Glc, it enables the formation of an $O_1{\rightarrow}O_2{\rightarrow}O_3{\rightarrow}O_4{\rightarrow}O_6$ H-bond series, for Deo of an $O_1{\rightarrow}O_2{\rightarrow}O_3$ series along with an $O_6{\rightarrow}O_5$ H-bond, and for Gal of an extended $O_1{\rightarrow}O_2{\rightarrow}O_3{\rightarrow}O_4{\rightarrow}O_6{\rightarrow}O_5$ series.

The results of Figure 3.10 for the solvent CCl_4 (blue) clearly evidence a conformational locking of the $(\phi, \chi_2, \chi_3, \chi_4)$ dihedral angles within either of the three H-bonding networks exemplified by the structures A, B or C, similar to what is observed in vacuum based on the available QM studies. In other words, in CCl_4, the correlation patterns in the dihedral angles can be entirely explained assuming that only these three networks are significantly populated. For Glc, pattern A involves the hydroxymethyl group in gg or gt and a $H_6{\rightarrow}O_5$ H-bond, whereas patterns B and C involve the hydroxymethyl group in tg and a $H_6{\rightarrow}O_4$ or a $H_4{\rightarrow}O_6$ H-bond, respectively. As seen earlier, for this solvent, the tg rotamer is overwhelmingly populated (80%), and the present analysis shows that it is not only associated with a very high occurrence of the $O_4{\leftrightarrow}O_6$ H-bond, but also of the compatible extended H-bonding patterns B and C (each corresponding to one of

the flip-flop variants of this H-bond). Similarly, the *gt* rotamer, which is also populated (15%), is not only associated with a high occurrence of the $H_6 \to O_5$ H-bond, but also of the compatible extended H-bonding pattern A. Analogous considerations apply to Gal, where pattern A involves again the hydroxymethyl group in *gg* or *gt* and a $H_6 \to O_5$ H-bond, whereas patterns B and C involve the hydroxymethyl group in *gg* and a $H_6 \to O_4$ or a $H_4 \to O_6$ H-bond, respectively. Here also, the different rotamers are not only associated with the formation of the indicated H-bond, but of the compatible extended H-bonding patterns A, B or C. For Deo, the three populated patterns are also A, B and C, but due to the lack of a hydroxyl group at C_4, the $(\phi, \chi_2, \chi_3, \chi_4)$ dihedral angles are only weakly correlated with the rotameric state determined by ω.

Considering in particular the correlation between ϕ and ω, it is seen that the three states of ω are compatible with any state of ϕ. In contrast, for Glc and Gal, the *trans*-state of ϕ is essentially precluded in the *gt* rotamer of both compounds.

In sharp contrast to the case of CCl_4, the results of Figure 3.10 for the solvent H_2O (black) evidence a high conformational flexibility of the $(\phi, \chi_2, \chi_3, \chi_4)$ dihedral angles, which can adopt either of their three staggered orientations in a largely uncorrelated fashion. Still, the areas of the correlation plots presenting a negligible density in CCl_4 tend to also present a reduced density in H_2O, indicating that the preference for the networks A, B and C persist, albeit in a highly damped form.

Also shown in Figure 3.10 are results for the artificial solvent presenting the lowest H-bonding capacity along with the lowest permittivity ($W_{0.3}^{0.5}$ in series S_p^h; $\varepsilon = 1$, $n_H = 1.4$; orange), or an intermediate permittivity ($W_{1.3}^{0.4}$ in series S_h^p; $\varepsilon = 12$, $n_H = 1.5$; red). For the first solvent, the correlations are very similar to those observed in CCl_4. In contrast, for the second solvent, the correlations are more similar to those observed in H_2O. Still, the areas of the correlation plots presenting a negligible density in CCl_4 tend to also

present a reduced density in this solvent compared to H_2O, *i.e.* the behavior is in this case somewhat intermediate between those in the two physical solvents. This suggests again that the influence of the permittivity on the conformational equilibrium is more important than that of the H-bonding capacity.

3.3.7   J-coupling constants

In the case of the physical solvents H_2O, DMSO and CH_3OH, it is possible to compare the simulation results for the rotameric probability distributions $p(\omega)$ to primary experimental NMR data in the form of scalar coupling constants (J-values). The relevant coupling constants are the vicinal coupling constants between the proton at carbon atom C_5, and the R or S protons at the methylene group C_6, noted J_R and J_S for simplicity.

An extensive list of experimental J_R and J_S values reported for Glc or Gal in H_2O, DMSO or CH_3OH can be found in Suppl. Mat. Table 3.S.2. The values are provided for the two compounds in either the α- or the β-anomeric form, and either as a free hexopyranose or as a O_1-methyl-hexopyranoside. For H_2O, this corresponds to a set of 42 and 29 independent measurements for Glc and Gal, respectively, covering the year range 1976-2008. The reported J-values clearly differ between the Glc and the Gal data subsets. However, the differences between the anomeric forms and functionalizations for either of the two compounds are not statistically significant compared to the spread in the values reported by different sources. This suggests that, at least in a highly polar solvent such as water, the lactol group is too distant from the hydroxymethyl group for exerting a significant influence on its conformational properties. This suggestion is in line with the results of independent Glc and Gal simulations[246] (see Table 7 therein) showing a limited influence of O_1-anomerization or/and -methylation on the calculated J_R and J_S values in water. It is also supported by the analysis of exocyclic dihedral-angle correlations

(previous section) suggesting an absence of significant correlation between ϕ and ω, except for the least polar solvents where extended H-bonding networks start to trigger longer-ranged exocyclic-dihedral correlations. Considering the Glc and the Gal data subsets without discriminating for anomerization and O_1-methylation, it also appears that the reported J_R and J_S values do not present a clear convergence trend as a function of the determination year, which would suggest an accuracy increase over the years due to the availability of improved NMR spectrometers.

For the above reasons, reference experimental values for J_R and J_S were averaged separately for Glc and Gal and for the three solvents over all the corresponding experimental determinations irrespective of the anomerization and O_1-methylation. These averages are reported in Table 3.5, along with a crude error estimate provided by the corresponding standard deviations. For DMSO and CH_3OH, the data sets are considerably smaller than for water, with 4 and 3 (DMSO) or 5 and 2 (CH_3OH) independent measurements for Glc and Gal, respectively. The estimated errors are on the order of 0.1-0.3 s^{-1} for the two

Compound	Solvent	N. data	Exp.		Sim.	
			J_R [s^{-1}]	J_S [s^{-1}]	J_R [s^{-1}]	J_S [s^{-1}]
	H_2O	42	5.8 (0.1)	2.1 (0.1)	6.2 (0.1)	2.7 (0.2)
Glc	DMSO	4	5.9 (0.1)	1.6 (0.1)	6.7 (0.1)	3.0 (0.2)
	CH_3OH	5	5.5 (0.1)	1.9 (0.1)	6.5 (0.1)	3.4 (0.2)
	H_2O	29	7.8 (0.2)	4.5 (0.3)	6.5 (0.1)	5.3 (0.1)
Gal	DMSO	3	6.3 (0.1)	6.1 (0.1)	4.5 (0.1)	7.3 (0.1)
	CH_3OH	2	6.4 (0.1)	5.6 (0.1)	5.1 (0.1)	6.4 (0.1)

Table 3.5: Comparison between calculated and experimental J-coupling constants for Glc and Gal in H_2O, DMSO and CH_3OH at 298.15 K and 1 bar. The J-values J_R and J_S correspond to the coupling between the proton H5 at C_5 and the protons H6R and H6S at C_6, respectively. The values from the simulations are calculated based on the 40 ns US phase of the LEUS simulations by application of the Karplus equations of Hasnoot *at al.*[241], Tafazzoli and Ghiasi[242], or Stenutz *et al.*[191] (Eqs. 3.1-3.5), and the averaged over the three results is reported along with the associated standard deviation. The values from experiments are based on Suppl. Mat. Table 3.S.2, where the average over the listed measurements is reported along with the associated standard deviation. The compounds considered are Glc and Gal (Figure 3.1). The solvents considered are H_2O (green), DMSO (blue), and CH_3OH (yellow). The force field considered is 56A6$_{CARBO}$[245]. The data is illustrated graphically in Figure 3.11.

compounds in the different solvents.

Corresponding estimates from the simulations were obtained based on the $p(\omega)$ distributions using the Karplus equations of Haasnoot *et al.*[241], Tafazzoli & Ghiasi[242], or Stenutz *et al.*[191] (Eqs. 3.1-3.5). The results are also reported in Table 3.5, in the form of averages over the three Karplus curves, along with the corresponding standard deviation as an error estimate. The estimated errors are on the order of 0.1-0.2 s^{-1} for the two compounds in the different solvents.

The comparison between experimental and simulated J-values is shown in Figure 3.11, in the form of correlation plots, and of a graph illustrating how the convolution of population profiles $p(\omega)$ with a Karplus equation leads to resulting averaged J_R and J_S estimates. For Glc, the effect of the solvent on J_R and J_S is limited, both experimentally and in the simulations (variations over a range of about 0.5-1 s^{-1}). This is mainly because the corresponding $p(\omega)$ variation among the three solvents is limited. In the simulations, a small population shift from gg (low values of J_R and J_S) to tg (higher values of J_R and

Figure 3.11: [Next page] Comparison between calculated and experimental J-coupling constants for Glc and Gal in H$_2$O, DMSO and CH$_3$OH at 298.15 K and 1 bar. In the left part, correlation between experimental J-coupling constants and values calculated based on the 40 ns US phase of the LEUS simulations. The J-values J_R and J_S correspond to the couplings between the proton H5 at C$_5$ and the protons H6R and H6S at C$_6$, respectively. The values from the simulations are calculated by application of the Karplus equations of Hasnoot *et al.*[241], Tafazzoli & Ghiasi[242], or Stenutz *et al.*[191] (Eqs. 3.1-3.5), and the average over the three results is reported (point) along with the associated standard deviation (horizontal bar). The values from experiments are from Suppl. Mat. Table 3.S.2, where the average over the listed measurements is reported (point) along with the associated standard deviation (vertical bar). The compounds considered are Glc and Gal (Figure 3.1). The solvents considered are H$_2$O (green), DMSO (blue) and CH$_3$OH (yellow). The force field considered is 56A6$_{\text{CARBO}}$[245]. The data is reported numerically in Table 3.5. In the right part, comparison between the Karplus curve by Hasnoot *et al.*[241] as a function of the dihedral angle ω and the normalized probability distributions $p(\omega)$ around the dihedral angle ω for Glc and Gal in the different solvents based on the simulations using the 56A6$_{\text{CARBO}}$[245] force field. The ranges of the Karplus curve relevant for the calculation of the average J-coupling values are marked with crosses.
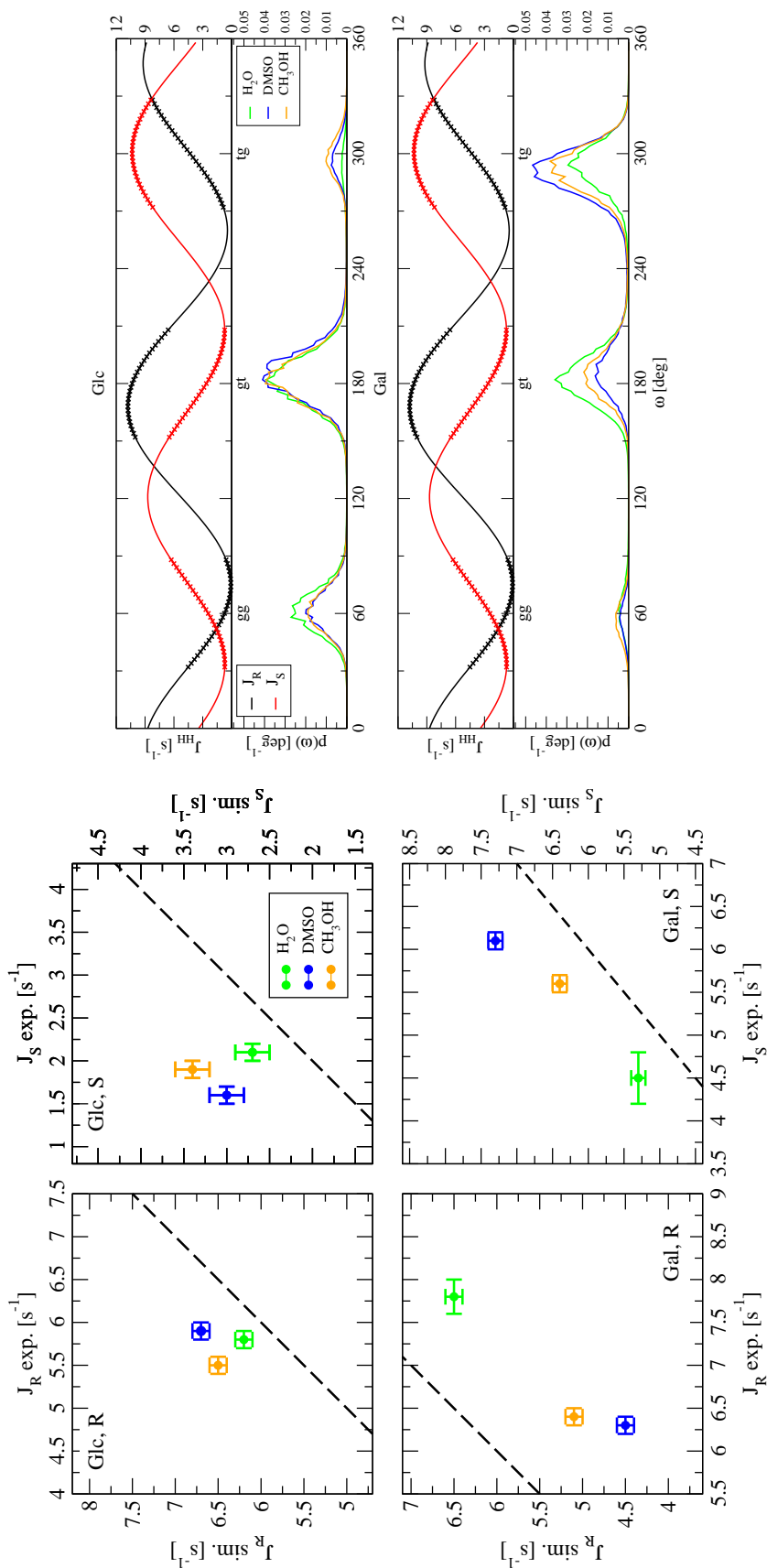
J_S) between H_2O and the two other solvents leads to a slight increase in both calculated J-values. The trend is not clearly visible in the corresponding experimental values, but these only vary over a range of about 0.5 s^{-1}. In contrast, for Gal, the effect of the solvent on J_R and J_S is larger (variations over a range of about 2 s^{-1}). In the simulations, a significant population shift from gt (high value of J_R, low value of J_S) to tg (lower value of J_R, higher value of J_S) when changing from H_2O to CH_3OH and then to DMSO leads to a systematic decrease of J_R and increase of J_S along the series. The same trend is observed experimentally.

Although the simulations capture qualitatively well the experimental trends, the deviations between calculated and experimental J-values remain on the order of $0.5\text{-}1.5 \text{ s}^{-1}$, smaller for Glc than for Gal. This is somewhat larger than the sum of the errors estimated based on the spread in the experimental data and Karplus-equation results. The discrepancy, which may be at least in part ascribed to force-field inaccuracies, appears in the form of a constant offset, especially for Gal. This suggests that another main source of inaccuracy may actually reside in the positions and widths of the peaks in $p(\omega)$, rather than in the corresponding populations. In fact, based on the results for H_2O shifting the positions of each of the three peaks in the calculated $p(\omega)$ to the left or to the right (eight possible combinations) by as little as $5°$ (or $10°$) may induce changes of up to 0.5 s^{-1} (or 1.0 s^{-1}) in J_R and J_S. Such small displacements are of the same order as the shifts relative to canonical values typically applied in simple population models used to interpret the NMR data[189,191,204] and the shifts observed between the two force-field versions (see Section 3.3.3).

3.4   Conclusions

Intra- or intermolecular H-bonding is commonly regarded as a major driving force in (bio)chemical processes such as conformational changes or host-guest binding. For example, considerations involving optimal H-bonding patterns have led to spectacular predictions concerning the native structures of proteins[126] and nucleic acids[127] in the 50's. However, the view proposed here and in Ref.[132] is that in an aqueous environment, the major conformational influence of H-bonding is limited to buried (as opposed to solvent-exposed) H-bonds and predominantly represents in this case a steering (as opposed to driving) force. More specifically, conformations involving mismatches among buried polar groups are strongly penalized, but buried H-bonded groups do not *per se* contribute significantly to the stability of a conformation.

In contrast, solvent-exposed H-bonds are considered to represent a minor (possible negligible or even, in some cases, adverse) conformational driving as well as steering force in an aqueous environment. They should be viewed as an opportunistic consequence of the close proximity of two H-bonding groups in a given conformation, and not as a factor contributing to the stability of this conformation. In the present context, this view is compatible with the experimental observation that the hydroxymethyl rotamers of Glc and Gal permitting the formation of a solvent-exposed H-bond between this group and the hydroxyl group at C_4 are not favored but, in the opposite, least populated in water.

The above considerations pertain to an aqueous environment where the solvent-exposed H-bonded interaction is nearly entirely screened by the solvent dielectric response (high solvent dielectric permittivity) as well as subject to intense H-bonding competition by the solvent molecules (high solvent H-bonding capacity). As the polarity of the solvent is decreased, solvent-exposed intramolecular H-bonding progressively evolves from a negligible (here, adverse) to a very significant (favorable) conformational driving force. This evolu-

tion is probed here using MD simulations to investigate the rotameric preferences of the hydroxymethyl group of Glc and Gal as well as Deo, an analogous compound lacking the hydroxyl group at C_4, along physical and artificial solvent series of decreasing polarity. The artificial series also permit to probe separately the effects of the solvent permittivity and H-bonding capacity, while factoring out a possible influence of the dispersive interactions, shape and size of the solvent molecule.

As the solvent polarity is decreased, the populations of the conformers compatible with the formation of a flip-flop $O_4 \leftrightarrow O_6$ intramolecular H-bond (*tg* of Glc and *gg* of Gal) indeed increase significantly. Because it lacks a hydroxyl group at C_4, Deo evidences comparatively much less pronounced trends, related in part to the possible formation of an alternative (weaker) $H_6 \rightarrow O_5$ H-bond. Considering the artificial solvent series, the effects of the solvent permittivity and H-bonding capacity appear to mutually enhance each other, *i.e.* the effect of varying one of the two solvent parameters is only significant when the other parameter is low. Thus, in water, the weak (slightly adverse) role of the solvent-exposed intramolecular H-bond on the hydroxyl rotameric preferences is due to both dielectric screening and H-bond competition by the solvent, either of the two being sufficient alone to suppress the conformational pressure towards the formation of this H-bond. In lower polarity solvents, however, the influence of the solvent permittivity appears to be intrinsically stronger than that of the solvent H-bonding capacity.

The analysis of the intramolecular H-bond occurrences and of the exocyclic dihedral-angle correlations also revealed two additional features.

First, even in high polarity solvents where the solvent-exposed H-bonding does not drive the hydroxymethyl rotameric equilibrium, the intramolecular H-bond is still found with a very high occurrence (*e.g.* about 50% for $O_4 \leftrightarrow O_6$ in water), a situation described above as opportunistic H-bond formation. As the solvent polarity is decreased, the occurrence

increases to nearly 100%, This is associated with of a raise of the conformational pressure towards H-bond formation, *i.e.* the H-bond starts to drive the rotameric equilibrium. Expectedly, this change is also correlated with (and, in part, caused by) a decrease of the number of solute-solvent H-bonds for the interacting partners. The key point here is that direct (*e.g.* in MD simulations) or indirect (*e.g.* experimentally *via* IR or NMR) observation of a persistent intramolecular H-bond is a necessary but by no means a sufficient condition for ascribing to this H-bond an important role in determining the preferred molecular conformation, a conclusion that is often too quickly reached in the literature.

Second, the hydroxymethyl conformation in hexopyranoses is a particularly sensitive probe for the above solvent effects compared *e.g.* to simpler organic diols[281]. The reason is that for these compounds, the $O_4 \leftrightarrow O_6$ H-bond (or the weaker $O_6 \rightarrow O_5$ H-bond) is only one out of up to five H-bonds constituting extended clockwise or counterclockwise networks around the ring, each of which also being affected by the solvent polarity (amplification effect).

The present study illustrates how MD simulations involving unphysical situations (here, artificial solvents) can be used to shed more light onto physical ones.

3.S.1   Supplementary Material

Glc

Svt.	Compound	Year/Ref.	gg [%]	gt [%]	tg [%]
H$_2$O	α	1984[180]	56	44	0
		1988[181]	56	44	0
		1988[181]	58	56	-14
		1988[181]	58	46	-4
		1992[179]	49	49	2
		2000[183]	47	54	-1
		2004[184]	40	53	7
		2004[184]	41	51	8
	β	1975[178(179)]	49	49	2
		1984[180]	53	45	2
		1988[181]	53	45	2
		1988[181]	54	57	-11
		1988[181]	54	48	-2
		1994[182(181)]	49	49	2
		2000[183]	45	62	-7
		2004[184]	31	61	8
		2006[185]	31	59	10
	α-Me	1974[282(189)]	56	39	5
		1984[180]	57	38	5
		1988[181]	58	38	4
		1988[181]	59	49	-8
		1988[181]	59	41	0
		1988[181(189)]	58	41	1
		1992[202]	57	38	5
		1994[189]	60(54)	51(46)	-11(0)
		1994[189]	63(54)	52(46)	-15(0)
		1994[189]	69(55)	55(45)	-24(0)
		1994[189]	53	47	1
		1994[189]	57(53)	50(47)	-7(0)
		1998[204]	48	48	4
		2002[190]	31	76	-7
		2002[190]	65	53	-18
	β-Me	1984[180(181)]	50	47	3
		1988[181]	50	47	3
		1988[181]	51	58	-9
		1988[181]	51	49	0
		1992[179]	55	45	1
		1994[182(181)]	48	50	2
		1994[189]	50	56	-6
		1995[283]	20	84	-4
		1995[283]	7	88	5
		2002[284]	41	52	7
		2002[190]	23	85	-8
		2002[190]	50	62	-13
DMSO	α-Me	1984[180]	55	45	0
		1984[180]	53	47	0
		1988[181]	55	45	0
		1988[181]	57	57	-14
		1988[181]	57	47	-4
		1992[202]	53	47	0
	β-Me	1992[179]	51	48	1
		1994[182]	51	47	2
CH$_3$OH	α-Me	1994[189]	56(52)	51(48)	-7(0)
		2002[190]	27	80	-7
		2002[190]	62	53	-15
	β-Me	2002[190]	24	81	-5
		2002[190]	59	56	-15
		1992[179]	52	48	0
		1994[182]	51	48	0
Force Fields	CHARMM/CSFF	2002[236]	66	31	1
	AMBER/Glycam (α-Me)	2001[235]	54	40	6
	OPLS/AA-SEI	2002[285]	69	27	4
	GROMOS/Spieser	1999[287]	48	47	5
	GROMOS 45A4	2005[248]	55	45	0
	GROMOS 45A4	2007[170]	57	43	0
	GROMOS 56A6$_{CARBO}$ (β-Me)	2011[245]	37	60	3
	GROMOS 56A6$_{CARBO}$	2015[246]	35	60	5

Gal

Svt.	Compound	Year/Ref.	gg [%]	gt [%]	tg [%]
H$_2$O	α	1983[186]	25	30	45
		1983[186]	18	38	44
		1983[186]	21	36	43
		1987[187]	21	54	25
		1987[187]	17	63	20
		1987[187]	20	60	20
		1988[181]	22	54	24
		1988[181]	17	63	20
		1988[181]	15	62	23
		2004[184]	3	74	23
		2004[184]	0	74	26
	β	1983[186]	27	56	17
		1983[186]	24	66	10
		1983[186]	27	62	11
		1984[180(181)]	22	53	25
		1987[187]	22	53	25
		1987[187]	18	62	20
		1987[187]	21	59	20
		1987[188]	14	65	21
		1988[181]	22	53	25
		1988[181]	18	62	20
		1988[181]	16	61	23
		1994[189]	12	56	32
		2001[190]	15	71	14
		2002[191]	3	67	30
		2004[184]	3	72	25
		2004[184]	0	71	29
	α-Me	1983[186]	19	57	24
		1983[186]	14	66	20
		1983[186]	17	62	21
		1984[180]	14	47	39
		1987[187]	21	61	18
		1987[187]	13	70	17
		1987[187]	20	67	13
		1988[181]	21	61	18
		1988[181]	13	70	17
		1988[181]	15	69	16
		1992[202]	14	47	39
		1994[182]	11	64	25
		2002[190]	40	69	-9
		2002[190]	23	63	14
	β-Me	1983[186]	25	52	23
		1983[186]	21	61	18
		1983[186]	25	57	18
		1984[180(181)]	22	55	23
		1987[187]	22	55	23
		1987[187]	17	65	18
		1987[187]	21	62	18
		1988[181]	22	55	23
		1988[181]	17	65	18
		1988[181]	15	64	21
		1994[182(202)]	17	58	24
		2002[190]	45	54	1
		2002[190]	29	53	18
DMSO	α-Me	1988[181]	27	30	43
		1988[181]	20	38	42
		1988[181]	21	39	40
	β-Me	1988[181]	32	27	41
		1988[181]	27	35	38
		1988[181]	27	37	36
CH$_3$OH	α-Me	2002[190]	57	46	-3
		2002[190]	36	19	45
	β-Me	2002[190]	57	47	-4
		2002[190]	33	28	38
Force Fields	CHARMM/CSFF	2002[236]	4	75	21
	AMBER/Glycam (α-Me)	2001[235]	8	64	28
	OPLS/AA-SEI	2002[285]	9	53	38
	GROMOS/Spieser	1999[287]	5	70	25
	GROMOS 45A4	2005[248]	34	41	25
	GROMOS 45A4	2007[170]	33	41	26
	GROMOS 56A6$_{CARBO}$ (β-Me)	2011[245]	7	67	26
	GROMOS 56A6$_{CARBO}$	2015[246]	6	52	42

Table 3.S.1: Experimentally inferred population estimates for the three staggered rotamers of the hydroxymethyl group of Glc and Gal, along with a few calculated values. The anomery (α or β) and possible O$_1$-methylation (α-Me or β-Me) of the compound considered as well as the solvent (H$_2$O, DMSO or CH$_3$OH) are indicated. Selected values from force-field simulations in water are also listed, referring to the unmethylated β-anomer unless otherwise specified. Values between parentheses represent alternative estimates with positiveness constraints on the populations. Multiple entries for the same literature reference correspond to the use of different raw J-values, Karplus equations or/and population models. Secondary references between parentheses indicate that the raw data of the primary reference is quoted as analyzed and reported in the secondary reference.

Glc				
Solvent	Compound	Year/Ref.	J_R [s⁻¹]	J_S [s⁻¹]
H$_2$O	α	1976[288]	5.8	2.0
		1976[289]	-	1.9
		1977[290]	5.8	2.0
		1977[291]	6.0	1.5
		1983[186]	5.7	2.8
		1984[180]	5.8	1.0
		1988[181]	5.8	1.9
		1995[292]	5.5	2.3
		2000[183]	5.40	2.20
		2004[184]	5.6	2.3
		2007[242]	5.4	2.3
		2008[293]	5.40	2.31
	β	1976[288]	5.8	2.0
		1977[290]	5.8	2.0
		1977[291]	5.4	1.6
		1983[186]	5.7	2.8
		1984[180]	6.0	2.1
		1988[181]	6.0	2.1
		1995[292]	6.0	2.3
		2000[183]	5.95	1.95
		2004[184]	6.2	2.3
		2007[242]	6.0	2.3
		2008[293]	5.95	2.27
	α-Me	1976[288]	5.4	2.2
		1977[291]	5.4	2.2
		1977[290]	5.4	2.2
		1983[186]	5.8	2.8
		1984[180]	5.4	2.3
		1988[181]	5.4	2.3
		1992[179]	5.62	2.24
		1994[189]	5.49	2.39
		2002[190]	5.4	2.2
		2008[293]	5.58	2.32
	β-Me	1976[288]	5.8	2.1
		1977[290]	5.8	2.1
		1977[291]	6.1	1.7
		1983[186]	6.4	2.4
		1988[181]	6.2	2.3
		1992[179]	5.89	2.00
		2002[191]	5.8	2.0
		2002[190]	5.9	2.0
		2008[293]	6.14	2.31
DMSO	α-Me	1984[180]	6.1	1.0
		1988[181]	5.9	1.9
	β-Me	1992[179]	5.56	1.93
5%H$_2$O/DMSO	α-Me	1984[180]	5.9	1.5
CH$_3$OH	α-Me	1992[179]	5.65	2.37
		1994[189]	5.64	2.40
		2002[190]	5.5	2.4
	β-Me	1992[179]	5.63	1.81
		2002[190]	5.25	1.95
Pyridine	β-Me	1992[179]	5.76	2.32

Gal				
Solvent	Compound	Year/Ref.	J_R [s⁻¹]	J_S [s⁻¹]
H$_2$O	α	1983[186]	6.4	6.4
		1987[187]	7.9	4.6
		1988[181]	7.9	4.6
		1995[292]	7.2	5.2
		2002[190]	8.3	4.0
		2004[184]	8.2	4.2
		2007[242]	8.2	4.2
	β	1974[282]	-	3.8
		1983[186]	7.8	-
		1987[187]	7.8	4.6
		1988[181]	7.8	4.6
		1995[292]	7.9	4.4
		2002[190]	7.6	4.4
		2004[184]	7.9	4.4
		2007[242]	7.9	4.4
	α-Me	1974[282]	-	4.6
		1976[288]	6.4	5.3
		1983[186]	8.2	4.6
		1984[180]	7.8	6.0
		1988[181]	8.3	4.0
		1987[187]	8.3	4.0
		2002[190]	8.3	4.0
	β-Me	1974[282]	-	4.4
		1976[288]	8.5	3.5
		1983[186]	7.6	4.4
		1987[187]	8.0	4.4
		1988[181]	8.0	4.4
		2002[190]	7.6	4.4
		2002[191]	8.0	4.3
DMSO	α-Me	1988[181]	6.3	6.2
	β-Me	1984[180]	6.3	-
		1988[181]	5.9	5.9
5%H$_2$O/DMSO	β-Me	1984[180]	6.6	-
CH$_3$OH	α-Me	2002[190]	6.2	5.7
	β-Me	2002[190]	6.6	5.5

Table 3.S.2: Experimentally measured J-coupling constants ($^3J_{HH}$) between the protons at C$_5$ (H5) and C$_6$ (H6R, H6S) for Glc and Gal. The values indicated as J_R and J_S correspond to $^3J_{H5,H6R}$ and $^3J_{H5,H6S}$, respectively. The anomery (α or β) and possible O$_1$-methylation (α-Me or β-Me) of the compound considered as well as the solvent (H$_2$O, DMSO, 5% H$_2$O/DMSO, CH$_3$OH or pyridine) are indicated.

	P [bar]	$\Delta V_{gt \to gg}$ [nm^3]	$\Delta V_{gt \to tg}$ [nm^3]
	1	-0.0014	0.0051
	$2 \cdot 10^3$	0.0001	0.0033
Glc	$5 \cdot 10^3$	-0.0013	0.0003
	$7 \cdot 10^3$	-0.0005	0.0001
	$10 \cdot 10^3$	-0.0010	0.0005
	1	0.0029	0.0026
	$2 \cdot 10^3$	0.0034	0.0025
Gal	$5 \cdot 10^3$	-0.0004	0.0004
	$7 \cdot 10^3$	0.0004	0.0008
	$10 \cdot 10^3$	-0.0006	-0.0004

Table 3.S.3: Partial molar volumes of the *gg* and *tg* rotamers of Glc and Gal relative to the *gt* rotamer at 298.15 K and at different pressures. The relative volumes $\Delta V_{gt \to gg}$ and $\Delta V_{gt \to tg}$ were calculated based on the box volumes in independent 100 ns NPT simulations of the compounds in SPC water (1200 molecules) at the indicated pressure, carried out with the hydroxymethyl group constrained to one of the three rotameric states. Numerical integration of $P\Delta V$ from 1 bar to 10 kbar suggests that the pressure increase induces free energy shifts $\Delta G_{gt \to gg}$ of -0.4 or 0.7 kJ·mol^{-1} and $\Delta G_{gt \to tg}$ of 0.9 or 0.7 kJ·mol^{-1} for Glc and Gal, respectively.

			53A6				56A6$_{CARBO}$			
	Series	Solvent	α_1 [°]	α_2 [°]	α_3 [°]	$^4C_1$ Pop. [%]	α_1 [°]	α_2 [°]	α_3 [°]	$^4C_1$ Pop. [%]
Glc	S_p^H	$W^{0.8}_{1.4}$	-35 (7)	-30 (8)	-35 (7)	91	-38 (7)	-36 (8)	-34 (7)	90
		$W^{0.9}_{1.2}$	-34 (11)	-29 (12)	-34 (11)	90	-38 (7)	-36 (8)	-34 (7)	89
		$W^{1.0}_{1.0}$	-35 (7)	-31 (8)	-35 (7)	92	-38 (7)	-36 (8)	-34 (7)	89
		$W^{1.5}_{0.5}$	-35 (7)	-32 (7)	-35 (7)	93	-38 (7)	-37 (8)	-34 (7)	89
	S_p^h	$W^{0.5}_{1.4}$	-35 (7)	-32 (7)	-35 (7)	93	-38 (7)	-37 (8)	-34 (7)	89
		$W^{0.8}_{0.9}$	-35 (7)	-32 (7)	-35 (7)	93	-37 (7)	-37 (8)	-34 (7)	88
		$W^{0.7}_{0.6}$	-34 (7)	-34 (7)	-34 (7)	95	-37 (7)	-39 (7)	-34 (7)	88
		$W^{0.5}_{0.3}$	-34 (7)	-35 (7)	-34 (7)	95	-37 (7)	-40 (7)	-34 (7)	87
	S_h^P	$W^{1.1}_{0.9}$	-35 (7)	-31 (7)	-35 (7)	92	-38 (7)	-36 (8)	-34 (7)	89
		$W^{1.0}_{1.0}$	-35 (7)	-31 (8)	-35 (7)	92	-38 (7)	-36 (8)	-34 (7)	89
		$W^{0.8}_{1.1}$	-35 (7)	-31 (8)	-35 (7)	97	-38 (7)	-37 (8)	-34 (7)	89
		$W^{0.7}_{1.2}$	-35 (7)	-31 (8)	-35 (7)	92	-38 (7)	-37 (8)	-34 (7)	88
		$W^{0.6}_{1.4}$	-35 (7)	-32 (8)	-35 (7)	92	-38 (7)	-37 (8)	-34 (7)	88
		$W^{0.5}_{1.5}$	-35 (7)	-32 (8)	-35 (7)	93	-38 (7)	-37 (8)	-34 (7)	89
	S_h^p	$W^{1.5}_{0.4}$	-35 (7)	-33 (7)	-34 (7)	94	-38 (7)	-38 (8)	-34 (7)	89
		$W^{1.2}_{0.5}$	-35 (7)	-33 (7)	-34 (7)	94	-37 (7)	-38 (7)	-34 (7)	89
		$W^{1.1}_{0.5}$	-35 (7)	-33 (7)	-34 (7)	94	-37 (7)	-38 (8)	-34 (7)	88
		$W^{0.8}_{0.7}$	-35 (7)	-33 (7)	-34 (7)	94	-37 (7)	-38 (7)	-34 (7)	88
		$W^{0.6}_{0.9}$	-35 (7)	-33 (7)	-34 (7)	94	-37 (7)	-38 (8)	-34 (7)	88
		$W^{0.4}_{1.3}$	-35 (7)	-33 (7)	-34 (7)	95	-37 (8)	-38 (7)	-34 (7)	89
	Series	Solvent	α_1 [°]	α_2 [°]	α_3 [°]	$^4C_1$ Pop. [%]	α_1 [°]	α_2 [°]	α_3 [°]	$^4C_1$ Pop. [%]
Deo	S_p^H	$W^{0.8}_{1.4}$					-36 (7)	-38 (7)	-36 (7)	91
		$W^{0.9}_{1.2}$					-36 (7)	-38 (7)	-36 (7)	90
		$W^{1.0}_{1.0}$					-36 (7)	-38 (7)	-36 (7)	90
		$W^{1.5}_{0.5}$					-36 (7)	-37 (7)	-36 (7)	89
	S_p^h	$W^{0.5}_{1.4}$					-36 (7)	-38 (7)	-36 (7)	90
		$W^{0.8}_{0.9}$					-36 (7)	-38 (7)	-36 (7)	89
		$W^{0.7}_{0.6}$					-37 (7)	-37 (7)	-34 (7)	89
		$W^{0.5}_{0.3}$					-38 (7)	-37 (7)	-33 (7)	88
	S_h^P	$W^{1.1}_{0.9}$					-36 (7)	-38 (7)	-36 (7)	90
		$W^{1.0}_{1.0}$					-36 (7)	-38 (7)	-36 (7)	89
		$W^{0.8}_{1.1}$					-36 (7)	-38 (7)	-36 (7)	89
		$W^{0.7}_{1.2}$					-36 (7)	-38 (7)	-36 (7)	89
		$W^{0.6}_{1.4}$					-36 (7)	-38 (7)	-36 (7)	90
		$W^{0.5}_{1.5}$					-36 (7)	-38 (7)	-36 (7)	89
	S_h^p	$W^{1.5}_{0.4}$					-36 (7)	-37 (7)	-36 (7)	89
		$W^{1.2}_{0.5}$					-37 (7)	-38 (7)	-35 (7)	89
		$W^{1.1}_{0.5}$					-37 (7)	-38 (7)	-35 (7)	89
		$W^{0.8}_{0.7}$					-37 (7)	-38 (7)	-35 (7)	89
		$W^{0.6}_{0.9}$					-37 (7)	-38 (7)	-35 (7)	89
		$W^{0.4}_{1.3}$					-37 (7)	-38 (7)	-35 (7)	89
	Series	Solvent	α_1 [°]	α_2 [°]	α_3 [°]	$^4C_1$ Pop. [%]	α_1 [°]	α_2 [°]	α_3 [°]	$^4C_1$ Pop. [%]
Gal	S_p^H	$W^{0.8}_{1.4}$	-31 (7)	-38 (6)	-37 (6)	93	-35 (7)	-38 (7)	-37 (6)	93
		$W^{0.9}_{1.2}$	-31 (7)	-38 (6)	-36 (6)	94	-36 (7)	-38 (7)	-36 (7)	92
		$W^{1.0}_{1.0}$	-31 (7)	-38 (6)	-37 (6)	94	-36 (7)	-38 (7)	-36 (7)	92
		$W^{1.5}_{0.5}$	-32 (7)	-38 (6)	-36 (6)	95	-36 (7)	-37 (7)	-36 (7)	93
	S_p^h	$W^{0.5}_{1.4}$	-32 (7)	-38 (6)	-36 (6)	94	-36 (7)	-38 (7)	-36 (7)	92
		$W^{0.8}_{0.9}$	-32 (7)	-38 (6)	-36 (6)	94	-36 (7)	-38 (7)	-36 (7)	92
		$W^{0.7}_{0.6}$	-33 (7)	-37 (6)	-35 (6)	95	-37 (7)	-37 (7)	-34 (7)	92
		$W^{0.5}_{0.3}$	-33 (6)	-36 (6)	-35 (6)	96	-38 (7)	-37 (7)	-33 (7)	91
	S_h^P	$W^{1.1}_{0.9}$	-31 (7)	-36 (6)	-36 (6)	94	-36 (7)	-38 (7)	-36 (7)	92
		$W^{1.0}_{1.0}$	-31 (7)	-38 (6)	-37 (6)	94	-36 (7)	-38 (7)	-36 (7)	92
		$W^{0.8}_{1.1}$	-31 (7)	-38 (6)	-36 (6)	94	-36 (7)	-38 (7)	-36 (7)	92
		$W^{0.7}_{1.2}$	-31 (7)	-38 (6)	-36 (6)	94	-36 (7)	-38 (7)	-36 (7)	92
		$W^{0.6}_{1.4}$	-31 (7)	-38 (6)	-36 (6)	94	-36 (7)	-38 (7)	-36 (7)	92
		$W^{0.5}_{1.5}$	-32 (7)	-38 (6)	-36 (6)	94	-36 (7)	-38 (7)	-36 (7)	92
	S_h^p	$W^{1.5}_{0.4}$	-32 (7)	-38 (6)	-36 (6)	95	-37 (7)	-37 (7)	-35 (7)	93
		$W^{1.2}_{0.5}$	-32 (7)	-38 (6)	-35 (6)	95	-37 (7)	-38 (7)	-35 (7)	92
		$W^{1.1}_{0.5}$	-32 (7)	-38 (6)	-35 (6)	95	-37 (7)	-38 (7)	-35 (7)	92
		$W^{0.8}_{0.7}$	-32 (7)	-38 (6)	-35 (6)	95	-37 (7)	-38 (7)	-35 (7)	92
		$W^{0.6}_{0.9}$	-32 (7)	-38 (6)	-35 (6)	95	-37 (7)	-38 (7)	-35 (7)	92
		$W^{0.4}_{1.3}$	-32 (7)	-38 (6)	-35 (6)	95	-37 (7)	-38 (7)	-34 (7)	96

Table 3.S.4: Ring-conformational properties of the three compound considered in the artificial solvents at 298.15 K and 998.3 kg·m^{-3}. The reported quantities are the Pickett & Strauss dihedral angles[271] α_1, α_2 and α_3, and the population of the $^4C_1$ chair conformation (defined by α_1, α_2 and α_3 simultaneously in the range between -17° and -50°). The values are calculated as reweighted averages over the 40 ns US phase of the different LEUS simulations. The corresponding standard deviations are reported between parentheses. The compounds considered are Glc, Deo and Gal (Figure 3.1). The solvents considered are artificial solvents in four series of decreasing permittivity or H-bonding capacity (Table 3.3). The force fields considered are the 53A6 force field[244] and the 56A6$_{CARBO}$ force field[245]. The simulations of Deo with 53A6 are omitted due to ambiguous torsional potential definitions. See Table 3.4 for corresponding results with the physical solvents.

	Solvent	ε	δH	53A6			$56A6_{CARBO}$		
				gg [%]	gt [%]	tg [%]	gg [%]	gt [%]	tg [%]
Glc	H_2O	61	42.3	56	44	0	36	59	4
	DMSO	46.7	10.2	47	53	0	26	64	10
	CH_3OH	33	22.3	46	54	0	27	59	14
	$CHCl_3$	4.8	5.7	40	35	24	6	15	79
	CCl_4	2.2	0	39	45	16	5	15	80
	VAC	1	0	46	45	9	6	13	80
Deo	H_2O	61	42.3				24	56	20
	DMSO	46.7	10.2				13	35	52
	CH_3OH	33	22.3				17	38	44
	$CHCl_3$	4.8	5.7				17	66	17
	CCl_4	2.2	0				13	70	17
	VAC	1	0				15	69	16
Gal	H_2O	61	42.3	34	41	25	7	53	40
	DMSO	46.7	10.2	10	33	56	5	27	68
	CH_3OH	33	22.3	23	36	41	8	35	57
	$CHCl_3$	4.8	5.7	91	4	5	45	33	21
	CCl_4	2.2	0	86	7	6	27	40	33
	VAC	1	0	77	16	7	36	36	28

Table 3.S.5: Populations of the three staggered hydroxymethyl rotamers for the three compounds considered in the physical solvents at 298.15 K and 1 bar. The values are calculated as reweighted averages over the 40 ns phase of the different LEUS simulations (or 1 μs plain SD simulation for VAC). The compounds considered are Glc, Deo and Gal, and the conformer populations refer to the canonical rotamers gg, gt and tg (Figure 3.1). The solvents considered are physical solvents of decreasing polarity (Table 3.2), with indicated values of the permittivity ϵ and Hansen parameter δH. The force fields considered are the 53A6 force field[244] and the $56A6_{CARBO}$ force field[245]. The simulations of Deo with 53A6 are omitted due to ambiguous torsional potential definitions. The data is illustrated graphically in Figure 3.4.

	Sol.	ε	n_H	53A6			56A6$_{\text{CARBO}}$		
				gg [%]	gt [%]	tg [%]	gg [%]	gt [%]	tg [%]
Glc	$W^{0.8}_{1.4}$	115	3.4	56	44	0	35	63	3
	$W^{0.9}_{1.2}$	89	3.5	56	44	0	38	59	4
	$W^{1.0}_{1.0}$	64	3.4	55	44	0	37	59	4
	$W^{1.5}_{0.5}$	25	3.4	55	45	0	35	61	5
	$W^{0.5}_{1.4}$	36	1.7	55	45	0	29	55	16
	$W^{0.8}_{0.9}$	33	2.0	55	44	0	30	56	14
	$W^{0.7}_{0.6}$	7	1.5	52	43	5	14	29	58
	$W^{0.5}_{0.3}$	1	1.4	42	38	20	5	13	82
Deo	$W^{0.8}_{1.4}$	113	3.4				21	58	21
	$W^{0.9}_{1.2}$	89	3.5				24	57	19
	$W^{1.0}_{1.0}$	64	3.4				24	57	19
	$W^{1.5}_{0.5}$	25	3.4				23	65	12
	$W^{0.5}_{1.4}$	36	1.7				20	55	25
	$W^{0.8}_{0.9}$	33	2.0				20	56	24
	$W^{0.7}_{0.6}$	7	1.5				20	56	24
	$W^{0.5}_{0.3}$	1	1.4				15	67	19
Gal	$W^{0.8}_{1.4}$	113	3.4	26	44	30	3	55	42
	$W^{0.9}_{1.2}$	89	3.5	31	43	26	4	55	41
	$W^{1.0}_{1.0}$	64	3.4	35	40	25	6	58	36
	$W^{1.5}_{0.5}$	25	3.4	48	38	14	10	67	23
	$W^{0.5}_{1.4}$	36	1.7	41	33	26	14	44	42
	$W^{0.8}_{0.9}$	33	2.0	43	34	23	15	47	38
	$W^{0.7}_{0.6}$	7	1.5	77	11	11	40	33	27
	$W^{0.5}_{0.3}$	1	1.4	87	7	7	33	36	31

Table 3.S.6: Populations of the three staggered hydroxymethyl rotamers for the three compounds considered in the series S_p^H and S_p^h of artificial solvents at 298.15 K and 998.3 kg·m^{-3}. The values are calculated as reweighted averages over the 40 ns phase of the different LEUS simulations. The compounds considered are Glc, Deo and Gal, and the conformer populations refer to the canonical rotamers gg, gt and tg (Figure 3.1). The solvents considered are artificial solvents of the series S_p^H and S_p^h (Table 3.3) in order of decreasing dielectric permittivity at water-like or lower H-bonding capacity, respectively, with indicated values of the solvent permittivity ϵ and number of H-bonds per molecule n_H. The force fields considered are the 53A6 force field[244] and the 56A6$_{\text{CARBO}}$ force field[245]. The simulations of Deo with 53A6 are omitted due to ambiguous torsional potential definitions. The data is illustrated graphically in Figure 3.5.

	Sol.	ε	n_H	53A6 gg [%]	gt [%]	tg [%]	56A6$_{\text{CARBO}}$ gg [%]	gt [%]	tg [%]
Glc	$W_{0.9}^{1.1}$	56	3.6	60	40	0	37	58	5
	$W_{1.0}^{1.0}$	64	3.4	55	44	0	37	59	4
	$W_{1.1}^{0.8}$	64	2.5	57	43	0	34	60	6
	$W_{1.2}^{0.7}$	65	2.2	54	46	0	33	61	6
	$W_{1.4}^{0.6}$	76	2.0	52	48	0	32	62	6
	$W_{1.5}^{0.5}$	49	1.7	53	47	0	32	58	10
Deo	$W_{0.9}^{1.1}$	56	3.6				24	60	16
	$W_{1.0}^{1.0}$	64	3.4				24	57	19
	$W_{1.1}^{0.8}$	64	2.5				21	56	23
	$W_{1.2}^{0.7}$	65	2.2				21	55	24
	$W_{1.4}^{0.6}$	76	2.0				20	56	24
	$W_{1.5}^{0.5}$	49	1.7				20	54	26
Gal	$W_{0.9}^{1.1}$	56	3.6	36	40	24	8	56	36
	$W_{1.0}^{1.0}$	64	3.4	35	40	25	6	58	36
	$W_{1.1}^{0.8}$	64	2.5	32	43	26	7	49	44
	$W_{1.2}^{0.7}$	65	2.2	32	41	27	7	50	43
	$W_{1.4}^{0.6}$	76	2.0	29	43	28	6	49	45
	$W_{1.5}^{0.5}$	49	1.7	33	39	28	10	45	45

Table 3.S.7: Populations of the three staggered hydroxymethyl rotamers for the three compounds considered in the series S_h^P of artificial solvents at 298.15 K and 998.3 kg·m^{-3}. The values are calculated as reweighted averages over the 40 ns phase of the different LEUS simulations. The compounds considered are Glc, Deo and Gal, and the conformer populations refer to the canonical rotamers gg, gt and tg (Figure 3.1). The solvents considered are artificial solvents of the series S_h^P (Table 3.3) in order of decreasing H-bonding capacity at water-like permittivity, with indicated values of the solvent permittivity ϵ and number of H-bonds per molecule n_H. The force fields considered are the 53A6 force field[244] and the 56A6$_{\text{CARBO}}$ force field[245]. The simulations of Deo with 53A6 are omitted due to ambiguous torsional potential definitions. The data is illustrated graphically in Figure 3.6.

	Sol.	ε	n_H	53A6			56A6$_\mathrm{CARBO}$		
				gg [%]	gt [%]	tg [%]	gg [%]	gt [%]	tg [%]
Glc	W$_{0.4}^{1.5}$	14	2.6	57	43	0	31	58	11
	W$_{0.5}^{1.2}$	15	2.2	56	44	1	28	50	22
	W$_{0.5}^{1.1}$	13	1.9	54	44	1	24	44	32
	W$_{0.7}^{0.8}$	15	1.7	54	45	1	22	43	35
	W$_{0.9}^{0.6}$	14	1.6	54	44	2	21	39	40
	W$_{1.3}^{0.4}$	12	1.5	53	45	2	19	36	45
Deo	W$_{0.4}^{1.5}$	14	2.6				23	63	14
	W$_{0.5}^{1.2}$	15	2.2				22	59	19
	W$_{0.5}^{1.1}$	13	1.9				22	58	20
	W$_{0.7}^{0.8}$	15	1.7				21	56	23
	W$_{0.9}^{0.6}$	14	1.6				20	55	25
	W$_{1.3}^{0.4}$	12	1.5				19	54	27
Gal	W$_{0.4}^{1.5}$	14	2.6	55	30	15	15	60	25
	W$_{0.5}^{1.2}$	15	2.2	58	25	16	24	46	30
	W$_{0.5}^{1.1}$	13	1.9	64	20	15	29	43	28
	W$_{0.7}^{0.8}$	15	1.7	61	21	18	30	39	31
	W$_{0.9}^{0.6}$	14	1.6	64	19	17	29	38	33
	W$_{1.3}^{0.4}$	12	1.5	65	18	17	34	34	32

Table 3.S.8: Populations of the three staggered hydroxymethyl rotamers for the three compounds considered in the series S_h^p of artificial solvents at 298.15 K and 998.3 kg·m^{-3}. The values are calculated as reweighted averages over the 40 ns phase of the different LEUS simulations. The compounds considered are Glc, Deo and Gal, and the conformer populations refer to the canonical rotamers gg, gt and tg (Figure 3.1). The solvents considered are artificial solvents of the series S_h^p (Table 3.3) in order of decreasing H-bonding capacity at low permittivity, with indicated values of the solvent permittivity ϵ and number of H-bonds per molecule n_H. The force fields considered are the 53A6 force field [244] and the 56A6$_\mathrm{CARBO}$ force field [245]. The simulations of Deo with 53A6 are omitted due to ambiguous torsional potential definitions. The data is illustrated graphically in Figure 3.7.

(a) Intramolecular H-bonds

Comp.	Sol.	53A6 — H₄→O₆				53A6 — H₆→O₄				53A6 — H₆→O₅				56A6ᴄᴀʀʙᴏ — H₄→O₆				56A6ᴄᴀʀʙᴏ — H₆→O₄				56A6ᴄᴀʀʙᴏ — H₆→O₅			
		tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg
Glc	H₂O	0.00	0.00	0.00	0.29	0.00	0.00	0.00	0.20	0.02	0.02	0.02	0.00	0.02	0.00	0.00	0.39	0.01	0.00	0.00	0.26	0.08	0.07	0.08	0.00
	DMSO	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.36	0.01	0.01	0.01	0.00	0.04	0.00	0.00	0.40	0.05	0.00	0.00	0.50	0.05	0.06	0.05	0.00
	CH₃OH	0.00	0.00	0.00	0.43	0.00	0.00	0.00	0.32	0.03	0.03	0.03	0.00	0.07	0.00	0.00	0.53	0.05	0.00	0.00	0.38	0.08	0.08	0.09	0.00
	CHCl₃	0.23	0.02	0.00	0.92	0.02	0.00	0.00	0.08	0.16	0.19	0.24	0.00	0.25	0.03	0.00	0.31	0.54	0.04	0.00	0.68	0.08	0.34	0.42	0.00
	CCl₄	0.13	0.02	0.00	0.79	0.03	0.00	0.00	0.21	0.20	0.24	0.23	0.00	0.10	0.02	0.00	0.12	0.70	0.04	0.00	0.87	0.08	0.36	0.42	0.00
	VAC	0.09	0.02	0.00	0.99	0.00	0.00	0.00	0.00	0.22	0.24	0.24	0.00	0.40	0.11	0.00	0.49	0.40	0.02	0.00	0.50	0.08	0.31	0.41	0.00
Deo	H₂O													-	-	-	-	-	-	-	-	0.06	0.08	0.07	0.00
	DMSO													-	-	-	-	-	-	-	-	0.02	0.05	0.03	0.00
	CH₃OH													-	-	-	-	-	-	-	-	0.04	0.08	0.07	0.00
	CHCl₃													-	-	-	-	-	-	-	-	0.34	0.45	0.41	0.00
	CCl₄													-	-	-	-	-	-	-	-	0.34	0.49	0.39	0.00
	VAC													-	-	-	-	-	-	-	-	0.34	0.49	0.39	0.00
Gal	H₂O	0.08	0.23	0.00	0.00	0.04	0.13	0.00	0.00	0.01	0.02	0.02	0.00	0.03	0.52	0.00	0.00	0.01	0.21	0.00	0.00	0.04	0.04	0.08	0.00
	DMSO	0.04	0.38	0.00	0.00	0.04	0.42	0.00	0.00	0.00	0.00	0.01	0.00	0.02	0.48	0.00	0.00	0.03	0.50	0.00	0.00	0.02	0.01	0.06	0.00
	CH₃OH	0.11	0.49	0.00	0.00	0.06	0.26	0.00	0.00	0.01	0.01	0.02	0.00	0.05	0.65	0.00	0.00	0.03	0.30	0.00	0.00	0.03	0.02	0.09	0.00
	CHCl₃	0.80	0.88	0.00	0.08	0.09	0.10	0.00	0.03	0.07	0.07	0.16	0.00	0.17	0.38	0.00	0.01	0.28	0.61	0.00	0.05	0.17	0.06	0.42	0.00
	CCl₄	0.76	0.88	0.00	0.03	0.09	0.10	0.00	0.08	0.12	0.13	0.15	0.00	0.11	0.42	0.00	0.00	0.18	0.57	0.00	0.06	0.18	0.10	0.39	0.00
	VAC	0.69	0.89	0.00	0.04	0.07	0.08	0.00	0.06	0.13	0.13	0.16	0.00	0.15	0.42	0.01	0.00	0.22	0.56	0.00	0.07	0.18	0.10	0.39	0.01

(b) Solute-solvent H-bonds

Comp.	Sol.	53A6 — O₄↔S				53A6 — O₆↔S				53A6 — S→O₅				56A6ᴄᴀʀʙᴏ — O₄↔S				56A6ᴄᴀʀʙᴏ — O₆↔S				56A6ᴄᴀʀʙᴏ — S→O₅			
		tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg
Glc	H₂O	1.70	1.71	1.69	1.69	2.17	2.16	2.19	2.15	0.68	0.67	0.68	0.70	1.74	1.78	1.76	1.28	2.21	2.18	2.27	1.71	0.44	0.45	0.43	0.47
	DMSO	0.82	0.83	0.82	0.55	0.98	0.98	0.97	0.65	-	-	-	-	0.71	0.75	0.74	0.44	0.89	0.93	0.93	0.49	-	-	-	-
	CH₃OH	1.09	1.13	1.06	0.68	1.74	1.73	1.75	1.13	0.39	0.38	0.40	0.38	1.05	1.17	1.10	0.56	1.69	1.78	1.80	1.00	0.13	0.14	0.13	0.13
	CHCl₃	0.04	0.04	0.04	0.03	0.08	0.09	0.11	0.02	0.04	0.04	0.03	0.06	0.03	0.06	0.06	0.02	0.11	0.11	0.13	0.10	0.02	0.02	0.01	0.02
Deo	H₂O													-	-	-	-	2.24	2.24	2.26	2.19	0.51	0.48	0.50	0.56
	DMSO													-	-	-	-	0.95	0.94	0.94	0.96	-	-	-	-
	CH₃OH													-	-	-	-	1.71	1.77	1.72	1.67	0.16	0.15	0.15	0.17
	CHCl₃													-	-	-	-	0.11	0.12	0.12	0.07	0.02	0.01	0.02	0.03
Gal	H₂O	1.70	1.71	1.69	1.69	2.17	2.16	2.19	2.15	0.68	0.67	0.68	0.70	1.74	1.78	1.76	1.28	2.21	2.18	2.27	1.71	0.44	0.45	0.43	0.47
	DMSO	0.82	0.83	0.82	0.55	0.98	0.98	0.97	0.65	-	-	-	-	0.71	0.75	0.74	0.44	0.89	0.93	0.93	0.49	-	-	-	-
	CH₃OH	1.09	1.13	1.06	0.68	1.74	1.73	1.75	1.13	0.39	0.38	0.40	0.38	1.05	1.17	1.10	0.56	1.69	1.78	1.80	1.00	0.13	0.14	0.13	0.13
	CHCl₃	0.04	0.04	0.04	0.03	0.08	0.09	0.11	0.02	0.04	0.04	0.03	0.06	0.03	0.06	0.06	0.02	0.11	0.11	0.13	0.10	0.02	0.02	0.01	0.02

Table 3.S.9: Detailed H-bond analysis for the three compounds in the series of physical solvents at 298.15 K and 1 bar. The values are calculated as averages over the 40 ns phase of the different LEUS simulations (1 μs plain SD simulation for VAC). The compounds considered are Glc, Deo, and Gal (Figure 3.1), along the series of physical solvents (Table 3.2) in order of decreasing polarity. The force fields considered are the 53A6 force field[244] and the 56A6$_{\text{CARBO}}$ force field[245]. In a, the average numbers of intramolecular H$_4\rightarrow$O$_6$, H$_6\rightarrow$O$_4$ and H$_6\rightarrow$O$_5$ H-bonds are reported. In b, the average numbers of solute-solvent O$_4\leftrightarrow$S, O$_6\leftrightarrow$S and S$\rightarrow$O$_5$ H-bonds are reported (CCl$_4$ and VAC are omitted as they are non-H-bonding). The average numbers are calculated either over the entire simulation (N) or over the three canonical rotamers separately, *i.e.* as N_{gg}, N_{gt} and N_{tg}. The data is illustrated graphically in Suppl. Mat. Figures 3.S.4 and 3.S.8.

Table 3.S.10: Detailed H-bond analysis for the three compounds in the series S_p^H and S_p^h of artificial solvents at 298.15 K and 998.3 kg·m^{-3}. The values are calculated as averages over the 40 ns phase of the different LEUS simulations. The compounds considered are Glc, Deo and Gal (Figure 3.1), along the series S_p^H and S_p^h of artificial solvents (Table 3.3) in order of decreasing dielectric permittivity at water-like or lower H-bonding capacity, respectively. The force fields considered are the 53A6 force field[244] (left) and 56A6$_{\text{CARBO}}$ force field[245] (right). In a, the average numbers of intramolecular H4→O6, H6→O4 and H6→O5 H-bonds are reported. In b, the average numbers of solute-solvent O4↔S, O6↔S and S→O5 H-bonds are reported, where S is a solvent donor or acceptor site. The average numbers are calculated either over the entire simulation (N) or over the three canonical rotamers separately, i.e. as N_{gg}, N_{gt} and N_{tg}. The data is illustrated graphically in Suppl. Mat. Figures 3.S.5 and 3.S.9.

Comp.	Sol.	53A6 H4→O6 tot	gg	gt	tg	53A6 H6→O4 tot	gg	gt	tg	53A6 H6→O5 tot	gg	gt	tg	56A6 H4→O6 tot	gg	gt	tg	56A6 H6→O4 tot	gg	gt	tg	56A6 H6→O5 tot	gg	gt	tg
Glc	$W_{1.4}$	0.00	0.00	0.00	0.15	0.00	0.00	0.00	0.06	0.01	0.01	0.01	0.00	0.01	0.00	0.00	0.25	0.00	0.00	0.00	0.11	0.03	0.03	0.03	0.00
	$W_{1.2}$	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.07	0.02	0.02	0.02	0.00	0.02	0.00	0.00	0.37	0.01	0.00	0.00	0.15	0.04	0.04	0.04	0.00
	$W_{1.0}$	0.00	0.00	0.00	0.26	0.00	0.00	0.00	0.20	0.03	0.03	0.03	0.00	0.02	0.00	0.00	0.42	0.01	0.00	0.00	0.24	0.07	0.07	0.08	0.00
	$W_{0.9}$	0.00	0.00	0.00	0.49	0.00	0.00	0.00	0.24	0.10	0.09	0.10	0.00	0.02	0.01	0.00	0.43	0.02	0.00	0.00	0.34	0.22	0.21	0.24	0.00
	$W_{0.8}$	0.00	0.00	0.00	0.56	0.00	0.00	0.00	0.29	0.05	0.05	0.05	0.00	0.09	0.01	0.00	0.53	0.06	0.00	0.00	0.39	0.11	0.13	0.14	0.00
	$W_{0.7}$	0.00	0.00	0.00	0.51	0.00	0.00	0.00	0.32	0.06	0.06	0.06	0.00	0.08	0.01	0.00	0.55	0.05	0.00	0.00	0.36	0.13	0.13	0.16	0.00
	$W_{0.6}$	0.05	0.01	0.00	0.86	0.01	0.00	0.00	0.13	0.16	0.15	0.19	0.00	0.25	0.02	0.00	0.43	0.33	0.02	0.00	0.56	0.14	0.27	0.35	0.00
	$W_{0.3}$	0.19	0.02	0.00	0.87	0.03	0.00	0.00	0.13	0.19	0.22	0.25	0.00	0.18	0.03	0.00	0.22	0.64	0.04	0.00	0.78	0.07	0.35	0.43	0.00
Deo	$W_{1.4}$									0.00	0.01	0.01	0.00	–	–	–	–	–	–	–	–	0.02	0.03	0.02	0.00
	$W_{1.2}$									0.01	0.01	0.02	0.00	–	–	–	–	–	–	–	–	0.03	0.04	0.04	0.00
	$W_{1.0}$									0.01	0.02	0.02	0.00	–	–	–	–	–	–	–	–	0.06	0.09	0.06	0.00
	$W_{0.9}$									0.05	0.05	0.07	0.00	–	–	–	–	–	–	–	–	0.19	0.24	0.21	0.00
	$W_{0.8}$									0.02	0.01	0.03	0.00	–	–	–	–	–	–	–	–	0.09	0.13	0.11	0.00
	$W_{0.7}$									0.02	0.02	0.03	0.00	–	–	–	–	–	–	–	–	0.10	0.14	0.12	0.00
	$W_{0.6}$									0.04	0.03	0.13	0.00	–	–	–	–	–	–	–	–	0.25	0.35	0.31	0.00
	$W_{0.3}$									0.10	0.10	0.18	0.00	–	–	–	–	–	–	–	–	0.34	0.47	0.40	0.00
Gal	$W_{1.4}$	0.03	0.11	0.00	0.00	0.01	0.04	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.09	0.00	0.00	0.01	0.44	0.00	0.00	0.02	0.03	0.03	0.00
	$W_{1.2}$	0.06	0.18	0.00	0.00	0.02	0.06	0.00	0.00	0.01	0.01	0.01	0.00	0.01	0.16	0.00	0.00	0.02	0.47	0.00	0.00	0.03	0.03	0.04	0.00
	$W_{1.0}$	0.09	0.25	0.00	0.00	0.04	0.10	0.00	0.00	0.01	0.02	0.02	0.00	0.01	0.20	0.00	0.00	0.03	0.51	0.00	0.00	0.05	0.04	0.08	0.00
	$W_{0.9}$	0.14	0.29	0.00	0.00	0.10	0.21	0.00	0.00	0.05	0.05	0.07	0.00	0.03	0.29	0.00	0.00	0.04	0.43	0.00	0.00	0.17	0.13	0.23	0.00
	$W_{0.8}$	0.18	0.45	0.00	0.00	0.11	0.28	0.00	0.00	0.02	0.01	0.03	0.00	0.05	0.35	0.00	0.00	0.05	0.58	0.00	0.00	0.06	0.03	0.13	0.00
	$W_{0.7}$	0.19	0.44	0.00	0.00	0.11	0.25	0.00	0.00	0.02	0.02	0.03	0.00	0.05	0.35	0.00	0.00	0.05	0.56	0.00	0.00	0.08	0.03	0.15	0.00
	$W_{0.6}$	0.57	0.74	0.00	0.03	0.16	0.21	0.00	0.01	0.04	0.03	0.13	0.00	0.20	0.49	0.00	0.01	0.20	0.49	0.00	0.01	0.13	0.04	0.36	0.00
	$W_{0.3}$	0.76	0.87	0.00	0.05	0.10	0.11	0.00	0.05	0.10	0.10	0.18	0.00	0.21	0.60	0.00	0.05	0.21	0.38	0.00	0.00	0.17	0.07	0.40	0.00

(a) Intramolecular H-bonds

Comp.	Sol.	53A6												56A6$_{CARBO}$												
		O$_4$↔S				O$_6$↔S				S→O$_5$				O$_4$↔S				O$_6$↔S				S→O$_5$			
		tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg
Glc	$W^{0.8}$	1.74	1.75	1.72	1.60	2.11	2.10	2.12	1.91	0.67	0.67	0.67	0.66	1.81	1.83	1.81	1.52	2.20	2.17	2.24	1.87	0.45	0.46	0.45	0.44
	$W^{0.9}$	1.73	1.74	1.71	1.65	2.15	2.14	2.17	1.92	0.68	0.67	0.69	0.67	1.78	1.78	1.80	1.36	2.21	2.17	2.26	1.77	0.45	0.46	0.45	0.46
	$W^{1.2}$	1.72	1.74	1.70	1.41	2.18	2.15	2.22	1.80	0.68	0.69	0.68	0.69	1.75	1.75	1.76	1.29	2.22	2.19	2.28	1.74	0.45	0.45	0.45	0.48
	$W^{1.0}$	1.75	1.76	1.74	1.48	2.38	2.32	2.45	1.94	0.75	0.75	0.74	0.81	1.76	1.76	1.76	1.49	2.37	2.30	2.44	1.99	0.50	0.52	0.48	0.63
	$W^{0.5}$	1.35	1.36	1.35	0.91	1.75	1.72	1.78	1.32	0.45	0.46	0.45	0.46	1.27	1.36	1.35	0.86	1.67	1.72	1.77	1.24	0.34	0.34	0.33	0.36
	$W^{0.4}$	1.48	1.50	1.47	1.01	1.97	1.93	2.01	1.43	0.58	0.58	0.57	0.59	1.43	1.53	1.50	0.95	1.91	1.93	2.04	1.39	0.41	0.42	0.40	0.44
	$W^{0.7}$	1.07	1.11	1.07	0.63	1.49	1.48	1.55	1.06	0.39	0.38	0.38	0.45	0.86	1.09	1.06	0.70	1.24	1.44	1.52	1.06	0.30	0.27	0.25	0.32
	$W^{0.3}$	0.69	0.74	0.75	0.46	0.95	0.96	1.02	0.80	0.22	0.21	0.21	0.25	0.60	0.74	0.72	0.57	0.74	0.93	0.97	0.69	0.20	0.19	0.17	0.21
Deo	$W^{0.8}$													-	-	-	-	2.16	2.16	2.18	2.09	0.44	0.42	0.44	0.46
	$W^{0.9}$													-	-	-	-	2.22	2.22	2.25	2.16	0.49	0.47	0.49	0.52
	$W^{1.2}$													-	-	-	-	2.25	2.24	2.28	2.18	0.52	0.49	0.51	0.57
	$W^{1.5}$													-	-	-	-	2.44	2.42	2.46	2.40	0.63	0.60	0.61	0.79
	$W^{0.5}$													-	-	-	-	1.25	1.32	1.26	1.24	0.09	0.09	0.09	0.10
	$W^{0.4}$													-	-	-	-	1.89	1.89	1.92	1.83	0.37	0.35	0.36	0.40
	$W^{0.7}$													-	-	-	-	1.23	1.21	1.24	1.23	0.18	0.18	0.17	0.21
	$W^{0.5}$													-	-	-	-	0.54	0.52	0.53	0.60	0.03	0.03	0.03	0.03
Gal	$W^{0.8}$	1.80	1.71	1.81	1.87	2.03	1.99	2.11	1.97	0.72	0.69	0.73	0.72	1.87	1.41	1.86	1.90	2.14	1.79	2.22	2.05	0.51	0.51	0.50	0.52
	$W^{0.9}$	1.78	1.66	1.82	1.87	2.06	1.96	2.16	2.01	0.73	0.72	0.74	0.75	1.85	1.34	1.85	1.89	2.15	1.74	2.24	2.06	0.51	0.52	0.49	0.54
	$W^{1.0}$	1.75	1.59	1.81	1.87	2.04	1.92	2.18	2.01	0.74	0.73	0.74	0.77	1.82	1.31	1.84	1.87	2.17	1.70	2.27	2.09	0.52	0.53	0.49	0.56
	$W^{1.5}$	1.73	1.66	1.79	1.82	2.23	2.12	2.39	2.16	0.83	0.82	0.81	0.89	1.76	1.55	1.78	1.80	2.35	2.02	2.44	2.22	0.62	0.61	0.58	0.75
	$W^{0.5}$	1.30	1.07	1.45	1.48	1.59	1.41	1.77	1.63	0.51	0.52	0.50	0.52	1.37	0.92	1.43	1.46	1.65	1.29	1.78	1.63	0.41	0.43	0.39	0.43
	$W^{0.8}$	1.43	1.20	1.58	1.65	1.76	1.57	1.99	1.79	0.66	0.66	0.65	0.68	1.51	1.02	1.58	1.63	1.87	1.45	2.03	1.83	0.50	0.52	0.46	0.53
	$W^{0.7}$	0.80	0.69	1.15	1.17	1.23	1.16	1.54	1.37	0.48	0.48	0.44	0.49	0.99	0.74	1.13	1.17	1.33	1.13	1.51	1.41	0.36	0.39	0.30	0.39
	$W^{0.3}$	0.50	0.46	0.76	0.77	0.87	0.86	1.01	0.90	0.25	0.25	0.23	0.24	0.69	0.56	0.75	0.77	0.90	0.75	0.99	0.95	0.22	0.23	0.19	0.23

(b) Solute-solvent H-bonds.

Table 3.S.11: Detailed H-bonds analysis for the three compounds in the series S_h^P of artificial solvents at 298.15 K and 998.3 kg·m^{-3}. The values are calculated as averages over the 40 ns phase of the different LEUS simulations. The compounds considered are Glc, Deo and Gal (Figure 3.1), along the series S_h^P of artificial solvents (Table 3.3) in order of decreasing H-bonding capacity at water-like dielectric permittivity. The force fields considered are the 53A6 force field[244] and 56A6$_{CARBO}$ force field[245]. In a, the average numbers of intramolecular $H_4\to O_6$, $H_6\to O_4$ and $H_6\to O_5$ H-bonds are reported. In b, the average numbers of solute-solvent $O_4\leftrightarrow S$, $O_6\leftrightarrow S$ and $S\to O_5$ H-bonds are reported. The average numbers are calculated either over the entire simulation (N) or over the three canonical rotamers separately, i.e. as N_{gg}, N_{gt} and N_{tg}. The data is illustrated graphically in Suppl. Mat. Figures 3.S.6 and 3.S.10.

| | | 53A6 | | | | | | | | | | | | 56A6$_{CARBO}$ | | | | | | | | | | | |
| | | $H_4\to O_6$ | | | | $H_6\to O_4$ | | | | $H_6\to O_5$ | | | | $H_4\to O_6$ | | | | $H_6\to O_4$ | | | | $H_6\to O_5$ | | | |
Comp.	Sol.	tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg
Glc	$W^{1.1}$	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.24	0.03	0.03	0.03	0.00	0.01	0.00	0.00	0.27	0.02	0.00	0.00	0.40	0.09	0.09	0.10	0.00
	$W^{1.0}$	0.00	0.00	0.00	0.26	0.00	0.00	0.00	0.20	0.03	0.03	0.03	0.00	0.01	0.00	0.00	0.24	0.02	0.00	0.00	0.42	0.07	0.07	0.08	0.00
	$W^{0.8}$	0.00	0.00	0.00	0.35	0.00	0.00	0.00	0.21	0.03	0.03	0.03	0.00	0.02	0.00	0.00	0.29	0.03	0.00	0.00	0.43	0.08	0.08	0.09	0.00
	$W^{0.7}_{0.2}$	0.00	0.00	0.00	0.36	0.00	0.00	0.00	0.22	0.03	0.03	0.04	0.00	0.02	0.00	0.00	0.31	0.03	0.00	0.00	0.45	0.08	0.08	0.09	0.00
	$W^{0.6}_{1.4}$	0.00	0.00	0.00	0.32	0.00	0.00	0.00	0.25	0.03	0.03	0.03	0.00	0.02	0.00	0.00	0.27	0.03	0.00	0.00	0.47	0.07	0.07	0.08	0.00
	$W^{0.5}_{1.5}$	0.00	0.00	0.00	0.46	0.00	0.00	0.00	0.30	0.04	0.04	0.04	0.00	0.04	0.00	0.00	0.36	0.05	0.01	0.00	0.52	0.10	0.10	0.12	0.00
Deo	$W^{1.1}$	–	–	–	–	–	–	–	–	0.02	0.02	0.02	0.00	–	–	–	–	–	–	–	–	0.08	0.10	0.08	0.00
	$W^{1.0}$	–	–	–	–	–	–	–	–	0.01	0.02	0.02	0.00	–	–	–	–	–	–	–	–	0.06	0.09	0.06	0.00
	$W^{0.8}$	–	–	–	–	–	–	–	–	0.01	0.01	0.02	0.00	–	–	–	–	–	–	–	–	0.05	0.08	0.06	0.00
	$W^{0.7}_{0.2}$	–	–	–	–	–	–	–	–	0.01	0.02	0.02	0.00	–	–	–	–	–	–	–	–	0.06	0.08	0.07	0.00
	$W^{0.6}_{1.4}$	–	–	–	–	–	–	–	–	0.01	0.01	0.02	0.00	–	–	–	–	–	–	–	–	0.05	0.08	0.06	0.00
	$W^{0.5}_{1.5}$	–	–	–	–	–	–	–	–	0.01	0.01	0.03	0.00	–	–	–	–	–	–	–	–	0.07	0.11	0.09	0.00
Gal	$W^{1.1}$	0.09	0.26	0.00	0.00	0.05	0.15	0.00	0.00	0.02	0.02	0.02	0.00	0.02	0.25	0.00	0.00	0.04	0.47	0.00	0.00	0.06	0.06	0.10	0.00
	$W^{1.0}$	0.09	0.25	0.00	0.00	0.04	0.10	0.00	0.00	0.01	0.02	0.02	0.00	0.01	0.20	0.00	0.00	0.03	0.51	0.00	0.00	0.05	0.04	0.08	0.00
	$W^{0.8}$	0.07	0.23	0.00	0.00	0.06	0.17	0.00	0.00	0.01	0.01	0.02	0.00	0.02	0.26	0.00	0.00	0.04	0.54	0.00	0.00	0.04	0.03	0.08	0.00
	$W^{0.7}_{0.2}$	0.09	0.28	0.00	0.00	0.06	0.18	0.00	0.00	0.01	0.02	0.02	0.00	0.02	0.29	0.00	0.00	0.04	0.53	0.00	0.00	0.04	0.03	0.08	0.00
	$W^{0.6}_{1.4}$	0.09	0.30	0.00	0.00	0.04	0.14	0.00	0.00	0.01	0.01	0.02	0.00	0.02	0.27	0.00	0.00	0.04	0.55	0.00	0.00	0.04	0.03	0.08	0.00
	$W^{0.5}_{1.5}$	0.11	0.35	0.00	0.00	0.08	0.24	0.00	0.00	0.01	0.01	0.03	0.00	0.03	0.31	0.00	0.00	0.06	0.60	0.00	0.00	0.06	0.02	0.12	0.00

(a) Intramolecular H-bonds

Comp.	Sol.	53A6												56A6$_{\text{CARBO}}$											
		O4↔S				O6↔S				S→O5				O4↔S				O6↔S				S→O5			
		tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg
Glc	$W^{1.1}_{0.9}$	1.73	1.73	1.72	1.36	2.17	2.15	2.20	1.74	0.68	0.68	0.69	0.73	1.76	1.77	1.79	1.29	2.20	2.17	2.27	1.72	0.43	0.44	0.41	0.49
	$W^{1.0}_{0.8}$	1.72	1.74	1.70	1.41	2.18	2.15	2.22	1.80	0.68	0.69	0.68	0.69	1.75	1.78	1.76	1.29	2.22	2.19	2.28	1.74	0.45	0.45	0.45	0.48
	$W^{1.0}_{0.7}$	1.63	1.64	1.61	1.30	2.05	2.05	2.06	1.67	0.63	0.63	0.62	0.63	1.63	1.68	1.65	1.17	2.09	2.05	2.16	1.55	0.44	0.46	0.44	0.44
	$W^{1.2}_{0.6}$	1.56	1.57	1.54	1.23	1.97	1.95	1.99	1.61	0.58	0.58	0.58	0.55	1.55	1.60	1.57	1.12	1.99	1.96	2.05	1.49	0.41	0.42	0.40	0.43
	$W^{1.4}_{0.5}$	1.52	1.53	1.52	1.22	1.90	1.88	1.92	1.56	0.53	0.53	0.53	0.52	1.51	1.56	1.53	1.08	1.91	1.88	1.96	1.47	0.39	0.40	0.39	0.40
	$W^{1.5}_{0.5}$	1.41	1.43	1.40	1.00	1.79	1.77	1.81	1.39	0.47	0.46	0.47	0.46	1.36	1.43	1.40	0.91	1.76	1.76	1.83	1.30	0.35	0.35	0.34	0.36
Deo	$W^{1.1}_{0.9}$	-	-	-	-									-	-	-	-	2.26	2.24	2.28	2.18	0.53	0.51	0.52	0.60
	$W^{1.0}_{0.8}$	-	-	-	-									-	-	-	-	2.25	2.24	2.28	2.18	0.52	0.49	0.50	0.57
	$W^{1.0}_{0.7}$	-	-	-	-									-	-	-	-	2.03	2.02	2.06	1.96	0.40	0.39	0.28	0.42
	$W^{1.2}_{0.6}$	-	-	-	-									-	-	-	-	1.80	1.81	1.51	1.48	0.31	0.30	0.16	0.32
	$W^{1.4}_{0.5}$	-	-	-	-									-	-	-	-	1.60	1.60	1.33	1.28	0.20	0.20	0.09	0.20
	$W^{1.5}_{0.5}$	-	-	-	-									-	-	-	-	1.29	1.30	1.29	1.22	0.09	0.09	0.04	0.10
Gal	$W^{1.1}_{0.9}$	1.74	1.56	1.82	1.87	2.03	1.88	2.17	2.02	0.74	0.74	0.72	0.78	1.80	1.33	1.84	1.86	2.14	1.71	2.24	2.08	0.52	0.51	0.48	0.58
	$W^{1.0}_{0.8}$	1.75	1.59	1.81	1.87	2.04	1.92	2.18	2.01	0.74	0.73	0.74	0.77	1.82	1.31	1.84	1.87	2.17	1.70	2.27	2.09	0.52	0.53	0.49	0.56
	$W^{1.0}_{0.7}$	1.66	1.49	1.70	1.77	1.94	1.82	2.06	1.89	0.69	0.69	0.69	0.70	1.71	1.18	1.73	1.76	2.01	1.56	2.13	1.95	0.53	0.53	0.52	0.54
	$W^{1.2}_{0.6}$	1.57	1.40	1.64	1.67	1.85	1.72	1.97	1.81	0.64	0.64	0.65	0.63	1.63	1.11	1.65	1.69	1.93	1.49	2.04	1.87	0.50	0.49	0.48	0.52
	$W^{1.4}_{0.5}$	1.54	1.37	1.59	1.64	1.80	1.69	1.91	1.76	0.59	0.60	0.60	0.59	1.59	1.06	1.60	1.65	1.85	1.47	1.95	1.80	0.47	0.46	0.46	0.48
	$W^{1.5}_{0.5}$	1.40	1.19	1.48	1.53	1.66	1.50	1.79	1.66	0.53	0.53	0.53	0.51	1.43	0.93	1.46	1.51	1.70	1.32	1.81	1.68	0.42	0.43	0.40	0.44

(b) Solute-solvent H-bonds.

Table 3.S.12: Detailed H-bonds analysis for the three compounds in the series S_h^p of artificial solvents at 298.15 K and 998.3 kg·m^{-3}. The values are calculated as averages over the 40 ns phase of the different LEUS simulations. The compounds considered are Glc, Deo and Gal (Figure 3.1), along the series S_h^p of artificial solvents (Table 3.3) in order of decreasing H-bonding capacity at low dielectric permittivity. The force fields considered are the 53A6 force field[244] (left) and 56A6CARBO force field[245] (right). In a, the average numbers of intramolecular H$_4$→O$_6$, H$_6$→O$_4$ and H$_6$→O$_5$ H-bonds are reported, where S is a solvent donor or acceptor site. In b, the average numbers of solute-solvent O$_4$↔S, O$_6$↔S and S→O$_5$ H-bonds are reported. The average numbers are calculated either over the entire simulation (N) or over the three canonical rotamers separately, *i.e.* as N_{gg}, N_{gt} and N_{tg}. The data is illustrated graphically in Suppl. Mat. Figures 3.S.7 and 3.S.11.

(a) Intramolecular H-bonds

Comp.	Sol.	53A6												56A6CARBO											
		H$_4$→O$_6$				H$_6$→O$_4$				H$_6$→O$_5$				H$_4$→O$_6$				H$_6$→O$_4$				H$_6$→O$_5$			
		tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg	tot	gg	gt	tg
Glc	W1.5	0.00	0.00	0.00	0.64	0.00	0.00	0.00	0.24	0.13	0.12	0.14	0.00	0.05	0.01	0.00	0.49	0.04	0.00	0.00	0.41	0.23	0.24	0.26	0.00
	W1.2	0.01	0.00	0.00	0.68	0.00	0.00	0.00	0.24	0.11	0.11	0.12	0.00	0.13	0.01	0.00	0.56	0.09	0.01	0.00	0.39	0.19	0.21	0.26	0.00
	W1.1	0.01	0.00	0.00	0.75	0.00	0.00	0.00	0.21	0.12	0.12	0.14	0.00	0.17	0.01	0.00	0.52	0.15	0.01	0.00	0.45	0.17	0.22	0.28	0.00
	W0.8	0.01	0.00	0.00	0.73	0.00	0.00	0.00	0.23	0.11	0.11	0.12	0.00	0.18	0.01	0.00	0.49	0.17	0.01	0.00	0.48	0.15	0.20	0.25	0.00
	W0.7	0.01	0.00	0.00	0.77	0.00	0.00	0.00	0.20	0.11	0.10	0.12	0.00	0.22	0.02	0.00	0.55	0.17	0.01	0.00	0.43	0.14	0.20	0.25	0.00
	W0.6	0.02	0.00	0.00	0.80	0.00	0.00	0.00	0.18	0.11	0.11	0.13	0.00	0.23	0.02	0.00	0.52	0.21	0.01	0.00	0.46	0.14	0.21	0.27	0.00
Deo	W1.5													-	-	-	-	-	-	-	-	0.22	0.29	0.25	0.00
	W1.2													-	-	-	-	-	-	-	-	0.18	0.25	0.22	0.00
	W1.1													-	-	-	-	-	-	-	-	0.20	0.37	0.24	0.00
	W0.8													-	-	-	-	-	-	-	-	0.17	0.24	0.22	0.00
	W0.7													-	-	-	-	-	-	-	-	0.18	0.25	0.23	0.00
	W0.6													-	-	-	-	-	-	-	-	0.18	0.26	0.24	0.00
Gal	W1.5	0.27	0.49	0.00	0.01	0.11	0.20	0.00	0.01	0.05	0.05	0.09	0.00	0.07	0.48	0.00	0.01	0.06	0.37	0.00	0.01	0.18	0.09	0.28	0.00
	W1.2	0.31	0.53	0.00	0.01	0.16	0.27	0.00	0.01	0.04	0.03	0.08	0.00	0.13	0.53	0.00	0.01	0.09	0.40	0.00	0.00	0.14	0.05	0.27	0.00
	W1.1	0.38	0.58	0.00	0.01	0.19	0.29	0.00	0.01	0.03	0.02	0.09	0.00	0.16	0.53	0.00	0.01	0.13	0.43	0.00	0.01	0.13	0.04	0.28	0.00
	W0.8	0.34	0.55	0.00	0.01	0.20	0.32	0.00	0.01	0.03	0.02	0.07	0.00	0.16	0.53	0.00	0.01	0.13	0.43	0.00	0.01	0.12	0.03	0.27	0.00
	W0.7	0.37	0.58	0.00	0.01	0.20	0.31	0.00	0.01	0.03	0.02	0.08	0.00	0.15	0.52	0.00	0.01	0.14	0.46	0.00	0.01	0.11	0.03	0.26	0.00
	W0.6	0.41	0.63	0.00	0.01	0.19	0.28	0.00	0.01	0.03	0.02	0.08	0.00	0.18	0.54	0.00	0.01	0.15	0.44	0.00	0.01	0.11	0.03	0.29	0.00

| Comp.	Sol.	53A6 $O_4{\leftrightarrow}S$ tot	gg	gt	tg	53A6 $O_6{\leftrightarrow}S$ tot	gg	gt	tg	53A6 $S{\to}O_5$ tot	gg	gt	tg	56A6$_{CARBO}$ $O_4{\leftrightarrow}S$ tot	gg	gt	tg	56A6$_{CARBO}$ $O_6{\leftrightarrow}S$ tot	gg	gt	tg	56A6$_{CARBO}$ $S{\to}O_5$ tot	gg	gt	tg
Glc	$W^{1.5}$	1.85	1.85	1.84	1.43	2.43	2.40	2.49	1.84	0.82	0.83	0.80	0.85	1.85	1.90	1.90	1.44	2.41	2.39	2.51	1.91	0.54	0.56	0.52	0.65
	$W^{1.2}_{0.4}$	1.58	1.59	1.58	1.04	2.17	2.14	2.21	1.48	0.69	0.70	0.68	0.75	1.50	1.63	1.62	1.05	2.06	2.13	2.24	1.54	0.47	0.48	0.44	0.54
	$W^{1.1}_{0.5}$	1.44	1.46	1.44	0.91	2.00	1.97	2.06	1.36	0.63	0.63	0.63	0.68	1.31	1.49	1.49	0.94	1.85	1.97	2.08	1.44	0.43	0.43	0.40	0.49
	$W^{0.8}_{0.5}$	1.30	1.32	1.30	0.78	1.78	1.76	1.82	1.25	0.50	0.51	0.50	0.54	1.13	1.33	1.30	0.82	1.61	1.73	1.84	1.25	0.37	0.36	0.34	0.41
	$W^{0.7}_{0.8}$	1.20	1.23	1.19	0.68	1.64	1.62	1.67	1.17	0.43	0.43	0.42	0.46	1.01	1.20	1.18	0.73	1.43	1.57	1.65	1.14	0.32	0.30	0.29	0.35
	$W^{0.6}_{0.9}$	1.10	1.12	1.10	0.61	1.48	1.48	1.50	1.12	0.35	0.34	0.36	0.38	0.90	1.10	1.07	0.68	1.28	1.42	1.48	1.05	0.27	0.27	0.25	0.29
Deo	$W^{1.5}$													-	-	-	-	2.50	2.52	2.52	2.42	0.67	0.64	0.65	0.80
	$W^{1.2}_{0.4}$													-	-	-	-	2.25	2.29	2.29	2.16	0.59	0.57	0.57	0.68
	$W^{1.1}_{0.5}$													-	-	-	-	2.10	2.13	2.13	2.02	0.52	0.49	0.50	0.61
	$W^{0.8}_{0.7}$													-	-	-	-	1.65	1.67	1.67	1.62	0.31	0.29	0.29	0.35
	$W^{0.6}_{0.9}$													-	-	-	-	1.24	1.24	1.24	1.25	0.14	0.13	0.13	0.15
	$W^{0.4}_{1.3}$													-	-	-	-	0.88	0.85	0.86	0.92	0.03	0.02	0.02	0.03
Gal	$W^{1.5}$	1.76	1.62	1.90	1.95	2.21	2.06	2.46	2.24	0.89	0.89	0.87	0.94	1.85	1.50	1.91	1.93	2.37	1.98	2.49	2.29	0.65	0.66	0.60	0.76
	$W^{1.2}_{0.4}$	1.43	1.24	1.67	1.74	1.86	1.69	2.19	1.95	0.80	0.80	0.76	0.83	1.57	1.13	1.70	1.72	2.03	1.64	2.24	2.02	0.56	0.61	0.50	0.62
	$W^{1.1}_{0.5}$	1.25	1.07	1.54	1.59	1.68	1.53	2.04	1.80	0.73	0.74	0.70	0.77	1.40	1.00	1.55	1.58	1.85	1.49	2.09	1.88	0.52	0.56	0.45	0.57
	$W^{0.8}_{0.5}$	1.13	0.94	1.39	1.43	1.51	1.37	1.82	1.64	0.61	0.61	0.58	0.61	1.25	0.89	1.39	1.42	1.62	1.32	1.83	1.66	0.45	0.48	0.39	0.48
	$W^{0.7}_{0.9}$	1.01	0.84	1.28	1.33	1.39	1.29	1.66	1.49	0.51	0.51	0.48	0.52	1.15	0.83	1.26	1.30	1.47	1.18	1.65	1.53	0.38	0.41	0.34	0.41
	$W^{0.4}_{1.3}$	0.91	0.75	1.18	1.22	1.28	1.18	1.51	1.39	0.43	0.44	0.40	0.43	1.02	0.74	1.16	1.18	1.32	1.09	1.46	1.40	0.33	0.35	0.28	0.37

(b) Solute-solvent H-bonds.

Figure 3.S.1: Normalized probability distributions around the dihedral angle ω for the three compounds considered in the series S_p^H and S_p^h of artificial solvents at 298.15 K and 998.3 kg·m^{-3}. The values are calculated as reweighted averages over the 40 ns US phase of the different LEUS simulations using a bin size of $2°$. The different panels from left to right correspond to Glc, Deo and Gal (Figure 3.1), and from top to bottom to the series of artificial solvents (Table 3.3) S_p^H with decreasing permittivity at water-like H-bonding capacity (four upper panels) and S_p^h with decreasing permittivity at lower H-bonding capacity (four lower panels). The black curves correspond to the 53A6 force field[244] and the red curves to the 56A6$_{\text{CARBO}}$ force field[245]. The simulations of Deo with 53A6 are omitted due to ambiguous torsional potential definitions. See Figure 3.3 for corresponding results with the physical solvents and Suppl. Mat. Figures 3.S.2 and 3.S.3 for the other series of artificial solvents.
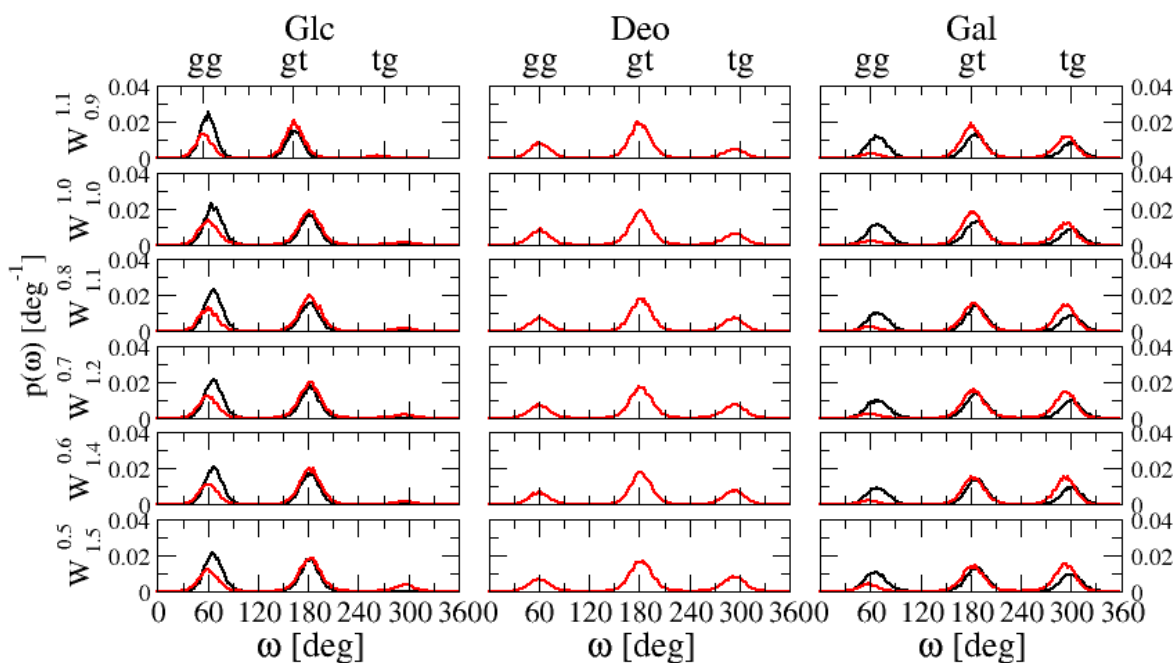
Figure 3.S.2: Normalized probability distributions around the dihedral angle ω for the three compounds considered in the series S_h^P of artificial solvents at 298.15 K and 998.3 kg·m^{-3}. The values are calculated as reweighted averages over the 40 ns US phase of the different LEUS simulations using a bin size of 2°. The different panels from left to right correspond to Glc, Deo and Gal (Figure 3.1), and from top to bottom to the series of artificial solvents (Table 3.3) S_h^P with decreasing H-bonding capacity at water-like permittivity. The black curves correspond to the 53A6 force field[244] and the red curves to the 56A6$_{CARBO}$ force field[245]. The simulations of Deo with 53A6 are omitted due to ambiguous torsional potential definitions. See Figure 3.3 for corresponding results with the physical solvents and Suppl. Mat. Figures 3.S.1 and 3.S.3 for the other series of artificial solvents.
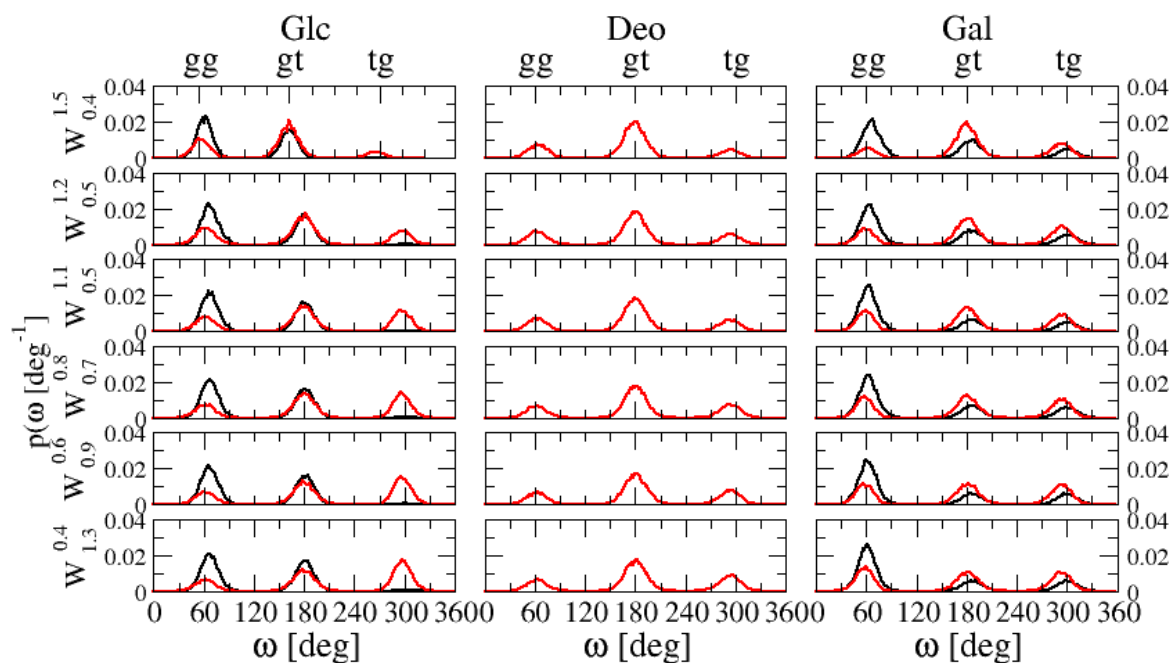
Figure 3.S.3: Normalized probability distributions around the dihedral angle ω for the three compounds considered in the series S_h^p of artificial solvents at 298.15 K and 998.3 kg·m^{-3}. The values are calculated as reweighted averages over the 40 ns US phase of the different LEUS simulations using a bin size of 2°. The different panels from left to right correspond to Glc, Deo and Gal (Figure 3.1), and from top to bottom to the series of artificial solvents (Table 3.3) S_h^p with decreasing H-bonding capacity at low permittivity. The black curves correspond to the 53A6 force field [244] and the red curves to the 56A6$_{\text{CARBO}}$ force field [245]. The simulations of Deo with 53A6 are omitted due to ambiguous torsional potential definitions. See Figure 3.3 for corresponding results with the physical solvents and Suppl. Mat. Figures 3.S.1 and 3.S.2 for the other series of artificial solvents.
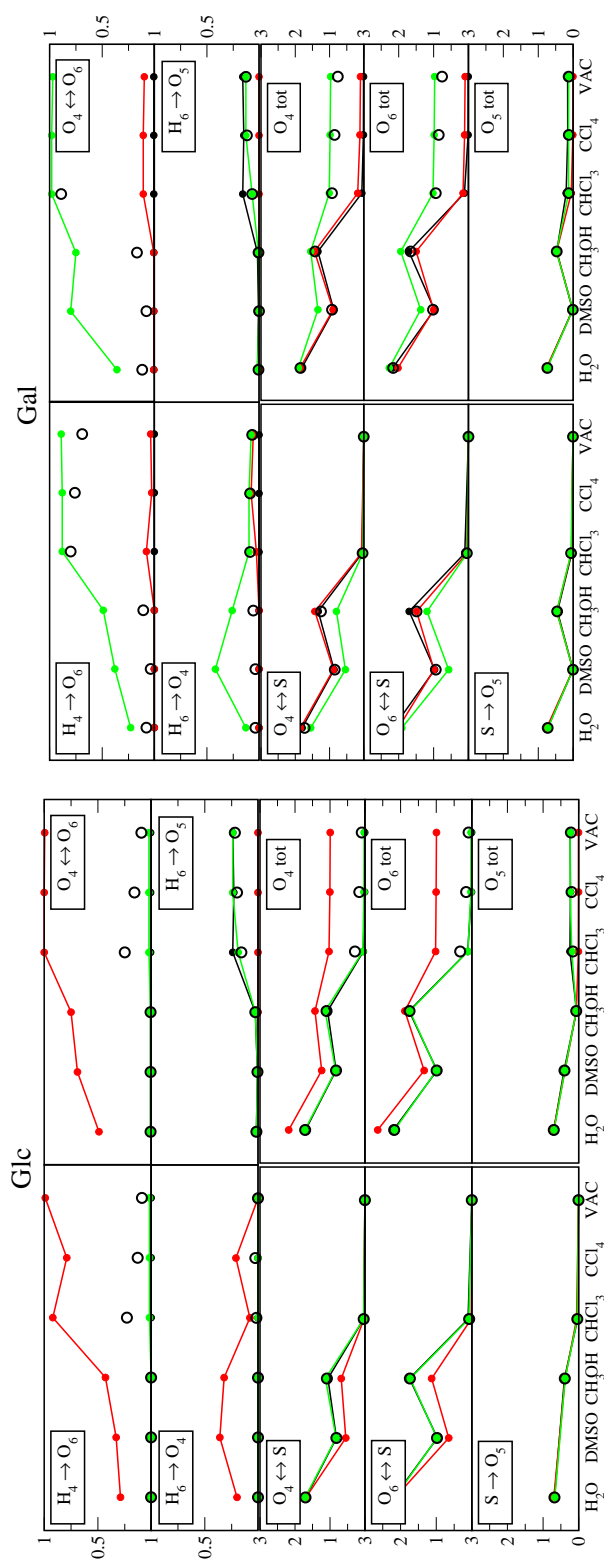
Figure 3.S.4: Detailed H-bond analysis for Glc and Gal in the series of physical solvents at 298.15 K and 1 bar using the 53A6 force field[244]. The values are calculated as averages over the 40 ns phase of the different LEUS simulations (1 μs plain SD simulation for VAC). The two columns correspond to Glc (left) and Gal (right). The 10 panels in each column refer to the average numbers of intramolecular $H_4{\rightarrow}O_6$, $H_6{\rightarrow}O_4$, $O_4{\leftrightarrow}O_6$ and $H_6{\rightarrow}O_5$ H-bonds, to the average numbers of solute-solvent $O_4{\leftrightarrow}S$, $O_6{\leftrightarrow}S$ and $S{\rightarrow}O_5$ H-bonds, where S and to the total numbers of H-bonds (sum of both types) involving atoms O_4, O_6 and O_5. The average numbers are calculated over the three canonical rotamers separately, *i.e.* as N_{gg} (green), N_{gt} (black) and N_{tg} (red), or over the entire conformational ensemble, *i.e.* as N (circles). The successive points follow the series of physical solvents (Table 3.2) in order of decreasing polarity. The data is reported numerically in Supp. Mat. Table 3.S.9.
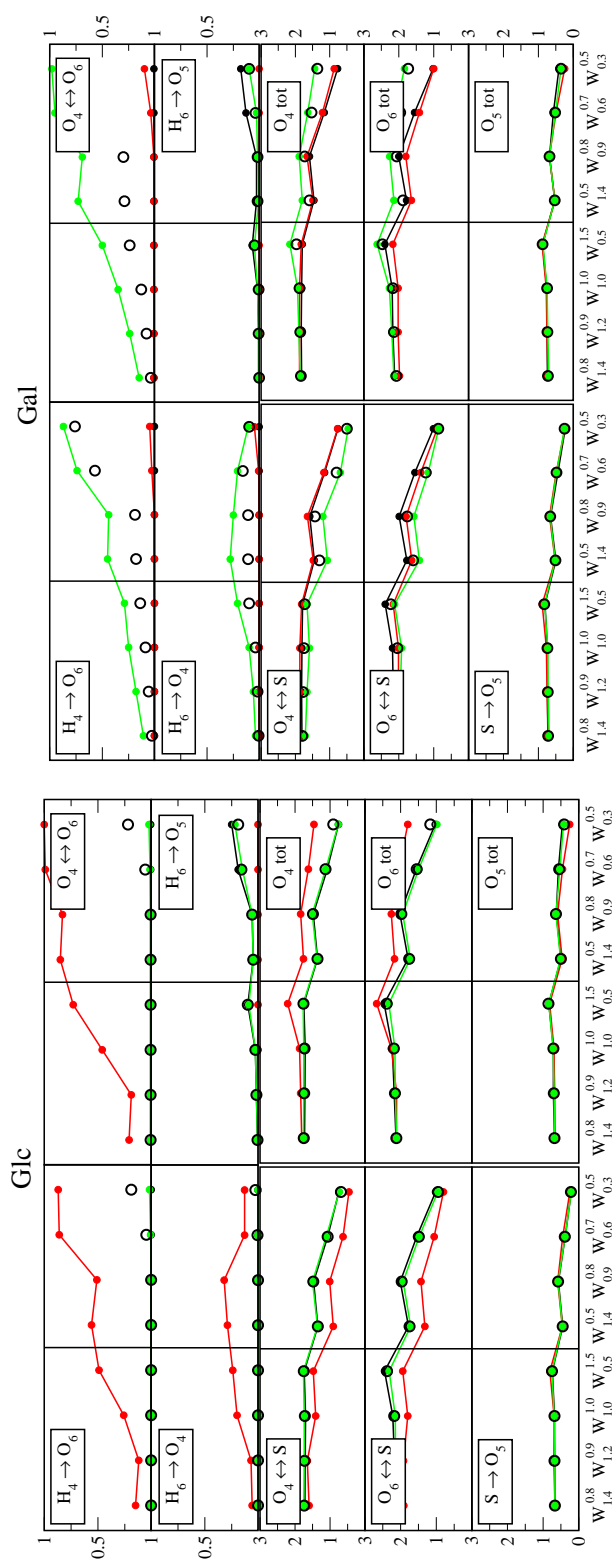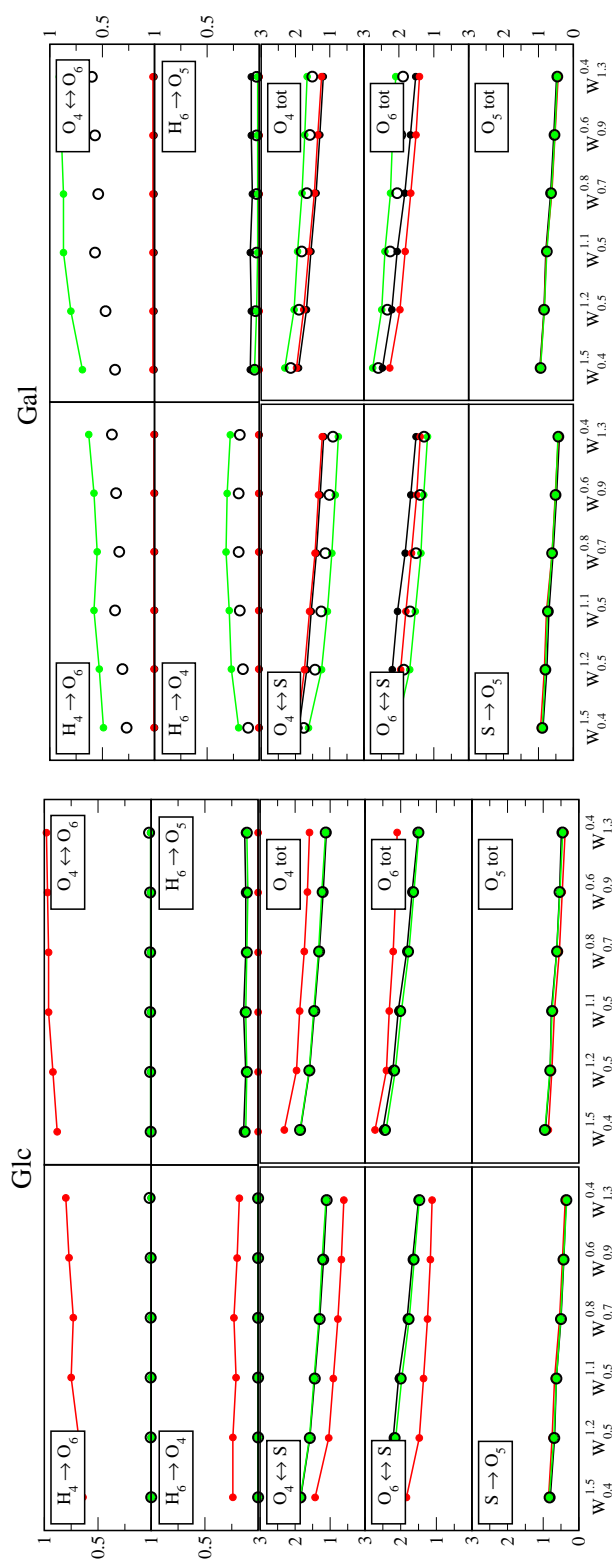
Figure 3.S.5: Detailed H-bond analysis for Glc and Gal in the series S_p^H and S_p^h of artificial solvents at 298.15 K and 998.3 kg·m^{-3} using the 53A6 force field[244]. The values are calculated as averages over the 40 ns phase of the different LEUS simulations. The two columns correspond to Glc (left) and Gal (right). The 10 panels in each column refer to the average numbers of intramolecular H$_4$→O$_6$, H$_6$→O$_4$, O$_4$↔O$_6$ and H$_6$→O$_5$ H-bonds, to the average numbers of solute-solvent O$_4$↔S, O$_6$↔S and S→O$_5$ H-bonds, where S is a solvent donor or acceptor site, and to the total numbers of H-bonds (sum of both types) involving atoms O$_4$, O$_6$ and O$_5$. The average numbers are calculated over the three canonical rotamers separately, *i.e.* as N_{gg} (green), N_{gt} (black) and N_{tg} (red), or over the entire conformational ensemble, *i.e.* as N (circles). The successive points follow the series S_p^H and S_p^h of artificial solvents (Table 3.3) in order of decreasing dielectric permittivity at water-like or lower H-bonding capacity, respectively. The data is reported numerically in Supp. Mat. Table 3.S.10.

Figure 3.S.6: Detailed H-bond analysis for Glc and Gal in the series S_h^P of artificial solvents at 298.15 K and 998.3 kg·m^{-3} using the 53A6 force field[244]. The values are calculated as averages over the 40 ns phase of the different LEUS simulations. The two columns correspond to Glc (left) and Gal (right). The 10 panels in each column refer to the average numbers of intramolecular H$_4$→O$_6$, H$_6$→O$_4$, O$_4$↔O$_6$ and H$_6$→O$_5$ H-bonds, to the average numbers of solute-solvent O$_4$↔S, O$_6$↔S and S→O$_5$ H-bonds, where S is a solvent donor or acceptor site, and to the total numbers of H-bonds (sum of both types) involving atoms O$_4$, O$_6$ and O$_5$. The average numbers are calculated over the three canonical rotamers separately, *i.e.* as N_{gg} (green), N_{gt} (black) and N_{tg} (red), or over the entire conformational ensemble, *i.e.* as N (circles). The successive points follow the series S_h^P of artificial solvent (Table 3.3) in order of decreasing H-bonding capacity at water-like dielectric permittivity. The data is reported numerically in Supp. Mat. Table 3.S.11.

Figure 3.S.7: Detailed H-bond analysis for Glc and Gal in the series S_h^p of artificial solvents at 298.15 K and 998.3 kg·m^{-3} using the 53A6 force field[244]. The values are calculated as averages over the 40 ns phase of the different LEUS simulations. The two columns correspond to Glc (left) and Gal (right). The 10 panels in each column refer to the average numbers of intramolecular H$_4$→O$_6$, H$_6$→O$_4$, H$_6$→O$_5$, O$_4$↔O$_6$ and H$_6$→O$_5$ H-bonds, to the average numbers of solute-solvent O$_4$↔S, O$_6$↔S and S→O$_5$ H-bonds, where S is a solvent donor or acceptor site, and to the total numbers of H-bonds (sum of both types) involving atoms O$_4$, O$_6$ and O$_5$. The average numbers are calculated over the three canonical rotamers separately, *i.e.* as N_{gg} (green), N_{gt} (black) and N_{tg} (red), or over the entire conformational ensemble, *i.e.* as N (circles). The successive points follow the series S_h^p of artificial solvent (Table 3.3) in order of decreasing H-bonding capacity at low dielectric permittivity. The data is reported numerically in Supp. Mat. Table 3.S.12.

Figure 3.S.8: Detailed H-bond analysis for the three compounds in the series of physical solvents at 298.15 K and 1 bar using the 56A6$_{CARBO}$ force field[245]. The values are calculated as averages over the 40 ns phase of the different LEUS simulations (1 μs plain SD for VAC). The two columns correspond to Glc (left) and Gal (right). The 10 panels in each column refer to the average numbers of intramolecular H$_4$→O$_6$, H$_6$→O$_4$, O$_4$↔O$_6$ and H$_6$→O$_5$ H-bonds, to the average numbers of solute-solvent O$_4$↔S, O$_6$↔S and S→O$_5$ H-bonds, where S is a solvent donor or acceptor site, and to the total numbers of H-bonds (sum of both types) involving atoms O$_4$, O$_6$ and O$_5$. The average numbers are calculated over the three canonical rotamers separately, *i.e.* as N_{gg} (green), N_{gt} (black) and N_{tg} (red), or over the entire conformational ensemble, *i.e.* as N (circles). The successive points follow the series of physical solvents (Table 3.2) in order of decreasing polarity. The data is reported numerically in Supp. Mat. Table 3.S.9.
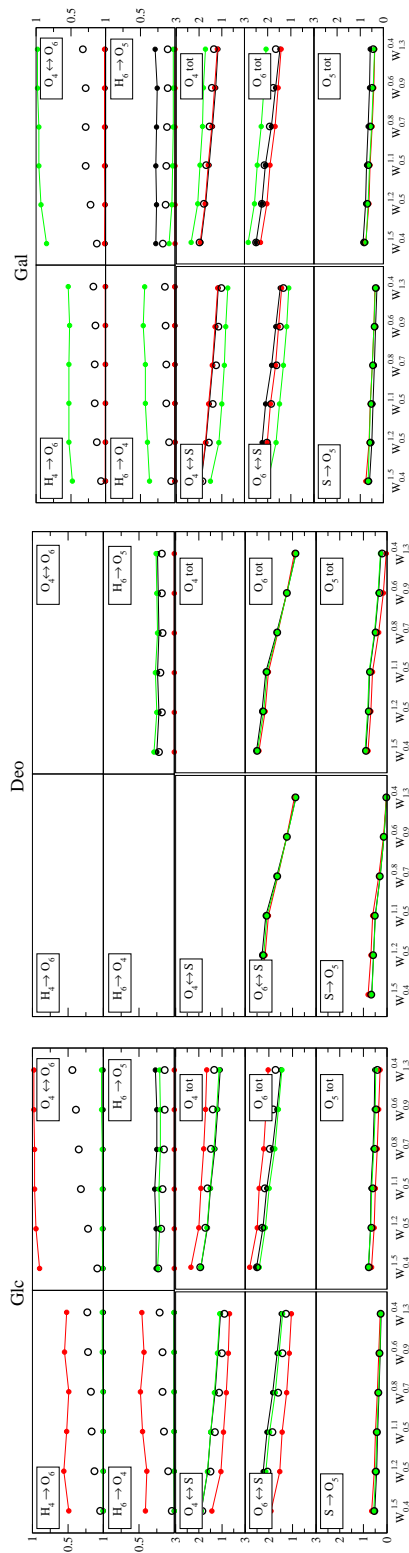
Figure 3.S.9: Detailed H-bond analysis for the three compounds in the series S_p^H and S_p^h of artificial solvents at 298.15 K and 998.3 kg·m^{-3} using the 56A6$_{\text{CARBO}}$ force field [245]. The values are calculated as averages over the 40 ns phase of the different LEUS simulations. The two columns correspond to Glc (left), Deo (middle) and Gal (right). The 10 panels in each column refer to the average numbers of intramolecular H$_4 \rightarrow$O$_6$, H$_6 \rightarrow$O$_4$, O$_4 \leftrightarrow$O$_6$ and H$_6 \rightarrow$O$_5$ H-bonds, to the average numbers of solute-solvent O$_4 \leftrightarrow$S, O$_6 \leftrightarrow$S and S$\rightarrow$O$_5$ H-bonds, where S is a solvent donor or acceptor site, and to the total numbers of H-bonds (sum of both types) involving atoms O$_4$, O$_6$ and O$_5$. The average numbers are calculated over the three canonical rotamers separately, *i.e.* as N_{gg} (green), N_{gt} (black) and N_{tg} (red), or over the entire conformational ensemble, *i.e.* as N (circles). The successive points follow the series S_p^H and S_p^h of artificial solvents (Table 3.3) in order of decreasing dielectric permittivity at water-like or lower H-bonding capacity, respectively. The data is reported numerically in Supp. Mat. Table 3.S.10.

Figure 3.S.10: Detailed H-bond analysis for the three compounds in the series S_h^P of artificial solvents at 298.15 K and 998.3 kg·m^{-3} using the 56A6$_{CARBO}$ force field[245]. The values are calculated as averages over the 40 ns phase of the different LEUS simulations. The two columns correspond to Glc (left), Deo (middle) and Gal (right). The 10 panels in each column refer to the average numbers of intramolecular H$_4{\rightarrow}$O$_6$, H$_6{\rightarrow}$O$_4$, O$_4{\leftrightarrow}$O$_6$ and H$_6{\rightarrow}$O$_5$ H-bonds, to the average numbers of solute-solvent O$_4{\leftrightarrow}$S, O$_6{\leftrightarrow}$S and S$\rightarrow$O$_5$ H-bonds, where S is a solvent donor or acceptor site, and to the total numbers of H-bonds (sum of both types) involving atoms O$_4$, O$_6$ and O$_5$. The average numbers are calculated over the three canonical rotamers separately, *i.e.* as N_{gg} (green), N_{gt} (black) and N_{tg} (red), or over the entire conformational ensemble, *i.e.* as N (circles). The successive points follow the series S_h^P of artificial solvent (Table 3.3) in order of decreasing H-bonding capacity at water-like dielectric permittivity. The data is reported numerically in Supp. Mat. Table 3.S.11.
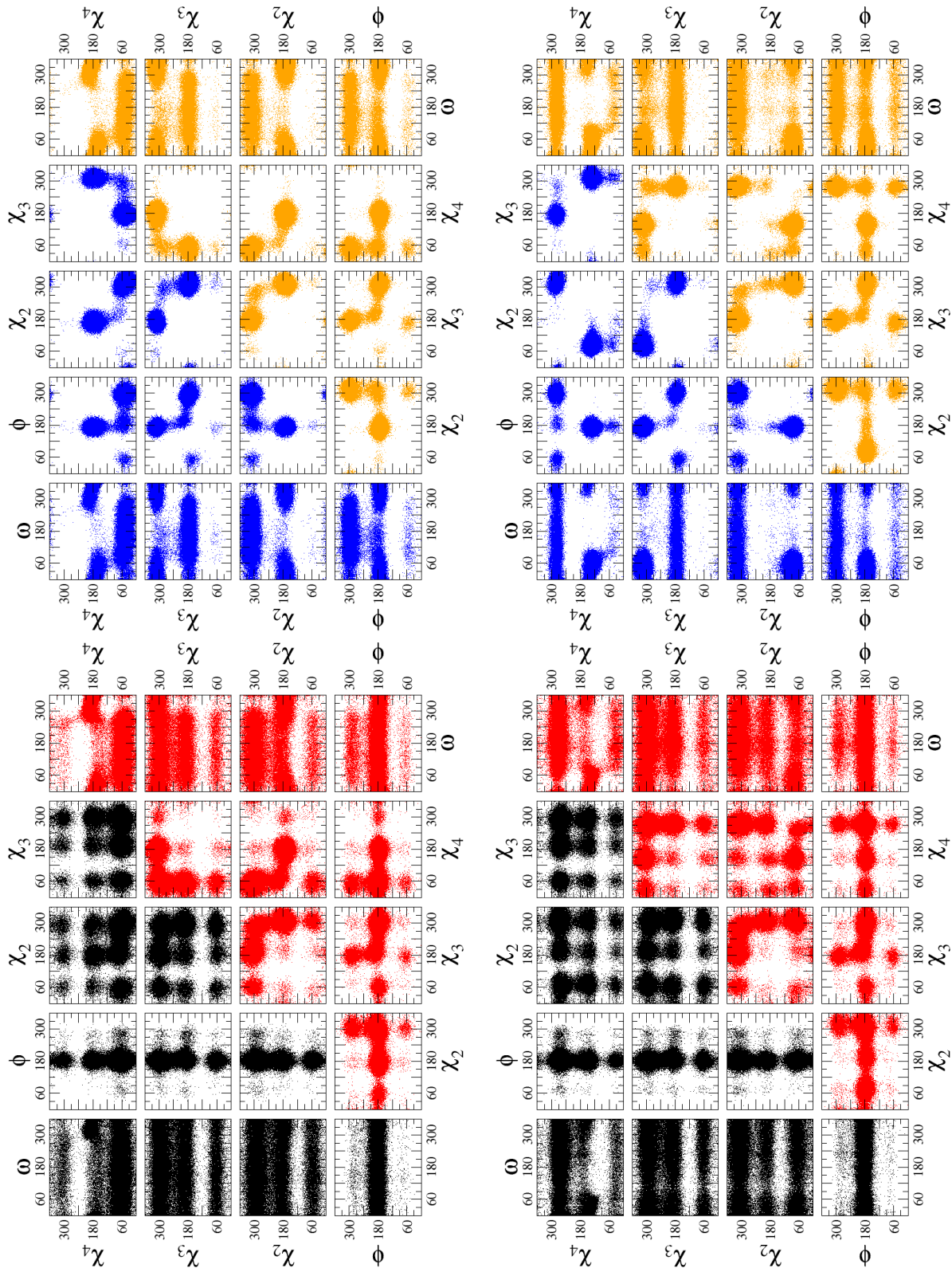
Figure 3.S.11: Detailed H-bond analysis for the three compounds in the series S_h^p of artificial solvents at 298.15 K and 998.3 kg·m^{-3} using the 56A6$_{CARBO}$ force field[245]. The values are calculated as averages over the 40 ns phase of the different LEUS simulations. The two columns correspond to Glc (left), Deo (middle) and Gal (right). The 10 panels in each column refer to the average numbers of intramolecular H$_4$→O$_6$, H$_6$→O$_4$, O$_4$↔O$_6$ and H$_6$→O$_5$ H-bonds, to the average numbers of solute-solvent O$_4$↔S, O$_6$↔S and S→O$_5$ H-bonds, where S is a solvent donor or acceptor site, and to the total numbers of H-bonds (sum of both types) involving atoms O$_4$, O$_6$ and O$_5$. The average numbers are calculated over the three canonical rotamers separately, *i.e.* as N_{gg} (green), N_{gt} (black) and N_{tg} (red), or over the entire conformational ensemble, *i.e.* as N (circles). The successive points follow the series S_h^p of artificial solvent (Table 3.3) in order of decreasing H-bonding capacity at low dielectric permittivity. The data is reported numerically in Supp. Mat. Table 3.S.12.

Figure 3.S.12: Pairwise correlations between the exocyclic dihedral angles of Glc and Gal in the physical solvents H_2O and CCl_4 at 298.15 K and 1 bar, and in the artificial solvents $W_{1.3}^{0.4}$ and $W_{0.3}^{0.5}$ at 298.15 K and 998.3 kg·m^{-3}. The individual points correspond to structures sampled at 0.5 ps intervals along the 40 ns US phase of the different LEUS simulations using the 53A6 force field[244]. The dihedral angles are defined as C_4-C_5-C_6-O_6 (ω), O_5-C_1-O_1-H_1 (ϕ), C_1-C_2-O_2-H_2 (χ_2), C_2-C_3-O_3-H_3 (χ_3) and C_3-C_4-O_4-H_4 (χ_4). The rows from top to bottom correspond to Glc and Gal (Figure 3.1). The solvents considered (Tables 3.2 and 3.3) are H_2O (black), CCl_4 (blue), $W_{1.3}^{0.4}$ (red) and $W_{0.3}^{0.5}$ (orange).

A molecular dynamics study of the solvent-modulated influence of intramolecular hydrogen-bonding on the conformational properties of disaccharides.

Abstract

Hydrogen-bonding (H-bonding) is often regarded as a determinant driving force in the conformational preferences of (bio)molecules. However, the influence of H-bonds depends on the solvent environment. In water, solvent-exposed H-bonds have only a very limited influence, being an opportunistic consequence of the close proximity of two H-bonding groups in a given conformation. In lower polarity solvents, the influence of solvent-exposed H-bonding increases, becoming a very significant conformational steering force.

Epimerization of disaccharides can modify their H-bonding patterns. When considering cellobiose and its epimers at C_2 and C'_3, the typical interresidue H-bond ($H'_3 \rightarrow O_5$), often considered responsible for the equilibrium conformation of cellobiose, is inhibited when the exocyclic hydroxyl group at C'_3 is in axial position. In previous work, MD simulations of the four compound in water revealed that no large conformational changes connected to the glycosidic dihedral angles ϕ and ψ occur to recover the missing intramolecular H-bond. In this case, the carbohydrates experience a compensation effect from the solvent molecules.

In the present study, MD simulations of the same four disaccharides are performed in 6 physical solvents (H_2O, DMSO, CH_3OH, $CHCl_3$, CCl_4 and vacuum) along with 19 artificial water-like solvent models, where the dielectric permittivity and H-bonding capacity can be varied independently *via* a scaling of the oxygen-hydrogen distance and of the atomic partial charges.

It is found that when several H-bonds are possible, both intra- and interresidue, their influence can act either in a cooperative way or in a competitive way. In the first case, specific H-bonds concurr to enhance the stability of a specific conformation, shifting the conformational preference towards this conformation. In the second case, incompatible H-bonds compete, representing in this case an adverse force for the specific conformations.

4.1 Introduction

In the conformational analysis of (bio)molecules, different qualitative or semi-quantitative effects are commonly invoked to rationalize the lower free energy of the experimentally dominant state(s) relative to alternative ones. Effects or classes of interactions stabilizing a specific conformation (*e.g.* native for a macromolecule or bound for a host-guest complex) relative to less structured ones (*e.g.* unfolded or unbound) are termed *driving* forces (controlling affinity), while those stabilizing one of the structured states relative to alternative ones (*e.g.* alternative folds or binding modes) have been termed[132] *steering* forces (controlling selectivity). In this kind of analysis, many different types of effects can be included, such as steric, stereoelectronic, electrostatic (*e.g.* hydrogen bonding) and hydrophobic effects[294–301].

Hydrogen bonding (H-bonding) is often considered to be a determinant driving force in the conformational preferences of (bio)molecules. Indeed, early predictions concerning the native state of proteins[126] and nucleic acids[127] were based on the consideration of optimal H-bonding patterns. However, the influence of H-bonding cannot be discussed without taking into account of the solute environment, *i.e.* the effect of the solvent. The importance of H-bond formation for the conformational affinity or selectivity depends on the degree of exposure of both the separated H-bonding partners (in one conformation) and the formed H-bond (in the other conformation) to the solvent. In particular, the importance of H-bond formation is different when two solvated H-bonding partners associate to form a buried or a solvent-exposed H-bond. When forming a buried H-bond, the electrostatic gain upon H-bond formation should be compared with the initial desolvation penalty of the H-bonding partners. When forming a solvent-exposed H-bond, the formed H-bond can be subject to H-bonding competition by the solvent molecules and the interaction screened by the solvent dielectric response.

In the context of carbohydrates in solution, the influence of solvent-exposed H-bonding on the stability of specific conformations is particularly relevant. In an aqueous environment, solvent-exposed H-bonds in carbohydrates can be considered to represent a minor (possibly negligible or even, in some cases, adverse) conformational driving as well as steering force (see Ref.[132] and Chapter 3). They should be viewed as an opportunistic consequence of the close proximity of two H-bonding groups in a given conformation, and not as a factor contributing to the stability of this conformation[132,171]. However, as the polarity of the solvent is decreased, solvent-exposed intramolecular H-bonding progressively evolves from a negligible (possibly adverse) to a very significant (favorable) conformational driving force.

This view is compatible with the experimental observation that the hydroxymethyl rotamers of glucose and galactose permitting the formation of a solvent-exposed H-bond between this group and the hydroxyl group at C_4 are not favored but, in the opposite, least populated in water[184,185,191]. The evolution of the rotameric preferences in lower polarity solvents was investigated in Chapter 3 As the solvent polarity is decreased, the populations of the conformers compatible with the formation of the intramolecular H-bond was found to increase significantly.

Evaluations concerning the importance of H-bonding are comparatively easy when considering a simple system such as a monosaccharide. On the other hand, when considering more complex systems such as oligosaccharides or polysaccharides, the number of factors involved in the stability of specific conformations can increase drastically. Additionally, this kind of molecules shows secondary-structure patterns analogous to the ones found in proteins, although often local and transient. For this reason, the corresponding conformations are commonly classified based on similar schemes[302–305], *e.g.* Ramachandran maps for the glycosidic linkages and secondary-structure elements such as regular helices). Note, however, that polysaccharides evidence a more pronounced constitutional variety

and conformational flexibility compared to proteins.

Disaccharides are the simplest carbohydrates involving all types of degrees of freedom present in longer chains, while still involving a limited number of effects. These molecules are thus particularly appealing for theoretical investigations. Numerous studies concerning the conformational preferences of disaccharides can be found in the literature, both experimental[306–314] and theoretical[132,169,315–321]. Many of these invoked solvent-exposed H-bond formation to rationalize specific conformational preferences of aqueous disaccharides[302,306,322–324]. In contrast, a smaller number of studies[132,171,245] (see also Chapter 3) suggested that, in the opposite, solvent-exposed H-bonding should not be considered as a major factor influencing those preferences in an aqueous environment.

The starting point of the present study is the article of Wang *et al.*[132] concerning the conformational preferences of cellobiose and its C_2 and C'_3 epimers in aqueous solution. The four epimers considered present a different stereochemistry for the potentially H-bonding groups neighboring the glycosidic linkage. The absence of significant differences in the distributions of the glycosidic dihedral angles ϕ and ψ supported the suggestion of a negligible influence of solvent-exposed H-bonding on the conformational properties of these four compounds in water. The present study revisits this system, now considering the influence of the solvent polarity on the conformational properties. The same four epimers are simulated here using explicit-solvent molecular dynamics (MD) simulations in physical solvents of various polarities, as well as in the artificial solvents developed in Chapter 2. These were previously employed (Chapter 3) to disentangle the influence of the solvent dielectric permittivity and H-bonding capacity, while preserving water-like dispersive interactions as well as molecular size and shape, on the rotameric equilibrium of the hydroxymethyl group in glucose and galactose.

4.2   Computational details

4.2.1   Simulated Systems

All MD simulations were carried out using the GROMOS MD++ simulation program[81–84] along with the $56A6_{\text{CARBO}}$ parameter set[245] and different solvent models (see below). The $56A6_{\text{CARBO}}$ parameter set of Hansen & Hünenberger[245] results from a complete re-optimization of the GROMOS 53A6 parameter set[244] for biomolecules in the context of hexopyranose-based carbohydrates, carried out in 2011. Note that this set was recently revised[246] to a new set $56A6_{\text{CARBO\_R}}$ improving the description of the ring-conformational
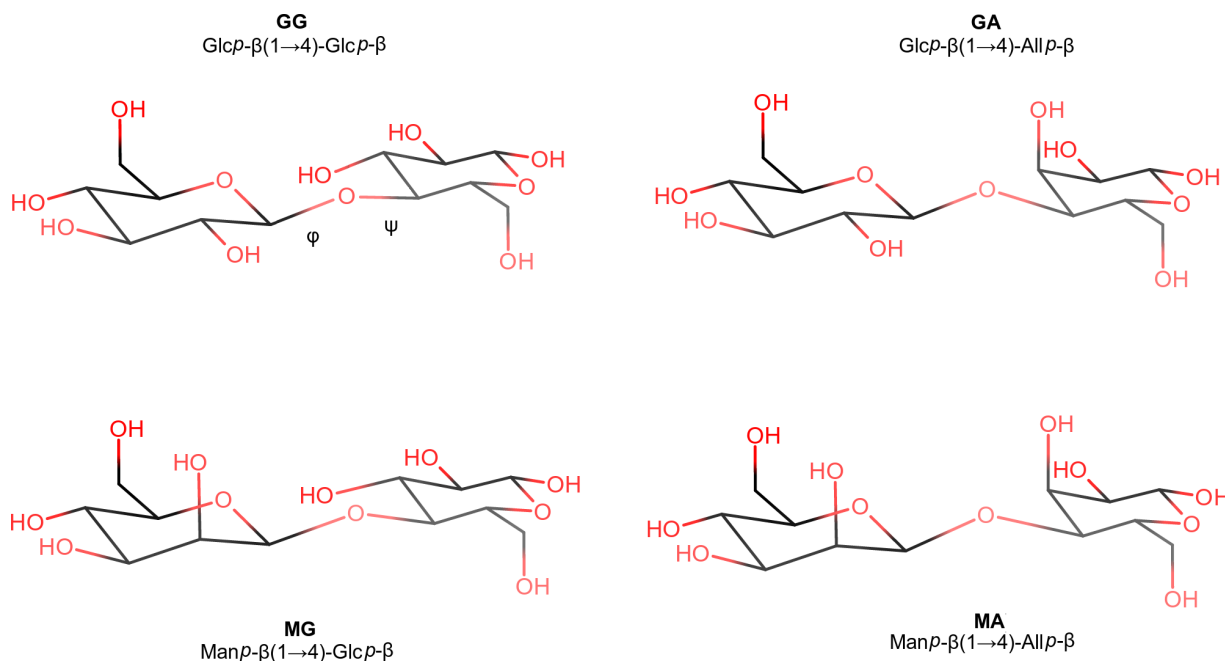
Figure 4.1:   Four $\beta(1{\to}4)$-linked D-aldohexopyranose disaccharides considered in the present study.   The four disaccharides considered are Glc$p$-$\beta(1{\to}4)$-Glc$p$-β (GG, β-cellobiose), and its C_2 or/and C'_3 epimers, namely Glc$p$-$\beta(1{\to}4)$-All$p$-β (GA), Man$p$-$\beta(1{\to}4)$-Glc$p$-β (MG) and Man$p$-$\beta(1{\to}4)$-All$p$-β (MA), where Glc$p$, Man$p$ and All$p$ stand for D-glucopyranose, D-mannopyranose and D-allopyranose, respectively. The disaccharides are simulated with β anomery at the reducing residue (primed atom labels) and the two residues are restrained to the $^4C_1$ chair conformation (as displayed) during the simulations. The glycosidic dihedral angles ϕ and ψ are defined by the atom sequences O_5-C_1-O_1-C'_4 and C_1-O_1-C'_4-C'_3, respectively.

properties for residues within chains. However, this change would not be expected to affect significantly the results presented here as dihedral-angle restraints are applied to prevent ring inversions (see Section 4.2.2).

The simulations involved cubic computational boxes containing one solute molecule and N_s solvent molecules. The solutes considered are the four $\beta(1{\rightarrow}4)$-linked D-aldohexo-pyranose disaccharides with β anomery at the reducing residue illustrated in Figure 4.1, namely Glc$p$-$\beta(1{\rightarrow}4)$-Glc$p$-β (GG, β-cellobiose), Glc$p$-$\beta(1{\rightarrow}4)$-All$p$-β (GA), Man$p$-$\beta(1{\rightarrow}4)$-Glc$p$-β (MG) and Man$p$-$\beta(1{\rightarrow}4)$-All$p$-β (MA), where Glc$p$, Man$p$ and All$p$ stand for D-glucopyranose, D-mannopyranose and D-allopyranose, respectively. Considering the most stable $^4C_1$ chair conformation of the two residues and labeling the atoms of the reducing residue with a prime, the four disaccharides present equatorial orientations of the hydroxyl groups at C_3, C_4, C_1' and C_2' as well as of the hydroxymethyl groups at C_5 and C_5', along with different orientations of the hydroxyl groups at C_2 and C_3', namely equatorial-equatorial (GG), equatorial-axial (GA), axial-equatorial (MG) or axial-axial (MA). Note that the parameters of the four disaccharides in the 56A6$_\text{CARBO}$ force field differ exclusively in terms of the sign of the reference improper-dihedral angle controlling the stereochemistry at carbon atoms C_2 and C_3', *i.e.*, the bond, bond-angle, torsional and non-bonded interaction parameters of the four compounds are otherwise rigorously identical. The four solutes were simulated in various physical and artificial solvents.

For the physical solvents, the models considered are the SPC water (H_2O) model of Berendsen *et al.*[136] ($N_s = 2081$), the dimethyl-sulfoxide (DMSO) model of Geerke *et al.*[149] ($N_s = 543$), the methanol (CH_3OH) model of Walser *et al.*[252] ($N_s = 952$), the chloroform ($CHCl_3$) model of Tironi and van Gunsteren[254] ($N_s = 479$), and the carbon-tetrachloride (CCl_4) model of Tironi *et al.*[256] ($N_s = 398$). The indicated values of N_s result in cubic computational boxes of about 4.0 nm edge length in all cases. The four solutes were also simulated in vacuum (VAC), using stochastic dynamics (SD)[100,101] instead of MD in this

Solvent	ρ [kg·m^{-3}]				μ [D]				ε				δH	
	Exp.	Ref.	Sim.	Ref.	Exp.	Ref.	Sim.	Ref.	Exp.	Ref.	Sim.	Ref.	Exp.	Ref.
H$_2$O	997	[147]	973	[138]	1.85	[249]	2.27	[136]	78.5	[138]	66.0	[138]	42.3	[250]
DMSO	1095	[251]	1096	[149]	3.96	[166]	5.25	[149]	46.0	[251]	38.0	[149]	10.2	[250]
CH$_3$OH	791	[251]	785	[252]	1.70	[253]	2.29	[252]	32.6	[253]	19.8	[252]	22.3	[250]
CHCl$_3$	1489	[251]	1520	[254]	4.81	[251]	2.30	[255]	4.8	[253]	2.4	[254]	5.7	[250]
CCl$_4$	1595	[253]	1601	[256]	0.00		0.00		2.2	[166]	1.0	[256]	-	
VAC	-		-		-		-		1.0		1.0		-	

Table 4.1: Experimental and simulated properties of the physical solvents considered in the present study at 298.15 K and 1 bar. The quantities reported are the density ρ, the gas-phase molecular dipole moment μ, the static relative dielectric permittivity ε, and the Hansen parameter[250] δH as a measure of the H-bonding propensity.

specific case. For convenience, basic properties characterizing the polarity of the five physical solvents, either experimental or based on previous simulations with the indicated models, are summarized in Table 4.1.

The artificial solvents considered are a subset of the tunable solvents introduced in Chapter 2 and used in Chapter 3 to be employed under constant-volume (NVT) conditions. These water-derived models were generated starting from the SPC water model of Berendsen *et al.*[136], by changing systematically the oxygen-hydrogen bond length (scaling factor s_b) and the atomic partial charges (scaling factor s_q), without any change in the Lennard-Jones interaction parameters. The scaling factors s_b and s_q were both varied systematically by increments of 0.1 in the range 0.1 to 1.5, leading to the definition of 195 water-derived models (35 extreme combinations of s_b and s_q were not retained as the resulting models could only be simulated using very small timesteps). These models have different dielectric permittivities and H-bonding capacities, but identical dispersive interactions as well as molecular shape and size. They are labeled $W_{s_q}^{s_b}$ according to the values of the two scaling factors, so that in particular $W_{1.0}^{1.0}$ is the SPC water model. The main thermodynamic, dynamic, dielectric and H-bonding properties of these 195 solvents as calculated based on pure-liquid simulations, can be found in Suppl. Mat. Table 2.S.1 of Chapter 2. Their dielectric permittivities ε and H-bonding capacities n_H, the latter value being defined as the average number of H-bonds per molecule in the pure liquid, are illustrated graphically in Figure 4.2. In this set of 195 solvents, a subset of 19 is considered here, corresponding to four series S_p^H, S_p^h, S_h^P, and S_h^p. These series permit to investigate specifically the effect of the following trends: (*i*) in series S_p^H, the progressive decrease of the permittivity (*p* subscript) at water-like H-bonding capacity (*H* superscript); (*ii*) in series S_p^h, the progressive decrease of the permittivity (*p* subscript) at lower H-bonding capacity (*h* superscript); (*iii*) in series S_h^P, the progressive decrease of the H-bonding capacity (*h* subscript) at water-like permittivity (*P* superscript); (*iv*) in series S_h^p, the

progressive decrease of the H-bonding capacity (h subscript) at lower permittivity (p superscript). The properties of the 19 artificial solvents along with the definition of the four series are summarized in Table 4.2. These models and the corresponding series are also shown graphically in Figure 4.2. Note that the SPC water model $W_{1.0}^{1.0}$ belongs to both series S_p^H and S_h^P. For the artificial solvents, the number of solvent molecules N_s in the computational box was always set to 2081 for a cubic box of edge length 4.0 nm.

Series	Solvent	s_b	s_q	ε	n_H
S_p^H	$W_{1.4}^{0.8}$	0.8	1.4	115	3.4
	$W_{1.2}^{0.9}$	0.9	1.2	89	3.5
	$W_{1.0}^{1.0}$ (SPC)	1.0	1.0	64	3.4
	$W_{0.5}^{1.5}$	1.5	0.5	25	3.4
S_p^h	$W_{1.4}^{0.5}$	0.5	1.4	36	1.7
	$W_{0.9}^{0.8}$	0.8	0.9	33	2.0
	$W_{0.6}^{0.7}$	0.7	0.6	7	1.5
	$W_{0.3}^{0.5}$	0.5	0.3	1	1.4
S_h^P	$W_{0.9}^{1.1}$	1.1	0.9	56	3.6
	$W_{1.0}^{1.0}$ (SPC)	1.0	1.0	64	3.4
	$W_{1.1}^{0.8}$	0.8	1.1	64	2.5
	$W_{1.2}^{0.7}$	0.7	1.2	65	2.2
	$W_{1.4}^{0.6}$	0.6	1.4	76	2.0
	$W_{1.5}^{0.5}$	0.5	1.5	49	1.7
S_h^p	$W_{0.4}^{1.5}$	1.5	0.4	14	2.6
	$W_{0.5}^{1.2}$	1.2	0.5	15	2.2
	$W_{0.5}^{1.1}$	1.1	0.5	13	1.9
	$W_{0.7}^{0.8}$	0.8	0.7	15	1.7
	$W_{0.9}^{0.6}$	0.6	0.9	14	1.6
	$W_{1.3}^{0.4}$	0.4	1.3	12	1.5

Table 4.2: Definition and simulated properties of the 19 artificial solvent models considered in the present study at 298.15 K and 968 kg·m^{-3}. The reported quantities are the scaling factors s_b and s_q applied to the oxygen-hydrogen bond length and to the atomic partial charges, respectively, relative to the SPC water model[136], the static relative dielectric permittivity ε of the liquid, and the average number n_H of H-bonds per molecule in the liquid as a measure of its H-bonding capacity. The models are grouped into four series as described in section 2.1. Note that the SPC water model $W_{1.0}^{1.0}$ belongs to both series S_p^H and S_h^P. These 19 models are a subset of the 195 models developed in Chapter 2, (see also Chapter 3) their main thermodynamic, dynamic, dielectric and H-bonding properties being reported in Suppl. Mat. Table 2.S.1 of Chapter 2.
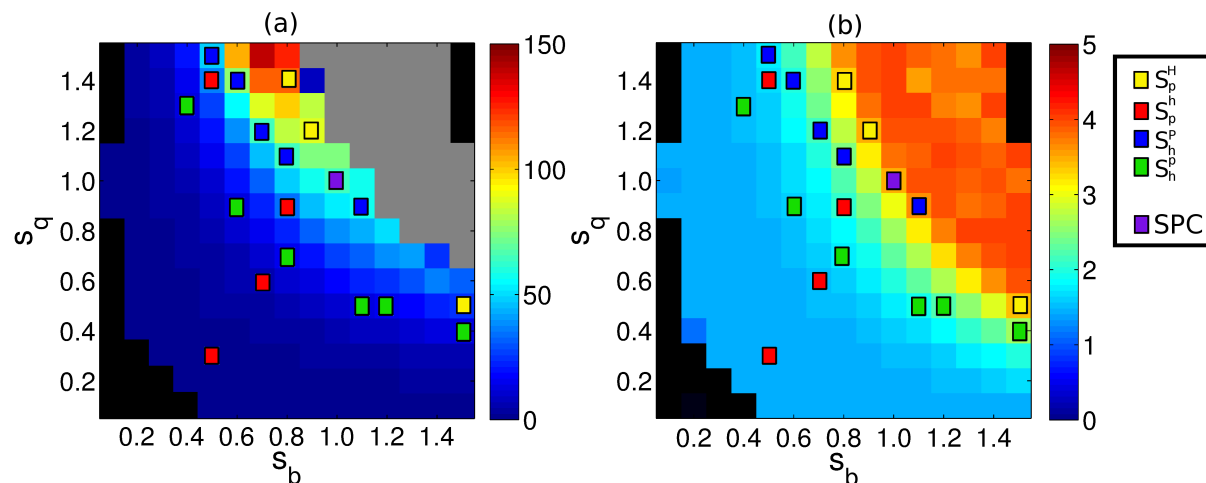
Figure 4.2: Simulated properties of the 195 artificial solvent models developed in Chapter 2 at 298.15 K and 968 kg·m^{-3}. The static relative dielectric permittivity ε of the liquid (a) and the average number n_H of H-bonds per molecule in the liquid (b) are shown as a function of the scaling factors s_b and s_q applied to the oxygen-hydrogen bond length and to the atomic partial charges, respectively, relative to the SPC water model[136]. The squares identify the subset of 19 models considered in the present study, colored according to the four series (S_p^H, S_p^h, S_h^P, and S_h^p; Table 4.2). The SPC model $W_{1.0}^{1.0}$ is shown in violet and belongs to both S_p^H and S_h^P. The areas in black and gray correspond to models that are either not in the set (30 s_b and s_q combinations were disregarded as the corresponding models could only be simulated with very short timesteps) or not liquid (glassy state), respectively, at the temperature and density considered. The main thermodynamic, dynamic, dielectric and H-bonding properties of the 195 models are reported numerically in Suppl. Mat. Table 2.S.1 of Chapter 2.

4.2.2 Simulations

The simulations performed include explicit-solvent MD simulations of the four solutes (GG, GA, MG, MA) in the two different sets of solvents (5 physical and 19 artificial), where the sampling along the dihedral angles ϕ and ψ was enhanced by the local-elevation umbrella-sampling (LEUS) approach[14,90,105], along with plain SD simulations[100,101] of the four compounds in vacuum. Additional plain MD simulations involving restraints to specific conformational regions were also performed in the physical solvents.

The MD simulations were performed under periodic boundary conditions based on cubic computational boxes containing one solute and N_s solvent molecules (values of N_s given in

Section 4.2.1). For the physical solvents, the simulations were carried out in the isothermal-isobaric (NPT) ensemble at 298.15 K and 1 bar. For the artificial solvents, the simulations were carried out in the canonical (NVT) ensemble at 298.15 K and an effective solvent density of 998.7 kg·m^{-3}. Here, the effective solvent density is estimated as $V^{-1}M(N_s+\tilde{N}_s)$, where V is the box volume, M the molecular mass of water, and $\tilde{N}_s$ an effective number of water molecules accounting for the solute volume (set here to $\tilde{N}_s = 24$).

The simulations involving the artificial solvent models must be carried out under NVT conditions because the Lennard-Jones interaction parameters of these models were not adjusted to reproduce the equilibrium density of water at 298.15 K and 1 bar. Accordingly, under NVT conditions at 298.15 K and 968 kg·m^{-3}, they are characterized by equilibrium pressures that range from -0.6 to 8.9 kbar for the 19 models considered here (see Suppl. Mat. Table 2.S.1 of Chapter 2). However, incorrect pressures are expected to have only a limited influence on the glycosidic dihedral-angle rotation in disaccharides, as was the case for the rotameric equilibrium of the hydroxymethyl group in monosaccharides (see Section 2.2.2 in Chapter 3).

In both the NPT and NVT simulations, the temperature was maintained close to its reference value of 298.15 K by weakly coupling[115] solute and solvent degrees of freedom jointly (to avoid solute damping[264]) to an external bath using a relaxation time of 0.1 ps. In the NPT simulations, the pressure was also maintained close to its reference value of 1 bar by weakly coupling[115] the atomic coordinates and the box dimensions (isotropic coordinate scaling, group-based virial) to an external bath using a relaxation time of 0.5 ps and an isothermal compressibility of 4.575·10^{-4} kJ^{-1}·mol·nm^3 as appropriate for aqueous biomolecular systems[101]. The compressibility was not adjusted for the different solvents, because it is combined with the arbitrary choice of a pressure relaxation time and does not affect the average thermodynamic properties of the system. The center of mass translation was removed every timestep.

The leap-frog algorithm[103] was used to integrate Newtonâs equations of motion with a timestep of 2 fs. Solute bond-length constraints as well as the full rigidity of the solvent molecules were enforced by application of the SHAKE procedure[53] with a relative geometric tolerance of 10^{-4}. The non-bonded interactions were calculated using a twin-range scheme[69], with short- and long-range cutoff distances set to 0.8 and 1.4 nm, respectively, and an update frequency of 5 timesteps for the short-range pairlist and intermediate-range interactions. A reaction-field correction[60,65] was applied to account for the mean effect of electrostatic interactions beyond the long-range cutoff distance, using the relative dielectric permittivity appropriate for the solvent model considered (Tables 4.1 and 4.2).

The LEUS method[14,90,105,132,171] was applied to enhance the sampling around the dihedral angles ϕ and ψ (O_5-C_1-O_1-C_4' and C_1-O_1-C_4'-C_3', see Figure 4.1) characterizing the conformation of the glycosidic linkage, which can present long-timescale transitions in unbiased simulations (on the order of[171] 10 ns up to a few μs). The LEUS calculations involved two steps: (i) a local elevation (LE) build-up phase[90] of duration $t_{LE} = 50$ ns, to progressively optimize a two-dimensional memory-based biasing potential in the space defined by ϕ and ψ; (ii) an umbrella sampling (US) phase[105] of duration $t_{US} = 60$ ns, using this preoptimized (now time-independent) biasing potential to enhance the sampling. The biasing potential was represented by means of a grid of 32×32 truncated-polynomial basis functions[270] with a spacing of 11.25° along both ϕ and ψ. The polynomial widths were set equal to the grid spacing. The LE build-up phase relied on a fixed force-constant increment per visit set to $k_{LE} = 10^{-4}$ kJ·mol^{-1}.

Five additional LEUS simulations were performed for the GG disaccharide in each of the five physical solvents in order to check the convergence of the method with respect to the memory build-up time. These involved biasing potentials obtained from LE build-up phases of different durations ($t_{LE} = 0$, 10, 20, 30 or 40 ns) along with a sampling phase of normal duration $t_{US} = 60$ ns. All the other parameters were the same as for the main

simulations. Together with the main simulation involving $t_{LE} = 50$ ns, the six simulations involving different t_{LE} are labelled t_{LE}-LEUS. The convergence of the method with respect to the US sampling time was also assessed, by truncating the US phase of the simulation with $t_{LE} = 50$ ns after different durations ($t_{US} = 10, 20, 30, 40, 50, 60$ ns). These six simulations involving different t_{US} are labelled t_{US}-LEUS

Four additional plain MD simulations (no LEUS) were also performed for each of the four disaccharides in each of the five physical solvents, involving restraints confining the ϕ and ψ dihedral angles to specific ranges. These ranges correspond to the areas labelled B, C_1, C_2 and C_3 in the Ramachandran map of Figure 4.3 (further discussed in Section 4.2.3). The restraint involved flat-bottom half-harmonic potentials applied starting from the edges of the corresponding regions with a force constant $k_{res} = 0.1$ kJ·mol^{-1}·deg^{-2}. These simulations were carried out for a duration $t_{MD} = 100$ ns. All the other parameters were the same as for the main simulations. These simulations are labelled MD-X, where X refers to the specific range (B, C_1, C_2 or C_3) considered.

The plain SD simulations (no LEUS) in vacuum were performed by integrating the Langevin equation of motion[100,101]. They relied on a reference temperature of 298.15 K and a friction coefficient of 91 ps^{-1}. The choice of the latter value, appropriate for water[100], has no effect on the thermodynamic properties of the system. The SD simulations were carried out for a duration $t_{SD} = 1$ μs.

Because ring conformational transitions (chair inversion from $^4C_1$ to $^1C_4$ or conversion to boat conformations) occur experimentally on the 0.05-1 μs timescale[269,325,326], and although the corresponding alternative conformations contribute negligibly to the conformational ensembles of most aqueous aldohexopyranoses under ambient conditions[261], their occasional occurrence may compromise the statistical accuracy of the simulated results on the timescale of a few tens of nanoseconds. For this reason, a restraining of the improper

dihedral angles α_1, α_2 and α_3 (defined according to Pickett and Strauss[271] as 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 decreased by 180°) was applied in all simulations to maintain the two rings in a $^4C_1$ conformation. The restraints relied on reference values of $-35°$ for α_1, α_2 and α_3, along with a force constant of 165.15 kJ·mol^{-1}. Additional simulations investigating the possible influence of the ring conformation on the glycosidic linkage preferences are reported and discussed in Appendix 4.A.

For the MD+LEUS, the plain MD and the SD simulations, the initial structure of the solute was in the $^4C_1$ chair conformation for both residues, and the glycosidic linkage was set in an initial conformation with $\phi = 300°$ and $\psi = 60°$. After filling the computational box with N_s molecules of the solvent considered, the equilibration consisted of a steepest-descent energy minimization, followed by a 0.5 ps thermalization MD (NVT, progressively increased temperature) and by a 1 ns plain MD simulation (NVT for the artificial solvents, NPT for the physical solvents, constant temperature). From this point, configurations (along with the value of the biasing potential in the US sampling phases of the LEUS simulations) were written to file every 0.5 ps for subsequent analysis.

4.2.3 Analysis

The analysis of the simulations was performed in terms of: (i) free-energy maps $G(\phi, \psi)$ in the space of the glycosidic dihedral angles ϕ and ψ, along with corresponding relative free energies G_X of different states (conformational regions) and $\tilde{G}_X$ of the associated minima at locations $\tilde{\phi}_X$ and $\tilde{\psi}_X$; (ii) populations P_X of the different states; (iii) average number n_i of intramolecular H-bonds; (iv) average number n_s of solute-solvent H-bonds; (v) relative populations p of the canonical rotamers (gg, gt, tg) of the two hydroxymethyl groups (dihedral angles ω and ω') and of the three staggered conformers (g^+, t, g^-) of the exocyclic dihedral angle χ_2 and χ'_3. For the MD+LEUS simulations, all the thermodynamic

and structural quantities analyzed were calculated based on the configurations generated during the US sampling phase, with a reweighting factor depending on the value of the biasing potential associated with each configuration, as detailed elsewhere[14].

The free-energy maps $G(\phi, \psi)$ in the space of the glycosidic dihedral angles ϕ and ψ were calculated using a grid spacing $\Delta\phi = 2°$ and $\Delta\psi = 2°$, and anchored ($G = 0$ kJ mol^{-1}) at their global minimum. In the t_{LE}-LEUS and t_{US}-LEUS simulations, the value of G at grid points that were never visited during a 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. For all other simulations, this value was set to $G_{max} = 50$ kJ mol^{-1}.

For the ease of discussion, the free-energy maps were partitioned into three distinct conformational regions (A, B, and C) as illustrated in Figure 4.3, using cutoff values similar to those employed in Ref.[171] (although a different labelling is used here). The region C was further partitioned into three subregions (C_1, C_2 and C_3), using cutoff values chosen by visual inspection of the free-energy maps. Whenever a region encompassed a local minimum, the location of this minimum was determined by interpolation based on the nearest-neighbor grid points. Minima occurred in regions B, C_1, C_2 or/and C_3, but never in region A. The location and free-energy of a minimum occurring in region X are noted ($\tilde{\phi}_X$, $\tilde{\psi}_X$) and $\tilde{G}_X$, respectively. Alternatively, the relative free energies of the states were calculated based on the corresponding integrated populations. Only states B, C_1, C_2 or/and C_3 were populated, never A. The relative free energy of state X is noted G_X, anchored to zero for the most populated state. Corresponding relative populations P_X are also reported, expressed in percent.

The occurrence of intramolecular H-bonds was analyzed considering all hydroxyl groups as potential H-bond donors and all hydroxyl or ring oxygen atoms as potential acceptors. An intramolecular H-bond is assumed to be present when the distance between the hy-
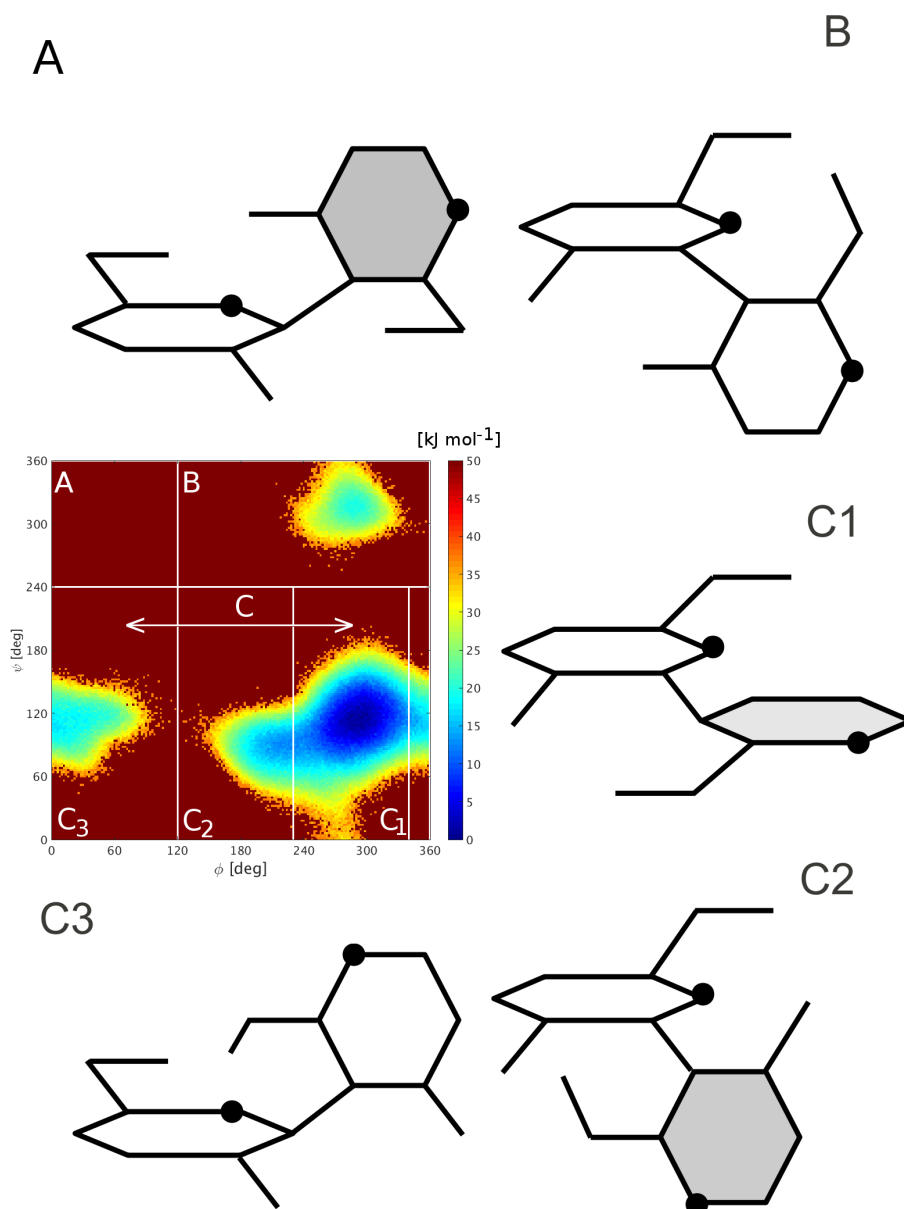
Figure 4.3: Definition of the five conformational regions in the Ramachandran (ϕ,ψ) maps of the disaccharides. The regions A, B, and C are defined by cutoff values of 120° and 240° for ϕ and ψ, respectively, according to Ref.[171] (with a different nomenclature: regions A, B and C correspond to the regions labelled B, D and A+C, respectively, in Ref.[171]). The region C is further partitioned in three subregions (C_1, C_2, and C_3), with cutoff values of 120°, 230° and 340° for ϕ. Note that region C_3 encompasses areas at the left and right of the map by periodicity. The regions are displayed superimposed to an illustrative free energy map corresponding to GG in water (first panel in Figure 4.6). Schematic representations of the ring orientations corresponding to the five regions are also shown. The darker shade indicates the α-face of the ring, the line pairs the hydroxymethyl groups, the single lines the hydroxyl groups at C_2 or C_3' and the black circles the ring oxygen atoms.

drogen and acceptor atoms is below 0.25 nm, and the angle between the donor, hydrogen and acceptor atoms is above $100°$. The use of a somewhat relaxed angular criterion[170] (a minimum angle of $135°$ is typically used instead[101]) is necessary to encompass H-bonds between the vicinal hydroxyl groups in the ring and between the hydroxymethyl group and the ring oxygen atom (see also Chapter 3).

The occurrence of solute-solvent H-bonds involving all the solute hydroxyl groups as well as the ring oxygen atom was also analyzed, using the normal angular criterion (minimum angle of $135°$). For the artificial solvent models involving a scaling of the oxygen-hydrogen bond length (scaling factor $s_b \neq 1.0$), the solute-solvent H-bond analysis of each trajectory configuration was preceded by a rescaling of the solvent oxygen-hydrogen distances as described in Chapter 2 and Chapter 3. This procedure involves a displacement of the hydrogen atoms of each solvent molecule along the oxygen-hydrogen bond vector, while keeping the oxygen position fixed, so that the value of 0.1 nm corresponding to the SPC water model is recovered.

Relative populations p of the canonical rotamers (gg, gt, tg) of the two hydroxymethyl groups (dihedral angles ω defined as C_4-C_5-C_6-O_6 and ω' defined as C_4'-C_5'-C_6'-O_6') and of the three staggered conformers (g^+, t, g^-) of the exocyclic dihedral angles χ_2 (C_1-C_2-O_2-H_2) and χ_3' (C_2'-C_3'-O_3'-H_3') were calculated from the probability distribution profiles around the corresponding four dihedral angles. For the two hydroxymethyl dihedral angles, values in the ranges $[0°; 120°[$, $[120°; 240°[$ and $[240°; 360°[$ define the relative populations of the gg, gt and tg rotamers, respectively. For the two exocyclic hydroxyl dihedral angles the corresponding ranges define the g^+, t and g^- rotamers.

4.3 Results

4.3.1 Convergence assessment

Biased (*i.e.* non-reweighted) probability distributions $P_N(\phi, \psi)$ in the space of the glycosidic dihedral angles ϕ and ψ along with free-energy maps $G(\phi, \psi)$ obtained from the $t_{US} = 60$ ns US phases of the six t_{LE}-LEUS simulations involving different build-up times t_{LE} are displayed in Figure 4.4 for the disaccharide GG in the solvent CCl_4. Similar results were obtained for the four other physical solvents considered (data not shown). These maps permit to assess the convergence of the results as a function of the time t_{LE} invested into the LE build-up of the memory-based biasing potential prior to the US sampling phase.

The comparison of the biased probability distributions P_N for the six runs reveals a large sampling enhancement upon increasing t_{LE} from 0 ns to 30 or 40 ns. The effect of increasing t_{LE} further to 50 ns is comparatively more limited. The free-energy maps calculated from the corresponding reweighted probability distributions show that all the relevant minima are already revealed with t_{LE} set to 10 ns. The maximal height above the global minimum up to which the free-energy surface is explored increases systematically with t_{LE}, from about 30 to about 50 kJ·mol^{-1}. The choice $t_{LE} = 50$ ns retained in the production simulations seems therefore to be adequate.

For $t_{LE} = 50$ ns, a convergence assessment was also performed in terms of the US sampling time t_{US} in the six t_{US}-LEUS simulations. The results are shown in Figure 4.5, again for GG in CCl_4. In this case, the sampling of the relevant region, particularly C, is almost completely performed in the first 20 ns of the US phase. The maximal height above the global minimum increases with t_{US}, from about 40 to about 50 kJ·mol^{-1}. The choice $t_{US} = 60$ ns retained in the production simulations was found sufficient.
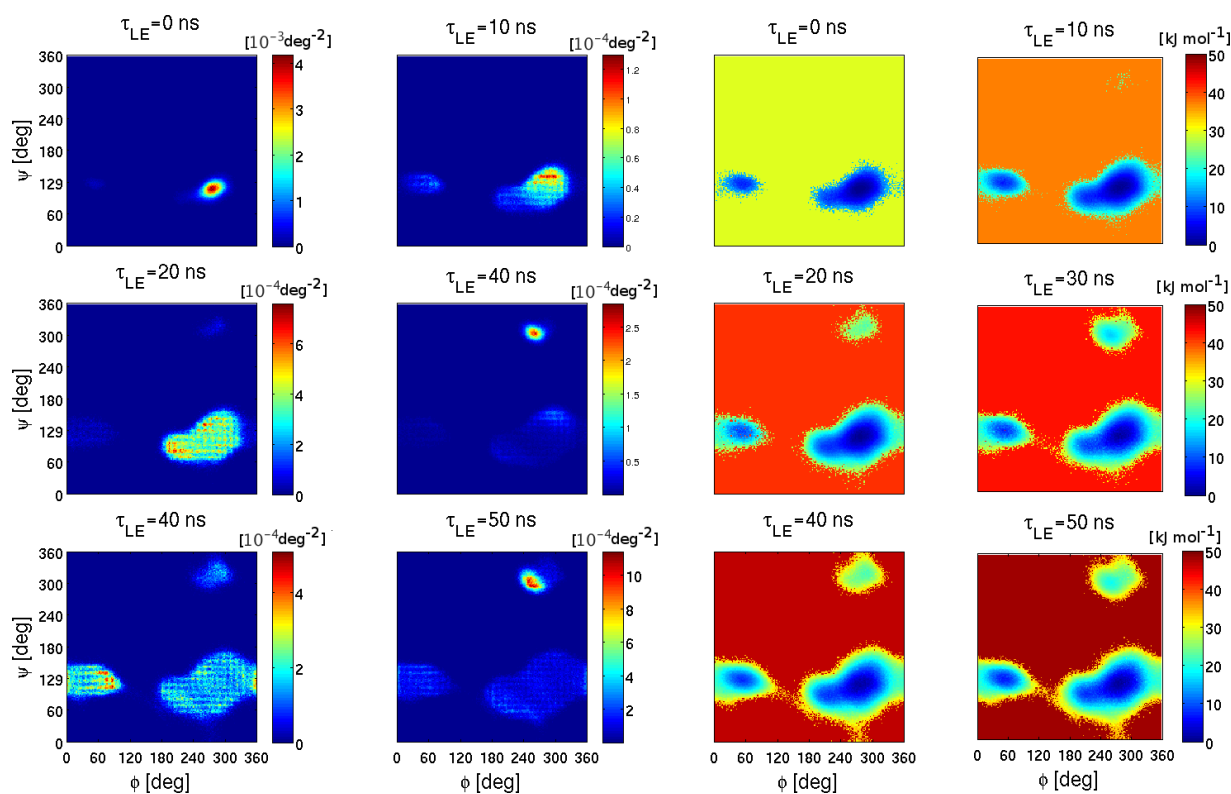
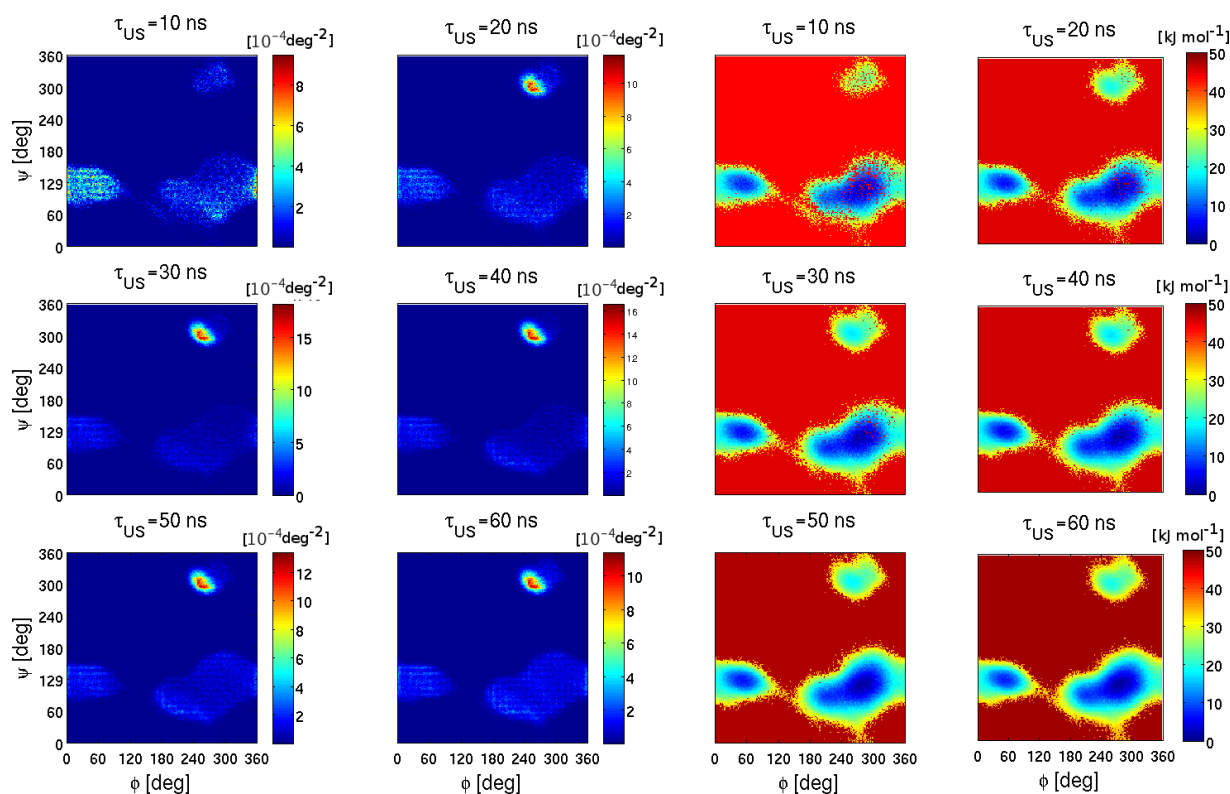**Figure 4.4:** [Next page] Convergence of the calculated biased probability distribution and free-energy map as a function of the LE build-up time. The results are shown for the compound GG (Figure 4.1) simulated in CCl_4 at 1 bar and 298.15 K. They are based on the $t_{US} = 60$ ns US sampling phases of LEUS simulations relying on a biasing potential built up in LE phases of different durations t_{LE}. The corresponding biased probability distributions (six panels on the left) and the free-energy maps (six panels on the right) in the space of the glycosidic dihedral angles ϕ and ψ are displayed. The free-energy maps are anchored to $G = 0$ kJ mol^{-1} at the location of their global minimum, and the value of G at grid points that were never visited during the simulation is arbitrarily set (for each panel separately) to the maximal value G_{max} of G over all grid points that were visited at least once.

Figure 4.5: Convergence of the calculated biased probability distribution and free-energy map as a function of the US build-up time. The results are shown for the compound GG (Figure 4.1) simulated in CCl_4 at 1 bar and 298.15 K. They are based on the $t_{US} = 10$, 20, 30, 40, 50, 60 ns US sampling phases of LEUS simulations relying on a biasing potential built up in $t_{LE} = 50$ ns LE phases. The corresponding biased probability distributions (six panels on the left) and the free-energy maps (six panels on the right) in the space of the glycosidic dihedral angles ϕ and ψ are displayed. The free-energy maps are anchored to $G = 0$ kJ mol^{-1} at the location of their global minimum, and the value of G at grid points that were never visited during the simulation is arbitrarily set (for each panel separately) to the maximal value G_{max} of G over all grid points that were visited at least once.

4.3.2 Conformational analysis in the physical solvents

Illustrative free-energy maps G(ϕ, ψ) for the four disaccharides considered obtained from the LEUS simulations (t_{LE} = 50 ns, t_{US} = 60 ns, after reweighting) in the physical solvents with highest and lowest polarities, namely H_2O and CCl_4, are shown in Figure 4.6. The other physical solvents represent intermediate situations between these two extremes (data not shown). The maps obtained in CCl_4 can also be compared with maps obtained from 1 μs SD simulations in vacuum (see Suppl. Mat. Figure 4.S.1). In the absence of LEUS sampling enhancement, the sampling in vacuum is less broad compared to the LEUS simulations and the maximum height above the global minimum reached in these simulations is about 30 kJ·mol^{-1}. Besides that, the maps are very similar.

The relative populations P_X of the different conformational regions X as defined in Figure 4.3 and the corresponding relative free energies G_X are displayed in Figure 4.7 for the entire series of physical solvents. The data can also be found numerically in Suppl. Mat. Tables 4.S.1 and 4.S.2, also including the heights $\tilde{G}_X$ and the locations ($\tilde{\phi}_X$, $\tilde{\psi}_X$) of the minima. To facilitate the interpretation of the conformational regions in terms of structures, illustrative schematic models of the disaccharides in the five regions considered are also shown in Figure 4.3. Note that these sketches are only meant to give a pictorial representation of the relative orientations of the two monosaccharide rings and of the proximity of specific exocyclic groups, and not a precise description of the molecular structure.

The two following subsections discuss in turn the conformational properties of the four disaccharides in water, and the changes observed when considering the four other (less polar) physical solvents.
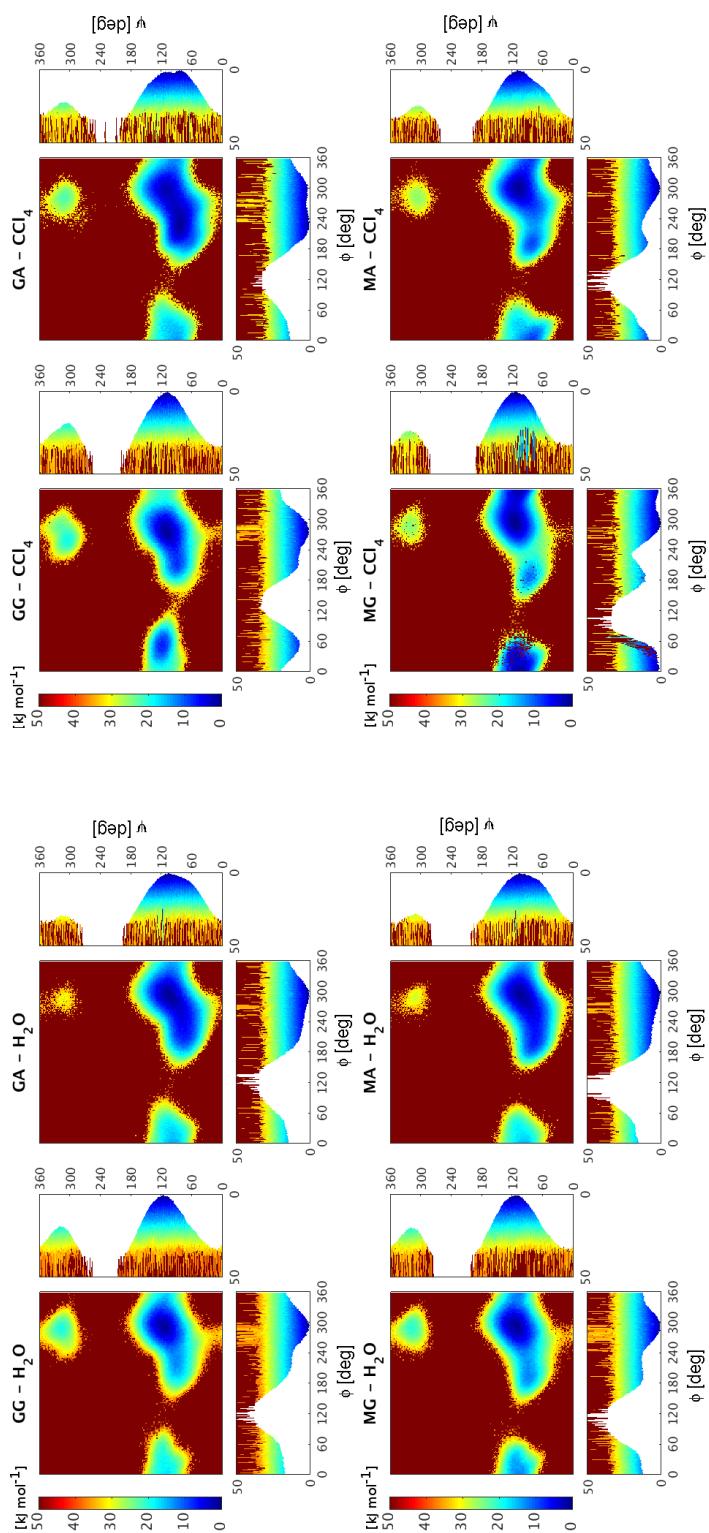
(a)

(b)

Figure 4.6: Free-energy maps $G(\phi, \psi)$ in the space of the glycosidic dihedral angles ϕ and ψ for the four disaccharides considered (Figure 4.1) in H_2O or CCl_4 at 1 bar and 298.15 K. The maps are based on the 60 ns sampling phases of the corresponding LEUS simulations in water (H_2O, four panels (a)) and in carbon-tetrachloride (CCl_4, four panels (b)). The maps are anchored to $G = 0$ 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 $G_{max} = 50$ kJ mol^{-1}. The maps are shown in the ϕ,ψ-plane as viewed from the top, along with lateral projections in the ϕ,G- and ψ,G-planes.

Conformational analysis in water

Visual inspection of the two-dimensional free energy maps for the four disaccharides in H_2O (Figure 4.6, panels (a)) reveals the following main features.

As was observed in previous work on the glucose-based disaccharides[132,171], large areas of the maps remain unexplored because they correspond to configurations involving steric clashes between the atoms of the two residues. All maps are characterized by a lowest free-energy basin in region C, along with one alternative metastable state in region B.

The global minimum is located for all the disaccharides in region C_1 at about $(\tilde{\phi}_{C_1},$ $\tilde{\psi}_{C_1})$=(290±5°, 111±5°), see Table 4.S.2. These values are compatible with the available experimental estimates for cellobiose (GG) in water as inferred from nuclear magnetic resonance (NMR) data[327], namely (272°, 99°), or in the crystal form as inferred from X-ray crystallographic data[306,322], namely (284°, 108°). They are also close to the values found in a previous simulation study of the same disaccharides with the GROMOS 45A4 force field[132] and with values typically found using other recent empirical force fields[328]. The region C_1 presents in all cases the largest population P_{C_1}, about 98% for GG and MG or about 93% for GA and MA, see Table 4.S.1.

Alternative low free-energy wells correspond to different ϕ (0-60° in region C_3, 180-220° in region C_2) or ψ (300-320° in region B) values. These alternative wells are associated with populations of at most about 7%. One should distinguish between wells that correspond to a local minimum and wells that can be seen as a broadening of the global minimum in region C_1 (without a separating free-energy barrier). The region B is associated with marginal populations (at most 0.4%) and represents a local minimum at about 19-20 kJ·mol^{-1} for GG and MG or 27-28 kJ·mol^{-1} for GA and MA. The regions C_2 and C_3, associated with small populations of about 1% or less for GG and MG or up to about 7% for GA and MA, represent broadenings of the global minimum at C_1. They present

free-energy values of 11-13 kJ·mol^{-1} except region C_2 for GA and MA, with values around 4-5 kJ·mol^{-1}.

These results confirm the observations made in Ref.[132]. The epimerization has only a very limited influence on the Ramachandran free-energy map of the GG disaccharide. More precisely, the epimerization at C_2 has no visible effect (no difference between GG and MG), while the epimerization at C_3' (difference from GG to GA) only broadens the C_1 basin in the direction of lower ϕ values (also slightly lower ψ), *i.e.* towards region C_2.

Conformational analysis in the other physical solvents.

The effect of drastically decreasing the polarity of the solvent can be seen by comparing the two-dimensional free energy maps of the four disaccharides in CCl$_4$ (Figure 4.6, panels (b)) with the maps in H$_2$O (top panels). The progressive trends along the series of five physical solvents of decreasing polarity can be followed in Figure 4.7 (numerical values in Suppl. Mat. Tables 4.S.1 and 4.S.2).

As observed in the case of H$_2$O, large areas of the maps remain unexplored in CCl$_4$. The conformational regions explored in the two solvents are similar. All maps are still characterized by a lowest free-energy basin in region C, along with an alternative metastable state in region B. Note that the area sampled in the latter region for GA and MA is noticeably broader in CCl$_4$ compared to H$_2$O. The subregion C_1 of C is still the most populated, but its population may decrease to 60% and, in contrast to H$_2$O, the subregions C_2 and C_3 may become alternative metastable states (true local minima).

In the case of GA and MA, a local minimum appears in subregion C_2, located at ($\tilde{\phi}_{C_2}$, $\tilde{\psi}_{C_2}$)=(232±5°, 80±5°) for GA and (194±5°, 80±5°) for MA. In the case of GA this minimum becomes the most stable state (global minimum), with the one in C_1 presenting a slightly higher free-energy value of about 1 kJ·mol^{-1}. The two minima are separated by

a barrier of about 2 kJ·mol^{-1}. In the case of MA the global minimum remains in C_1 and becomes narrower, while the minimum in C_2 is at 8 kJ·mol^{-1}.

Along the series of physical solvents in order of decreasing polarity (Figure 4.7), the region C_2 for GA is associated with an increasing population, from 7% in H_2O to 24% in CCl_4, with a simultaneous decrease in the population of C_1, from 93% in H_2O to 76% in CCl_4. The population of C_3 remains below 0.6% in all cases. The region C_2 for MA shows populations that change little upon decreasing the polarity of the solvent, remaining below 7% for all the physical solvents considered, while the single overwhelmingly populated state is the one in the region C_1 (relative population above 90% in all cases). For this disaccharide, a slight increase in the population of the region C_3 can be noticed, associated with a lowering of the free energy of that region (no formation of a real minimum can be observed in this area).

In the case of GG and MG, a second minimum appears in C_3. The minimum is located at $(\tilde{\phi}_{C_3}, \tilde{\psi}_{C_3})$=(54±5°, 118±5°) for GG and (6±5°, 106±5°) for MG. In the case of GG, this local minimum is separated from the global minimum by a very high barrier of 25 kJ·mol^{-1} (relative to the global minimum) and it presents a free energy of about 6 kJ·mol^{-1}. In the case of MG, the second minimum is at 0.5 kJ·mol^{-1} and is separated from the global minimum by a very low barrier of about 2 kJ·mol^{-1}. In addition, MG presents a third minimum, located in region C_2 at about 10 kJ·mol^{-1}, with a higher barrier of 20 kJ·mol^{-1} (relative to the global minimum). In contrast, for GG, the region C_2 represent a broadening of the global minimum (no local minimum).

Along the series of physical solvents in order of decreasing polarity (Figure 4.7), for GG, the populations of the region C_2 and C_3 both increase slightly from about 0.5% in H_2O to about 4% in CCl_4. The population of C_1 remains above 92% in all solvents. For MG, the region C_3 is associated with an increasing population upon decreasing the polarity of the
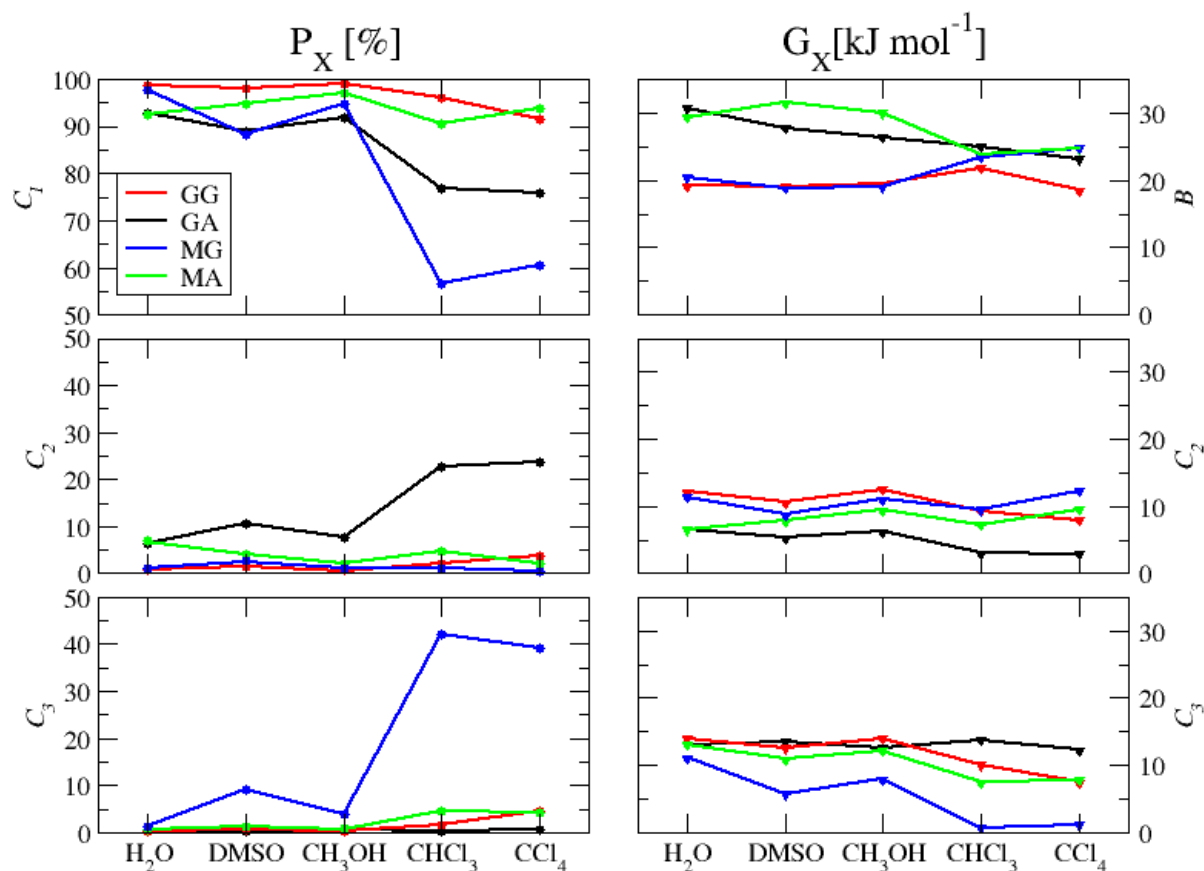
Figure 4.7: Populations and relative free energies of specific regions (Figure 4.3) in the Ramachandran free-energy maps for the four disaccharides considered (Figure 4.1) along the series of pysical solvents at 298.15 K and 1 bar. For a subset of the six regions X of the Ramachandran map, the populations P_X (three panels on the left) and the relative free energies G_X (three panels on the right) are displayed along the solvent series. The populations P_X are calculated by integrating the reweighted probability distributions in the space of the glycosidic dihedral angles ϕ and ψ over the selected regions. The corresponding free energies are reported relative to the region with the lowest value (in general in the region C_1, for this reason omitted). The region A has zero population and is therefore also omitted. The data are based on the 60 ns sampling phases of the corresponding LEUS simulations. The data points follow the series of physical solvents (Table 4.1) in order of decreasing polarity. The different panels (top to bottom) correspond to the different regions. The regions C_1, C_2, and C_3 are considered for the populations, while the regions B, C_2 and C_3 are considered for the relative free energies. The data is reported numerically in Suppl. Mat. Tables 4.S.1 and 4.S.2. Full free-energy maps for H_2O and CCl_4 are displayed in Figure 4.6.

solvent, from 1% in H_2O to about 40% in $CHCl_3$ and CCl_4. Accordingly, the population of the region C_1 decreases from 98% in H_2O to 60% in $CHCl_3$ and CCl_4. The population of the region C_2 remains constant and very low, despite the formation of a third minimum in that area.

As mentioned in Section 4.3.2, there is no dramatic differences between the four disaccharides in water, in line with observations made in previous work on either the same or similar disaccharides[132,171]. However, significant conformational changes occur in solvents of lower polarity. These can be summarized as follows. When the hydroxyl group at C_3' is equatorial (GG and MG), a local minimum appears in the region C_3. This minimum represents a metastable state when the hydroxyl group at C_2 is equatorial (GG), whereas it becomes a stable state comparable with the global minimum when this the hydroxyl group is axial (MG). When the hydroxyl group at C_3' is axial (GA and MA), a second minimum appears in region C_2. This minimum represents a metastable state when the hydroxyl group at C_2 is axial (MA), whereas it becomes a stable state comparable with the global minimum when this hydroxyl group is equatorial (GA).

4.3.3   Hydrogen bonding analysis in the physical solvents

The analysis of the H-bonds was performed considering all hydroxyl groups as potential H-bond donors and all hydroxyl or ring oxygen atoms as potential acceptors, and is reported in terms of average number n between specific potentially H-bonding partners. Note that when a single H-bond is considered, n can be viewed as a relative occurrence of the H-bond. The numbers n_i of intramolecular H-bonds and the number n_s of H-bonds with the solvent are considered separately. The intramolecular H-bonds analysis also distinguishes between numbers n_{ir} of intraresidue (*i.e.* within the same ring) and n_{it} of interresidue (*i.e.* *trans*-glycosidic) H-bonds. The former are only significant when the angular criterion used

to define the presence of an H-bond is relaxed[170] (from 135° to 100°, see Section 4.2.3). In addition, for these intraresidue H-bonds, the flip-flop alternatives are also listed as an indicator of the presence of either a clockwise or a counterclockwise (as viewed from the β-side of the ring) H-bond network (pattern 4→3→2→1 or pattern 1→2→3→4).

The H-bond analysis is provided in full details in Suppl. Mat., graphically in Figures 4.S.2-4.S.4 and numerically in Table 4.S.3. For a further analysis of the H-bonding patterns, additional plain MD simulations, labelled MD-X (see Section 4.2.2), were performed restraining the ϕ and ψ dihedral angles in the four region of Figure 4.3 (B, C, C_1, C_2 or C_3). The data regarding the H-bonding in the specific regions of Ramachandran maps refer to these simulations, unless otherwise specified. To facilitate the discussion, a subset of all these results is presented synoptically in Figures 4.8-4.11, along with Table 4.3.

Figure 4.8 gives a summary of the H-bond analysis, showing graphically the total number n_{it} of interresidue H-bonds, the total number n_{ir} of intraresidue H-bonds and the total number n_s of H-bonds with the solvent. Figure 4.9 provides a more detailed analysis of the interresidue H-bonds in the four regions of Ramachandran map, together with information on the conformations of the four exocyclic dihedral angles compatible with these H-bonds.

A preliminary analysis revealed that six possible interresidue H-bonds can occur, namely $H'_3 \rightarrow O_5$, $O_6 \leftrightarrow O'_3$, $O_2 \leftrightarrow O'_6$, $H'_6 \rightarrow O_5$, $O_6 \leftrightarrow O'_6$ and $O_2 \leftrightarrow O'_3$, the double arrow indicating the possibility of two filp-flop alternatives. These H-bonds are compatible with different possible conformations of the ϕ and ψ dihedral angles, and involve specific orientations of the four exocyclic dihedral angles ω, ω', χ_2 and χ'_3. To better understand this correlation, disaccharide structures were generated varying systematically the ϕ and ψ dihedral angles by increments of 10°, and considering the three standard orientations (gg, gt, tg) of the hydroxymethyl groups (dihedral angles ω and ω') and of the hydroxyl groups at C_2 and C'_3 (dihedral angles χ_2 and χ'_3). Figure 4.10 shows the results of this analysis. Additionally,

the list of all possible interresidue H-bonds for the different regions of the Ramachandran map is given in Table 4.3. Finally, the occurring interresidue H-bonds are also shown in illustrative structures presented in Figures 4.11.

Hydrogen bonding analysis in water

Figure 4.8 shows the total number of H-bonds for the four disaccharides considered, distinguishing between H-bonds with the solvent (dark green), intraresidue H-bonds (orange) and interresidue H-bonds (gray). A rectangle with a color code consistent with that in Figure 4.7 also identifies each disaccharide. The values for the simulations in H_2O correspond to the set of four bars on the left.

In water, the average total number of H-bonds is about 17.5 (GA, MA) or 18 (GG, MG) for all the disaccharides considered. The disaccharides have several potentially H-bonding sites, namely six hydroxyl and two hydroxymethyl groups (that can be both donors and acceptors), plus two ring and one glycosidic oxygen atoms (only acceptors). On average, each site participates in 1.5 H-bonds. However, the distribution of these 17.5 or 18 H-bonds among the three categories is different for the different compounds. For GG and MG, one has about 15 H-bonds with the solvent, 2 intraresidue H-bonds and 1 interresidue H-bond. The epimerization at C_2 does not affect this pattern. On the other hand, epimerization at C'_3 (GA and MA) leads to a slightly different balance in the distribution of H-bonds. When being in an axial position (MA), the hydroxyl group at C'_3 is more solvent exposed and the number of H-bonds with the solvent increases to about 16. At the same time, the number of intraresidue H-bonds is decreased to about 1.5, and that of interresidue H-bonds becomes nearly zero.

As can be seen in Figure 4.10, the typical interresidue H-bond present in the dominant conformations at equilibrium for GG and MG is $H'_3 \rightarrow O_5$. The epimerization at C'_3 clearly
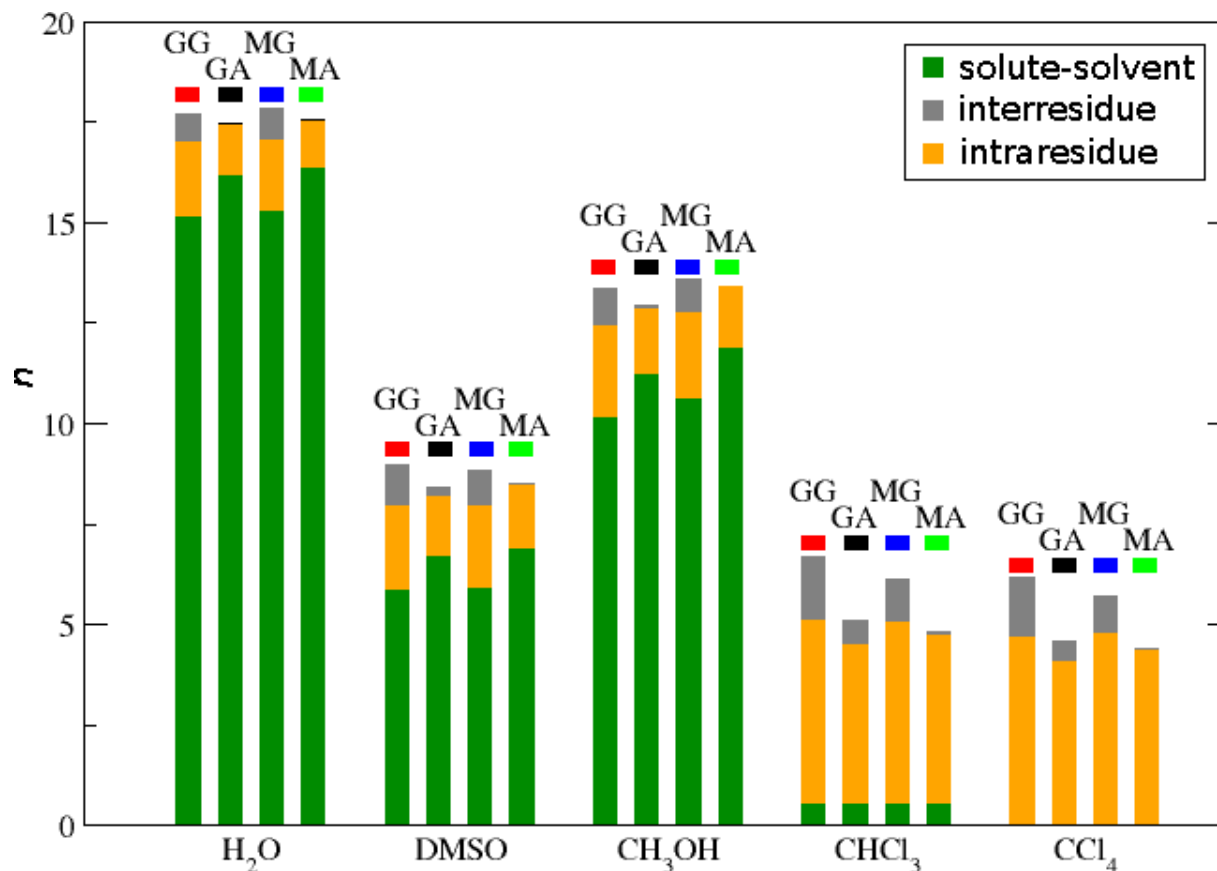
Figure 4.8: Numbers of H-bonds for the four disaccharides considered in the physical solvents at 298.15 K and 1 bar. The results are based on the 60 ns sampling phase of the corresponding LEUS simulations. Each bar represents the total number n of H-bonds, with different colors distinguishing interresidue H-bonds, intraresidue H-bonds, and solute-solvent H-bonds. The different bars for each solvent refer to one of the four disaccharides (Figure 4.1), that are identified by a box with the same color code as in Figure 4.7. The bars follow the series of the physical solvents (4.1) in order of decreasing polarity. The data is reported numerically in Suppl. Mat. Table 4.S.3

affects the feasibility of such a H-bond. The most relevant alternative to this H-bond for GG and GA is $H'_2 \rightarrow O_6$, compatible with the conformations in the region C_1. Note that this H-bond is no longer feasible for MG and MA in this region, due to the epimerization at C'_3. As already observed in previous work on either the same or similar disaccharides[132,171], the intramolecular H-bonding partners are in a highly polar solvent (water), and their interaction is screened by the solvent dielectric response as well as subject to H-bonding competition by the solvent molecules. The lack of a specific intramolecular H-bond is

easily compensated for by the interaction of the potential H-bonding partners with the solvent molecules. This compensation prevents large conformational changes that would be necessary to recover the lost intramolecular H-bond in the case of GA and MA, as shown in the free energy maps discussed in the previous sections.
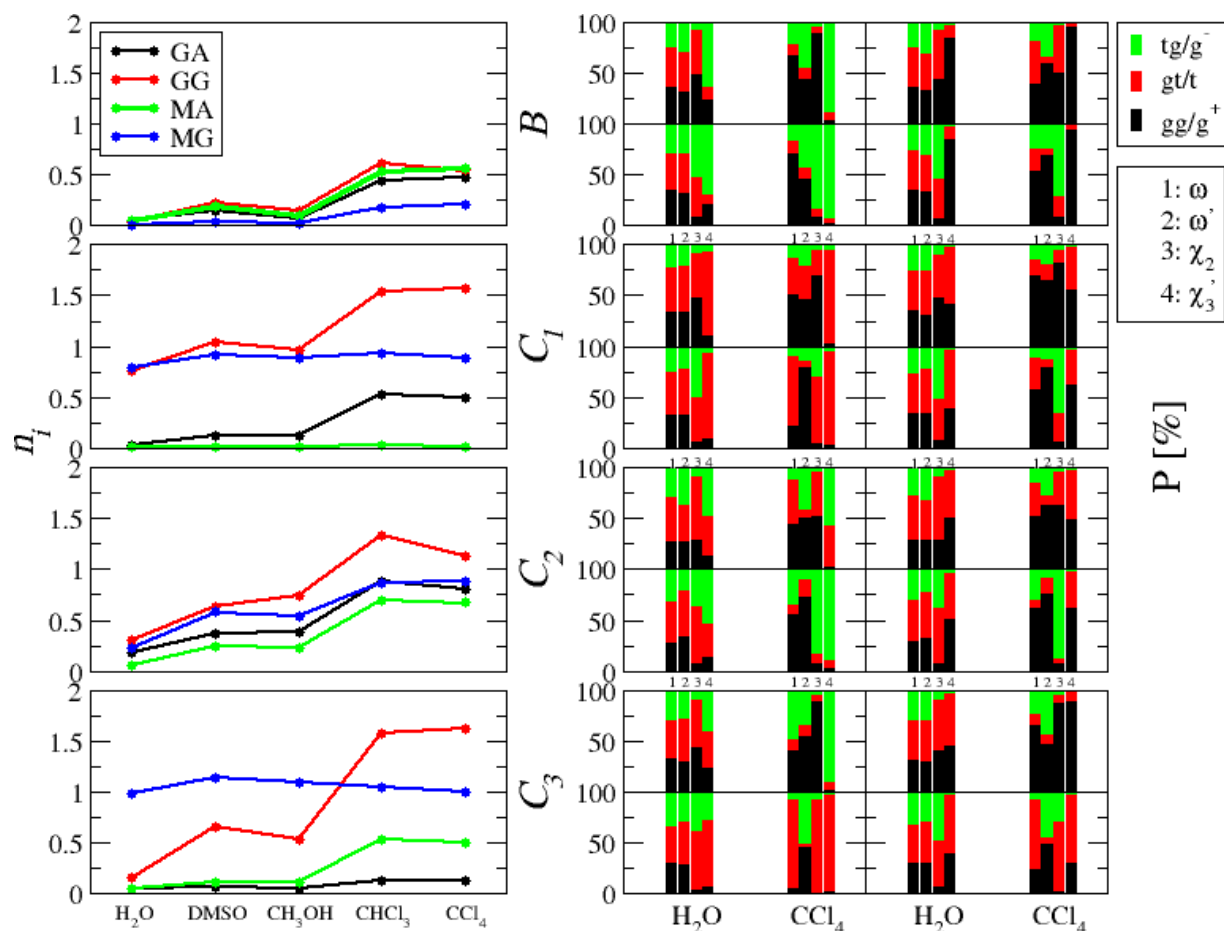
Figure 4.9: Numbers of interresidue H-bonds and exocyclic rotamer populations for the four disaccharides considered in the physical solvents at 298.15 K and 1 bar. The results are based on plain MD simulations restrained to four regions (B, C_1, C_2 or C_3) in the space of the glycosidic dihedral angles ϕ and ψ (Figure 4.3). The graphs on the left show the total numbers n_i of interresidue H-bonds. The colors refer to the four disaccharides (GG, MG, GA and MA). The graphs on the right show the relative populations p (in percent) of three canonical rotamers (gg, gt, tg) of the two hydroxymethyl groups (dihedral angles ω and ω') and of the three staggered conformers (g^+, t, g^-) of the exocyclic dihedral angle χ_2 and χ_3' considering the solvents H_2O and CCl_4 only. The bars correspond to the relative populations of ω (1), ω' (2), χ_2 (3) and χ_3' (4) and the colors refers to the three possible conformations of the dihedral angles (gg or g^+, gt or t and tg or g^-).

Hydrogen bonding analysis in the other physical solvents

The effect of a decrease in the polarity of the solvent on the H-bonding pattern can be see in Figure 4.8. An increase in the number of intramolecular H-bonds compensates for the lower H-bonding propensity of the solvent. In $CHCl_3$ and CCl_4 the average number of H-bonds with the solvent decreases to values close to zero and the average number of intraresidue H-bonds increases to about 7 (GG), 6 (MG) or 4.5 (GA, MA). Note that the average number of H-bonds with DMSO is lower than the one with CH_3OH in spite of the comparable dielectric permittivities of these two solvents (Table 4.1), because DMSO is only an H-bond acceptor whereas CH_3OH is both a donor and an acceptor

Differences between the four disaccharides considered in terms of intramolecular H-bonding become pronounced in low polarity solvents. In $CHCl_3$ and CCl_4, GG shows the highest number of both intraresidue (about 5) and interresidue (about 2) H-bonds. The epimerization at C_2 affects the interresidue H-bond formation, lowering the average number to 1 for MG. The effect of the epimerization at C'_3 is even more pronounced, interfering with both types of intramolecular H-bonds. For both GA and MA the total number of intraresidue H-bonds is about 4.5. In addition, the double epimerization in MA also inhibits almost completely the formation of interresidue H-bonds.

The conformational changes described in the Section 4.3.2 upon decreasing the solvent polarity can be explained by looking at Figure 4.10 and Table 4.3. This figure displays the areas where each of the six possible interresidue H-bonds are possible for the four different disaccharides. This permit to assess the compatibility between stable (or metastable) states and the formation of normal (or alternative) H-bonds. The following patterns can be identified. First, regions that are stabilized upon decreasing the solvent polarity (lowering of the free energy and formation of a barrier to other states) are compatible with one specific H-bond. This is the case for the minimum at C_2 in MG and MA, compatible with

Comp.	Conf.	H-bonds					
		H'$_3$→O$_5$	O$_6$↔O'$_3$	O$_2$↔O'$_6$	H'$_6$→O$_5$	O$_6$↔O'$_6$	O$_2$↔O'$_3$
GG	B				$\cdot\ g^-\ \cdot\ \cdot$	$t\ \substack{g^- \\ g^+}\ \cdot\ \cdot$	$\cdot\ \cdot\ \substack{g^+ \\ g^-}\ \substack{g^+ \\ t}$
	C_1	$\cdot\ \cdot\ \cdot\ t$	$t\ \cdot\ \substack{t \\ g^+}\ \cdot$	$\cdot\ g^-\ g^+\ \cdot$	$\cdot\ g^+\ \cdot\ \cdot$	$t\ g^+\ \cdot\ \cdot$	
	C_2	$*$		$\cdot\ g^-\ \substack{g^+ \\ g^-}\ \cdot$			
	C_3				$\cdot\ \substack{g^+ \\ g^-}\ \cdot\ \cdot$	$\substack{g^+ \\ t}\ \substack{g^+ \\ g^-}\ \cdot\ \cdot$	$\cdot\ \cdot\ \substack{g^+ \\ g^-}\ \substack{g^+ \\ t}$
GA	B	$\cdot\ \cdot\ \cdot\ t$			$\cdot\ g^-\ \cdot\ \cdot$	$t\ \substack{g^- \\ g^+}\ \cdot\ \cdot$	
	C_1	$*$	$*$	$\cdot\ g^-\ g^+\ \cdot$			
	C_2	$*$		$\cdot\ g^-\ g^+\ \cdot$			
	C_3				$\cdot\ \substack{g^+ \\ g^-}\ \cdot\ \cdot$	$\substack{g^+ \\ t}\ \substack{g^+ \\ g^-}\ \cdot\ \cdot$	
MG	B			$\cdot\ g^-\ \substack{g^+ \\ g^-}\ \cdot$	$\cdot\ g^-\ \cdot\ \cdot$	$t\ \substack{g^+ \\ g^-}\ \cdot\ \cdot$	
	C_1	$\cdot\ \cdot\ \cdot\ t$	$t\ \cdot\ \substack{t \\ g^+}\ \cdot$				
	C_2			$\cdot\ g^-\ \substack{g^- \\ g^+}\ \cdot$	$\cdot\ g^+\ \cdot\ \cdot$		
	C_3	$*$			$\cdot\ \substack{g^+ \\ g^-}\ \cdot\ \cdot$	$\substack{g^+ \\ t}\ \substack{g^+ \\ g^-}\ \cdot\ \cdot$	$\cdot\ \cdot\ \substack{g^+ \\ g^-}\ \substack{g^+ \\ t}$
MA	B	$\cdot\ \cdot\ \cdot\ t$			$\cdot\ \substack{g^+ \\ g^-}\ \cdot\ \cdot$	$t\ \substack{g^+ \\ g^-}\ \cdot\ \cdot$	
	C_1	$*$	$*$				
	C_2	$*$		$\cdot\ g^-\ \substack{g^- \\ g^+}\ \cdot$			
	C_3				$\cdot\ \substack{g^+ \\ g^-}\ \cdot\ \cdot$	$t\ \substack{g^+ \\ g^-}\ \cdot\ \cdot$	$\cdot\ \cdot\ \substack{g^+ \\ g^-}\ \substack{t \\ g^+}$

Table 4.3: Interresidue H-bonds compatible with specific areas of the Ramachandran map given specific exocyclic group orientations for the four disaccharides considered. The feasibility of the six possible interresidue H-bonds (either of the two flip-flop possibilities considered together if possible) is indicated for the four region of the (ϕ,ψ) conformational space (Figure 4.3). For each possible H-bond, the implied conformation of the four exocyclic angles ω, ω', χ_2 and χ'_3 is given. When a certain angle is not relevant for the formation of a given H-bond, a dot is inserted. Two conformations for a pair of angles refer to the two alternative flip-flop H-bonds, while two conformations for a single angle refer to two possible alternative. A star indicates denoted a possibility that implies a larger deformation of the (ϕ,ψ) conformation relative to the values representative of the free-energy minimum in the specific region.

an alternative O$_2$↔O$'_6$ H-bond. Second, regions that become comparable to the global minimum (lowest free-energy) are compatible with two or more specific H-bonds. This is the case for the minimum at C_1 in GG, compatible with both H$'_3$↔O$_5$ and O$_2$↔O$'_6$, or the minimum at C_3 in MG, compatible with H$'_6$→O$_5$, O$_6$↔O$'_6$ and O$_2$↔O$'_3$. Third, regions that show a lowering of the free energy but not the formation of an alternative minimum are either compatible with two or more alternative H-bonds that are mutually excluding,
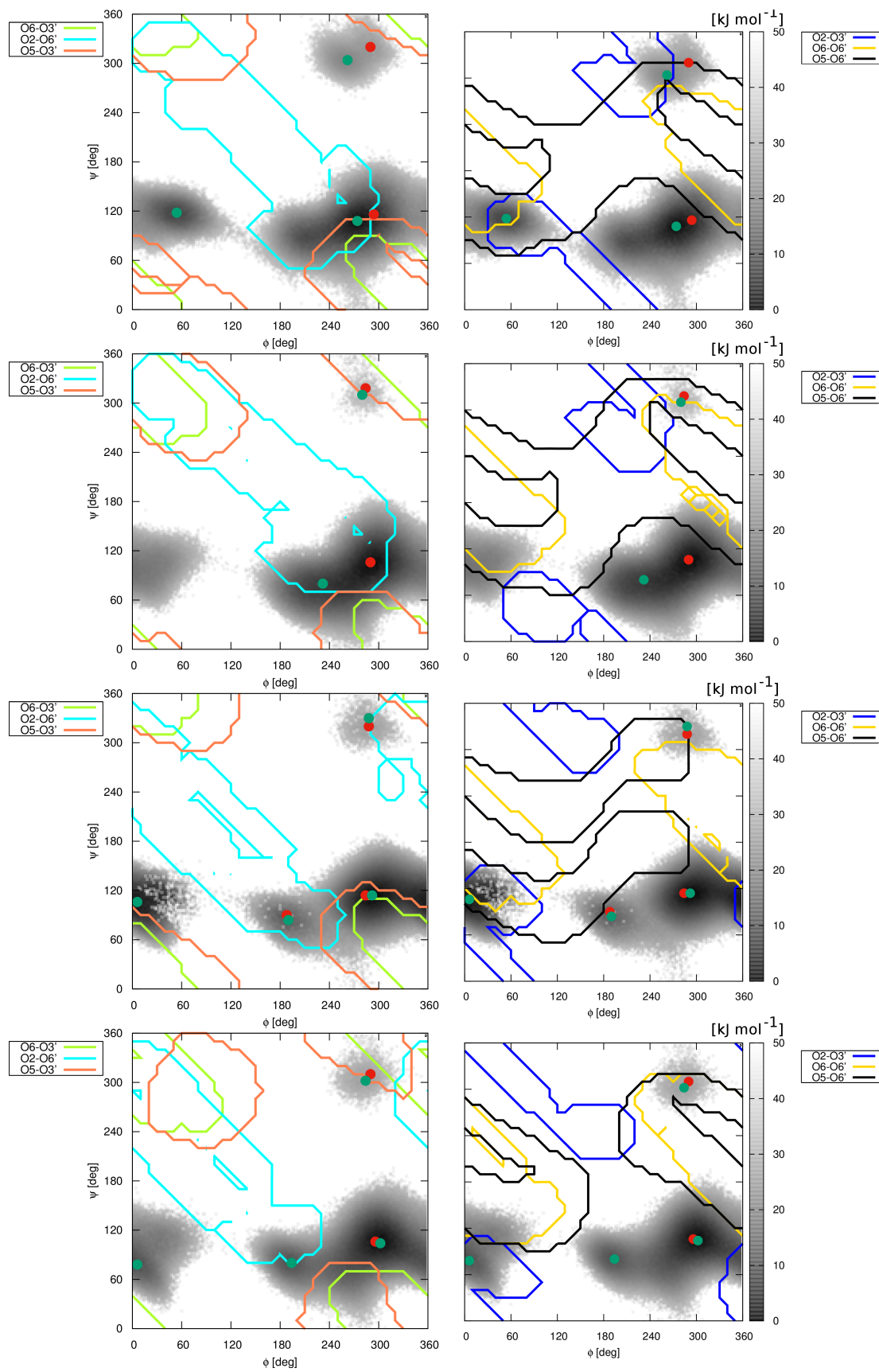
Figure 4.10: [Previous page] Interresidue H-bonds compatible with specific configurations of the glycosidic linkage for the four disaccharides considered. The contour of the areas compatible with a specific interresidue H-bond are displayed in the space of the glycosidic dihedral angles ϕ and ψ for the four disaccharides (Figure 4.1). They are superimposed to the free energy maps $G(\phi, \psi)$ obtained by LEUS simulations in CCl_4 at 298.15 K and 1 bar (see Figure4.6). The analysis is based on model conformations generated varying systematically the ϕ and ψ dihedral angles by increments of 10° and considering the three standard orientations (gg, gt, tg) of the hydroxymethyl groups (dihedral angles ω and ω') and the three standard orientations (g^+, t, g^-) of the hydroxyl groups at C_2 and C_3' (dihedral angles χ_2 and χ_3'). The dots identify the local free-energy minima in H_2O (red) and in CCl_4 (green). The different panels correspond to the four disaccharides, GG, GA, MG and MA (top to bottom). The six possible interresidue flip-flop H-bonds are illustrated, distringuishing between the H-bonds compatible with a conformation of the two rings with the opposite orientation of the hydroxymethyl groups (left panels) and with the same orientation of the hydroxymethyl groups (right panels).

or partially compatible with different H-bonds given a slight shift of the ϕ and ψ dihedral angles. This is the case for the region B of MG, compatible with either $O_6 \leftrightarrow O_6'$ or $O_2 \leftrightarrow O_6'$ (two H-bonds that cannot occur in the same conformation) or region C_3 of MA, compatible with $H_6' \rightarrow O_5$, $O_6 \leftrightarrow O_6'$ and $O_2 \leftrightarrow O_3'$ (but with slightly different values of ϕ and ψ).

Based on this analysis, it can be suggested that with a system as complex as a disaccharide, the effect of H-bonding in low-polarity solvents is either cooperative (two or more H-bonds can strongly stabilize a conformational change, mutually enhancing their effect), or stabilizing (one H-bond can stabilize a metastable state), or adverse (two or more H-bonds compatible with slightly different conformations belonging to the same region but that cannot be formed simultaneously within a single conformation can destabilize each other and the specific conformational region they are both compatible with).
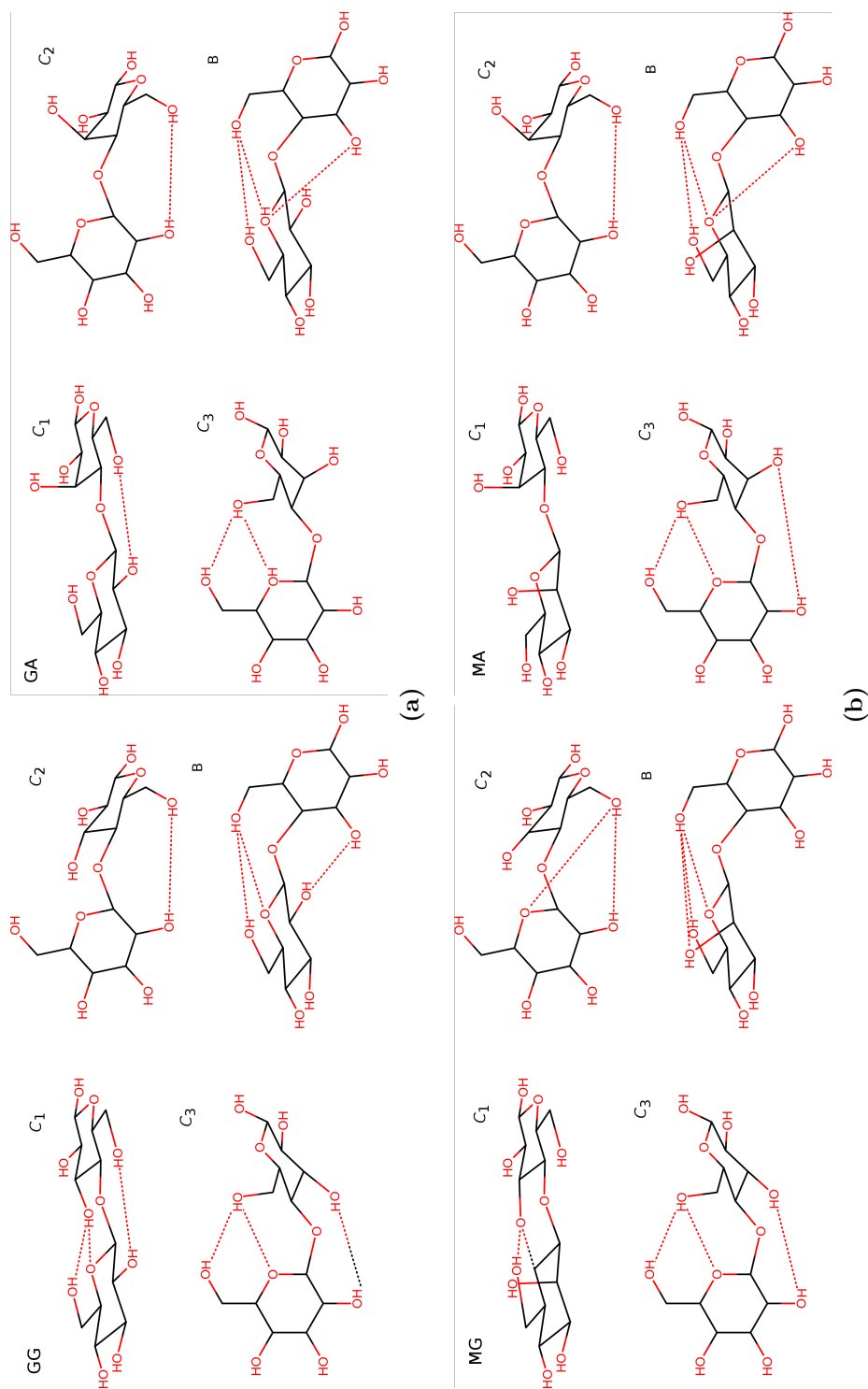
Figure 4.11: Illustrative structures of the four disaccharides in the four regions considered (Figure 4.3). The configuration of the glycosidic linkages displayed is representative of the relative orientation of the two residues for each region. The dotted lines represent possible interresidue H-bonds based on the analysis of Figure 4.10 (see also Table 4.S.2 for a complete description).

4.3.4   Simulations in artificial solvents

The four disaccharides were also simulated in artificial solvents derived from the SPC water[136] model by systematic variation of the oxygen-hydrogen bond length and of the partial charges, to allow for a separate modulation of the dielectric permittivity and H-bonding capacity, while preserving water-like dispersive interactions as well as molecular size and shape (Chapter 2).

The results of these simulations are illustrated in Figure 4.12 which shows the fractional populations and the relative free energies of the different regions of the Ramachandran map. The corresponding numerical values can be found in Suppl. Mat. Tables 4.S.4 and 4.S.5 for the populations, and Tables 4.S.6 and 4.S.7 for the relative free energies. The H-bonding analysis is illustrated in Figure 4.13. More details are shown in Suppl. Mat. Figures 4.S.5, 4.S.6 and 4.S.7 and the corresponding numerical values are reported in Tables 4.S.8, 4.S.9, 4.S.10 and 4.S.11.

In the series S_p^H, the dielectric permittivity of the solvent is decreased at water-like H-bonding capacity. A slight stabilization of the global minimum in C_1 can be noticed for the four disaccharides, with an increase of the population of this region above 95% for all disaccharides. Additionally, one observes an increase in the relative free energies of the regions C_2 and C_3, most pronounced for GG and MG. For GA and MA this results also in a noticeable decrease of the populations of region C_2. The H-bonding patterns follow these tendencies, with an increase in the numbers of both interresidue (stabilization of the $O_5 \rightarrow O_3'$ H-bond) and intraresidue H-bonds for GG and MG (see also Figure 4.S.5), and an increase in the number of intraresidue H-bonds for GA and MA (see also Figure 4.S.6).

In the series S_h^P, the H-bonding capacity of the solvent is decreased at water-like permittivity. The disaccharides MG, GA and MA show a general broadening of the global minimum in C_1, with a slight increase in the populations of the regions C_2 (GA and MA)
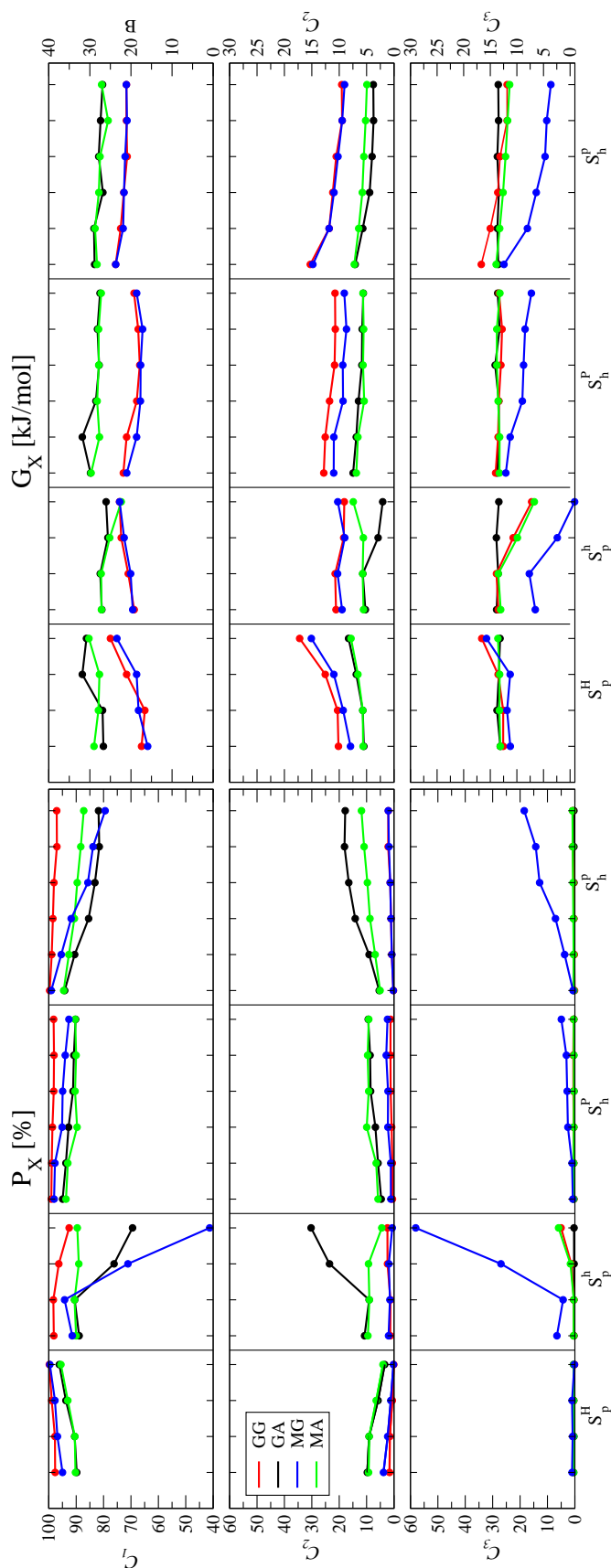
Figure 4.12: Populations and relative free energies of specific regions (Figure 4.3) in the Ramachandran free-energy maps for the four disaccharides considered (Figure 4.1) along the four series of artificial solvents (Table 4.2) at 298.15 K and 968 kg·m^{-3}. For a subset of the six regions X of the Ramachandran map, the populations P_X (three panels on the left) and the relative free energies G_X (three panels on the right) are displayed along the four solvent series. The populations P_X are calculated by integrating the reweighted probability distributions in the space of the glycosidic dihedral angles ϕ and ψ over the selected regions. The corresponding free energies are reported relative to the region with the lowest value (in general in the region C_1, for this reason omitted). The region A has zro population and is therefore also omitted. The data are based on the 60 ns sampling phases of the corresponding LEUS simulations. The data points follow the four series of artificial solvents (Table 4.2), S_p^H, S_p^H, S_p^h, S_h^P and S_h^p. The different panels (top to bottom) correspond to the different regions. The regions C_1, C_2 and C_3 are considered for the populations, while the regions B, C_2 and C_3 are considered for the relative free energies. The data is reported numerically in Suppl. Mat. Tables 4.S.4 and 4.S.5 (populations) and 4.S.6 and 4.S.7 (free energies).

and C_3 (MG). In the opposite, GG seems to stabilize the global minimum in C_1, that keeps a population above 98% in all the solvents of the series. The H-bonding patterns of GG, MG and GA are consistent with the conformational analysis, with a slight increase in the numbers of both inter- and intraresidue H-bonds compensating for the decrease of H-bonding with the solvent. For MA, the H-bonding pattern does not show significant variations.

In the series S_p^h, the dielectric permittivity is decreased while the H-bonding capacity is lower than that of water. For the four disaccharides, the trends along this series is analogous to those observed along the series of physical solvents with decreasing polarity. The appearance of a second minimum in region C_3 for GG and MG, with the additional third minimum in region C_2 for the latter (metastable state separated by a high barrier in the first case, stable state with a free energy comparable to the global minimum in the second case) is compatible with the formation of a double H-bond between the rings. The increase in the number of interresidue H-bonds and the additional increase in the number of intraresidue H-bonds are again similar to the behaviour in the physical solvents of low polarity. The formation of the second minimum in the region C_2 for GA and MA (stable state with a free energy comparable to the global minimum in the first case, metastable state separated by a high barrier in the second case) is again associated with the increase in the number of interresidue H-bonds (GA) and with an increase in the number of all intraresidue H-bonds for both disaccharides.

In the series S_h^p, the H-bonding capacity is decreased while the dielectric permittivity is lower than that of water. This evolution seems to have a similar qualitative effect on the four compounds. However, a more careful analysis reveals that in this case, the increase in the populations and the decrease in the relative free energy in regions C_2 and C_3 (effects anyway less significant in this series than the changes occurring when decreasing the polarity for the physical solvents) are not always the consequence of the formation of

a separate minimum, but can be seen as a broadening of the global minimum or of the low free-energy basin present in that region. The variations in the H-bonding patterns of the different disaccharides are similar, although less pronounced than the ones associated to the decrease of the dielectric permittivity of the solvent. An important difference is the constant occurrence of intraresidue H-bonding.
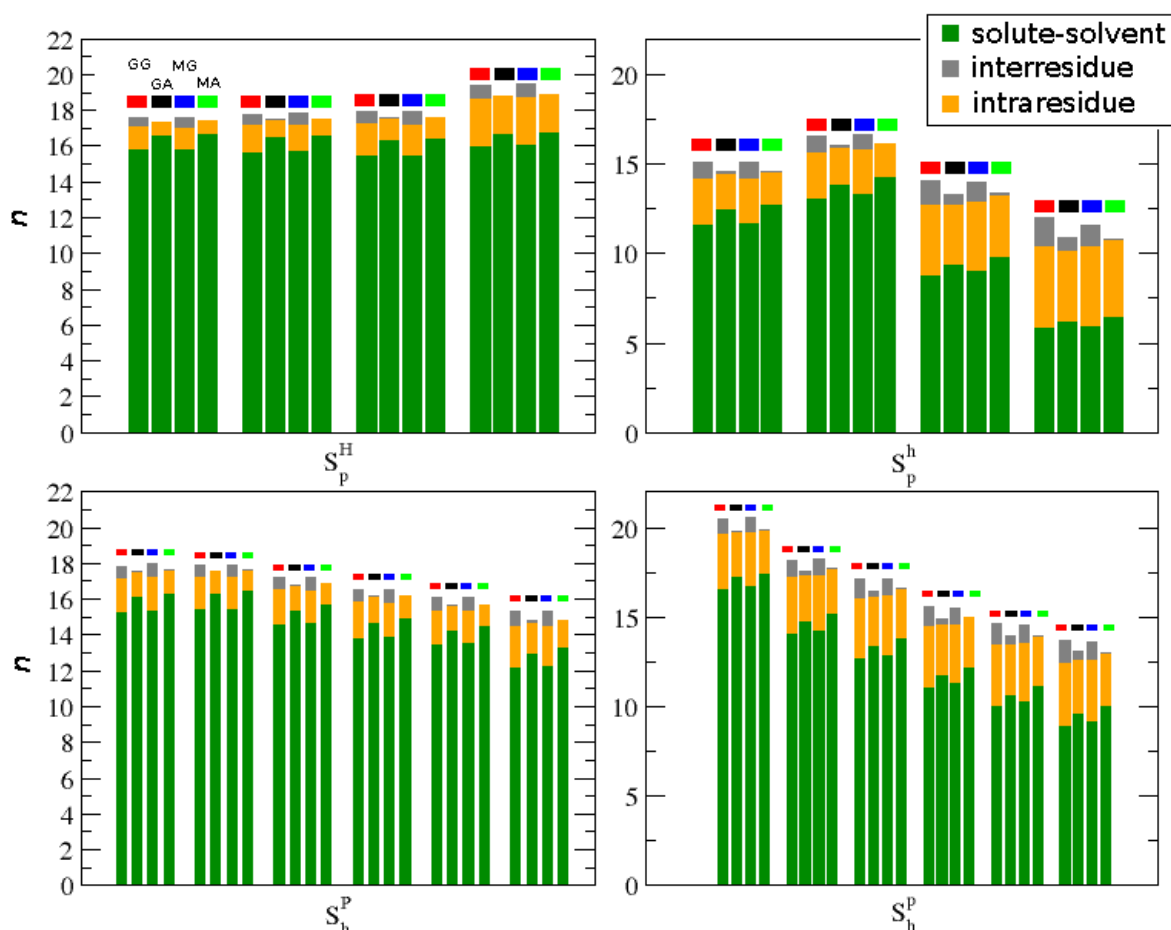

Figure 4.13: Numbers of H-bonds for the four disaccharides considered in the artificial solvents at 298.15 K and 968 kg·m^{-3}. The results are based on the 60 ns sampling phase of the corresponding LEUS simulations. Each bar represents the total number n of H-bonds, with different colors distinguishing interresidue H-bonds, intraresidue H-bonds, and solute-solvent H-bonds. The different bars for each solvent refer to one of the four disaccharides (Figure 4.1), that are identified by a box with the same color code as in Figure 4.7. The bars follow the series of the artificial solvents, the series (a) S_p^H, (b) S_p^h, (c) S_h^P and (d) S_h^p (Table 4.2). The data is reported numerically in Suppl. Mat. Tables 4.S.8, 4.S.9, 4.S.10 and 4.S.11.

4.4 Conclusions

Hydrogen-bonding (H-bonding) is often regarded as a determinant driving force in the conformational preferences of (bio)molecules. However, the influence of H-bonding cannot be discussed without consideration of the solute environment. When focusing on solvent-exposed H-bonds in an aqueous environment, competition by the solvent molecules and screening of the H-bonded interaction by the solvent dielectric response reduces their influence to that of a minor (possibly negligible or even, in some cases, adverse) conformational driving as well as steering force. However, when the polarity of the solvent is decreased, solvent-exposed H-bonds can become a very significant conformational steering force.

Qualitative or semi-quantitative evaluations of the influence of H-bonding on the conformational preferences of carbohydrates support this view in the context of monosaccharides (see Chapter Chapter 3). However, for more complex systems such as oligosaccharides or polysaccharides, the number of factors involved in the stability of specific conformations can increase drastically. In the present study, MD simulations were used to analyze the influence of H-bonding on the conformational properties of disaccharides. To this purpose cellobiose and its epimers at C_2 and C'_3 were considered. The distributions of the glycosidic dihedral angles ϕ and ψ were calculated in physical and artificial solvents of decreasing polarity.

In an aqueous environment, the epimerization has only a very limited influence on the conformational properties of the disaccharides considered, with no visible effects in case of epimerization at C_2. On the other hand, intramolecular the H-bonding pattern of the four disaccharides is strongly modified by the epimerization, which inhibits the formation of the typical interresidue H-bond ($H'_3 \rightarrow O_5$) often considered responsible for the equilibrium conformation of cellobiose. In water, the lack of this intramolecular H-bond is easily compensated for by a stronger interaction with the solvent molecules. This compensation

prevents possible large conformational changes necessary to recover the intramolecular H-bond. The H-bonding is in this case an opportunistic consequence of the close proximity of two H-bonding groups in a given epimer and conformation, and not a factor determining the stability of this conformation for the epimer where it is possible.

When solvents of lower polarity are considered, significant conformational changes occur. Alternative local minima appear in the free-energy maps, representing either metastable or stable states depending on the specific epimer. The conformational changes can be explained by the influence of intramolecular the H-bonding. However, several H-bonds are possible, both intra- and interresidue. Their influence can act either in a cooperative way or in a competitive way. In the first case, specific H-bonds concur to enhance the stability of a specific conformation, shifting the conformational preference towards this specific state. In the second case, incompatible H-bonds compete with each other, representing in this case an adverse force towards this specific conformation. Therefore, in the case of low-polarity solvents, solvent-exposed H-bonding can become as a very significant steering force.

These considerations focus on the regions of the free-energy maps of the glycosidic dihedral angles ϕ and ψ that were sampled during the simulations, while vast areas of the Ramachandran maps were not explored. The H-bond analysis reveals that the conformations corresponding to these excluded region would be the most favorable ones in terms of the number of H-bonds compatible with these values of the dihedral angles. However, the same conformations involve steric clashes between atoms of the two residue that drastically increase their free energy. The evaluation of this aspect requires additional work.
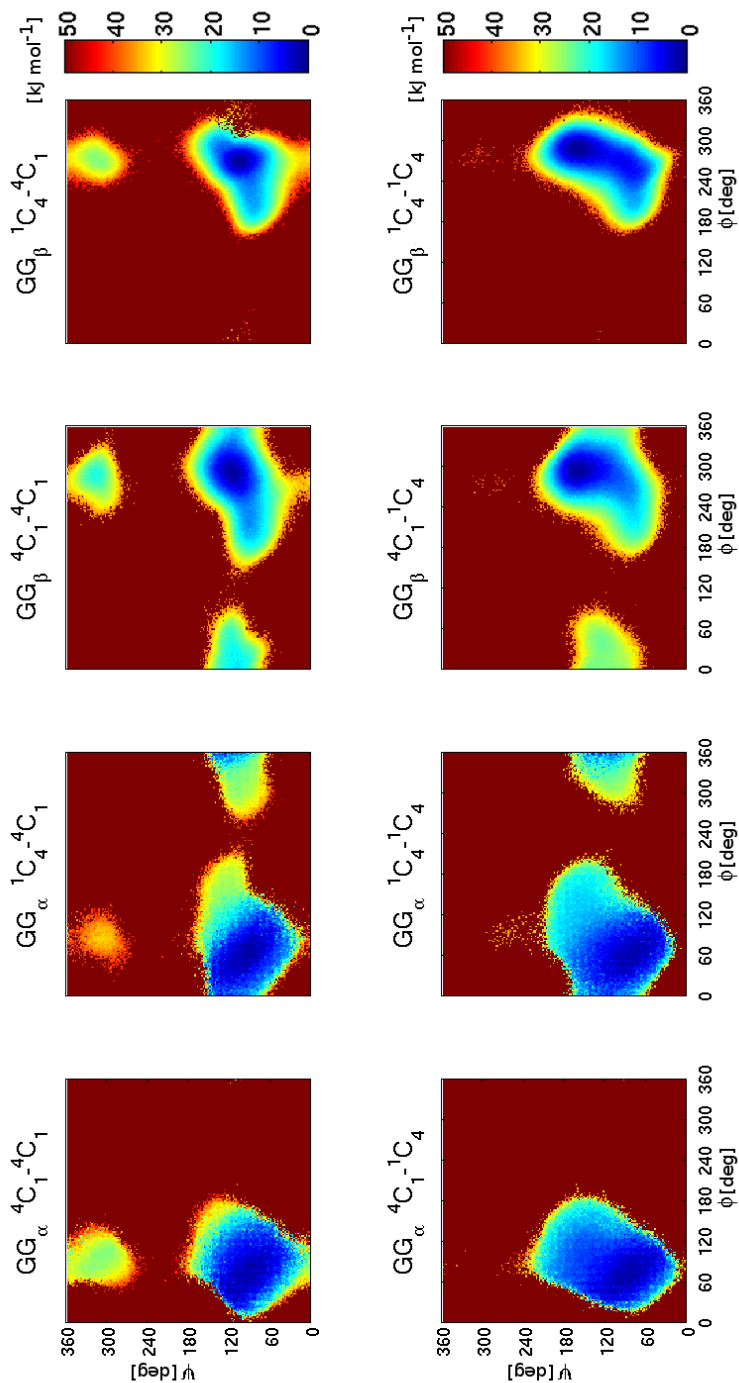
Figure A.14: Free-energy maps $G(\phi, \psi)$ in the space of the glycosidic dihedral angles ϕ and ψ. The maps are displayed for Glc$p$-α(1→4)-Glc$p$-α (GG$_\alpha$, α-maltose; four panels on the left) and Glc$p$-β(1→4)-Glc$p$-β (GG$_\beta$, β-cellobiose; four panels on the right). In both cases, the four panels correspond to simulations involving restraining of the two rings in specific chair conformations, namely $^4C_1$-$^4C_1$, $^1C_4$-$^4C_1$, $^4C_1$-$^1C_4$ and $^1C_4$-$^1C_4$, where the two symbols specify the conformation of the first and the second ring, respectively. The free energies were calculated based on the 60 ns sampling phase of the corresponding LEUS simulations at 1 bar and 298.15 K in water. The maps are anchored to $G = 0$ 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} (which is about 50 kJ mol^{-1} here in all cases).

4.A   Influence of the ring conformation

The simulations reported in this chapter were all performed with dihedral-angle restraints applied to the two rings so as to ensure a $^4C_1$ chair conformation throughout the simulations. The influence of possible flips in the ring conformations is analyzed in this appendix focusing on the GG disaccharide (β-cellobiose, named here GG_β). The $\alpha(1{\to}4)$-linked Glc-based disaccharide with α-anomery at the reducing residue, namely Glc$p$-$\alpha(1{\to}4)$-Glc$p$-α (GG_α, α-mannose) is also considered.

LEUS simulations of the GG_β and GG_α disaccharides with the two rings restrained in different conformations ($^4C_1$-$^4C_1$, $^1C_4$-$^4C_1$, $^4C_1$-$^1C_4$ or $^1C_4$-$^1C_4$) were performed in water. All the parameters in the LEUS simulations were the same as for the main simulations. These simulations are labelled LEUS-R_1-R_2, R_1 and R_2 referring to the conformation of the two rings.

The free-energy maps obtained from simulations of GG_α and GG_β in water with rings restrained in different conformations are shown in Figure A.14. As suggested in Ref.[329], the influence of the ring conformation on the glycosidic linkage properties is limited when focusing on the equilibrium configuration ($\phi \simeq 300°$ and $\psi \simeq 60°$). The only significant difference is observed for GG_β, where the global minimum is shifted by about $60°$ in ψ when restraining the second ring in an inverted-chair conformation. However, the extent of the space visited and the details of free-energy map in the alternative metastable states reveal that the effect of the puckering is not completely negligible.

In a study concerning the effect of the solvent on the conformational properties of the glycosidic linkage, any discussion would be complicated by the presence of alternative ring conformations. For this reason, the ring conformations were restrained to $^4C_1$ for both rings during the LEUS simulations as reported in the main text (Section 4.2.3). In the absence of restraints (data not shown), several simulations evidenced the occurrence of a

transition, with a higher propensity of the second ring to flip. The reasons for the tendency to a distortion in the second ring were further investigated and the results of this analysis led to the publication of a patch for the $56A6_{CARBO}$ force field ($56A6_{CARBO\_R}$). For more details see Ref[246].

Results from unrestrained simulations that did not evidence a transition in the ring conformation and from restrained simulations were compared and suggested that the effect of the restraints on the glycosidic-linkage properties can be considered negligible (relative to a situation without restraints where no flip occurs).

4.S.1 Supplementary Material

Comp.	Reg.	H_2O	DMSO	CH_3OH	$CHCl_3$	CCl_4
GG	A	0.00	0.00	0.00	0.00	0.00
	B	0.04	0.04	0.04	0.01	0.05
	C	99.96	99.96	99.96	99.99	99.95
	C_1	98.88	97.95	98.97	96.10	91.56
	C_2	0.73	1.38	0.64	2.20	3.76
	C_3	0.35	0.63	0.35	1.69	4.63
GA	A	0.00	0.00	0.00	0.00	0.00
	B	0.00	0.00	0.00	0.00	0.01
	C	100.00	100.00	100.00	100.00	99.99
	C_1	93.00	89.04	91.89	76.93	75.86
	C_2	6.51	10.58	7.53	22.77	23.59
	C_3	0.49	0.38	0.58	0.29	0.54
MG	A	0.00	0.00	0.00	0.00	0.00
	B	0.03	0.04	0.05	0.00	0.00
	C	99.97	99.96	99.95	100.00	100.00
	C_1	97.84	88.33	94.89	56.58	60.45
	C_2	1.00	2.61	1.12	1.23	0.42
	C_3	1.13	9.02	3.95	42.18	39.12
MA	A	0.00	0.00	0.00	0.00	0.00
	B	0.00	0.00	0.00	0.01	0.00
	C	100.00	100.00	100.00	99.99	100.00
	C_1	92.68	94.76	97.07	90.63	93.90
	C_2	6.84	4.05	2.19	4.86	1.97
	C_3	0.47	1.20	0.73	4.51	4.13

Table 4.S.1: Populations (in percent) of specific regions (Figure 4.3) in the Ramachandran free-energy maps for the four disaccharides considered (Figure 4.1) along the series of physical solvents at 298.15 K and 1 bar. The results are based on the 60 ns sampling phase of the corresponding LEUS simulations. The populations P_X are calculated by integrating the reweighted probability distributions in the space of the glycosidic dihedral angles ϕ and ψ over the selected regions. The entries follow the series of physical solvents (Table 4.1) in order of decreasing polarity and refer to the four disaccharides considered (GG, GA, MG, MA). The regions considered are A, B, and C and C_1, C_2, as well as the subregions C_3. Note that the populations of C_1, C_2 and C_3 sum up to that of C, and the populations of A, B and C sum up to 100%. The data is illustrated graphically in Figure 4.7.

Comp.	Reg.	H$_2$O			DMSO			CH$_3$OH			CHCl$_3$			CCl$_4$		
		G_X	$\tilde{G}_X$	$(\tilde{\phi}_X, \tilde{\psi}_X)$	G_X	$\tilde{G}_X$	$(\tilde{\phi}_X, \tilde{\psi}_X)$	G_X	$\tilde{G}_X$	$(\tilde{\phi}_X, \tilde{\psi}_X)$	G_X	$\tilde{G}_X$	$(\tilde{\phi}_X, \tilde{\psi}_X)$	G_X	$\tilde{G}_X$	$(\tilde{\phi}_X, \tilde{\psi}_X)$
GG	B	19.2	18.9	(290,320)	19.1	18.7	(294,318)	19.4	19.6	(288,320)	21.8	22.5	(260,302)	18.5	18.7	(262,304)
	C_1	0.0	0.0	(294,116)	0.0	0.0	(284,112)	0.0	0.0	(292,116)	0.0	0.0	(270,112)	0.0	0.0	(274,108)
	C_2	12.2	-	-	10.6	-	-	12.5	-	-	9.4	-	-	7.9	-	-
	C_3	14.0	-	-	12.5	12.7	(52,114)	14.0	15.6	(38,118)	10.0	9.3	(56,116)	7.4	6.4	(54,118)
GA	B	30.7	28.5	(284,318)	27.7	24.9	(282,304)	26.4	25.2	(282,312)	25.0	23.8	(282,312)	23.1	21.7	(280,310)
	C_1	0.0	0.0	(290,106)	0.0	0.0	(290,104)	0.0	0.0	(290,102)	0.0	-	-	0.0	-	-
	C_2	6.6	-	-	5.3	-	-	6.2	-	-	3.0	0.0	(236, 82)	2.9	0.0	(232, 80)
	C_3	13.0	-	-	13.5	-	-	12.6	-	-	13.8	-	-	12.2	-	-
MG	B	20.4	19.7	(288,320)	18.9	18.2	(286,320)	19.0	18.5	(288,314)	23.4	21.9	(288,320)	24.8	23.3	(288,330)
	C_1	0.0	0.0	(284,114)	0.0	0.0	(292,110)	0.0	0.0	(286,110)	0.0	0.0	(286,112)	0.0	0.0	(292,114)
	C_2	11.4	11.4	(188, 90)	8.7	8.0	(186, 88)	11.0	9.7	(192, 90)	9.5	7.7	(188, 90)	12.3	9.7	(190, 84)
	C_3	11.1	-	-	5.6	6.2	(34,114)	7.9	7.6	(350,110)	0.7	0.7	(34,116)	1.1	0.5	(6,106)
MA	B	29.4	27.5	(290,310)	31.5	29.4	(288,306)	30.0	28.6	(280,300)	23.7	22.1	(280,314)	24.7	24.0	(284,302)
	C_1	0.0	0.0	(296,106)	0.0	0.0	(302,108)	0.0	0.0	(302,106)	0.0	0.0	(298,106)	0.0	0.0	(302,104)
	C_2	6.4	-	-	7.8	8.1	(202, 76)	9.4	-	-	7.2	5.6	(190, 82)	8.4	8.4	(194, 80)
	C_3	13.1	-	-	10.8	-	-	12.1	-	-	7.4	7.1	(36, 82)	7.7	8.2	(6, 78)

Table 4.S.2: Free energies (in kJ mol^{-1}) and positions of the minima (in degrees) of specific regions (Figure 4.3) defined in the Ramachandran free-energy maps for the four disaccharides considered (Figure 4.1) along the series of pysical solvents at 298.15 K and 1 bar. The results are based on the 60 ns sampling phase of the corresponding LEUS simulations. The free energies, calculated either for the state X (G_X) or at the corresponding local minimum ($\tilde{G}_X$ at the position $\tilde{\phi}_X, \tilde{\psi}_X$), are reported relative to the corresponding lowest value (in the region C_1 except for GA in CCl$_4$, where it is in region C_2). The region A never presents a minimum and is therefore omitted. The minimum in the region C correspond to the lowest of the minima in the subregions C_1, C_2 or C_3. The regions considered are B, C_1, C_2 and C_3. The entries follow the series of physical solvents (Table 4.1) in order of decreasing polarity and refer to the four disaccharides considered (GG, GA, MG, MA). The data is illustrated graphically in Figure 4.7.

Table 4.S.3: H-bonding analysis for the four disaccharides considered in the series of pysical solvents at 298.15 K and 1 bar. The results are based on the 60 ns sampling phase of the corresponding LEUS simulations. For each disaccharide, the first set of lines reports the occurence (in percent) of each of the possible interresidue H-bonds. The second set of lines refers to the the occurence (in percent) of the intraresidue H-bonds considered separately for the two residues (the first column corresponds to the data for the first residue, the second to the data for the second residue). In parenthesis the percentage of the hydrogen bonds that involve the first atom as a donor is reported. The third set of lines describes the total number n_i of intramolecular H-bonds for each hydroxyl group of the two residues (the first column corresponds to the data for the first residue, the second to the data for the second residue). The fourth and fifth set of lines describe the number n_s of solute-solvent H-bonds for each hydroxyl group of the two residues (the first column corresponds to the data for the first residue, the second to the data for the second residue). The calculation of all H-bonds involved appropriate reweighting of the configurations so as to remove the effect of the biasing potential energy term. The data points follow the series of physical solvents (Table 4.1) in order of decreasing polarity. The data is illustrated graphically in Figure 4.8 and Suppl. Mat. Figures 4.S.2-4.S.4.

Left portion of the table (compounds GG and MG)

Cmp.	A.	H_2O (1)	H_2O (2)	DMSO (1)	DMSO (2)	CH_3OH (1)	CH_3OH (2)	$CHCl_3$ (1)	$CHCl_3$ (2)	CCl_4 (1)	CCl_4 (2)
GG	O2-O3'	0.0		0.3		0.1		1.5		4.1	
	O6-O3'	1.7		3.5		2.5		1.8		1.9	
	O5-O3'	67.9		86.4		86.0		91.0		85.5	
	O1-O3'	50.3		51.7		55.9		56.0		55.2	
	O2-O6'	1.8		10.8		8.7		63.3		53.5	
	O5-O6'	0.0		0.0		0.0		0.4		1.0	
	O6-O6'	0.0		0.1		0.0		0.9		2.3	
	O1-O6'	0.5		2.3		2.4		6.3		2.0	
	O2-O3	34.0	38.2	38.1	52.3	42.5	55.1	78.6	84.4	80.8	85.0
	O2-O1	3.7	4.4	3.6	3.8	3.1	4.5	19.4	34.2	17.2	39.2
	O2-O5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	O3-O4	39.4	0.0	41.4	0.0	45.8	0.0	84.2	0.0	84.8	0.0
	O4-O6	2.6	0.0	3.7	0.0	5.0	0.0	79.4	0.0	78.6	0.0
	O6-O5	3.9	8.0	8.7	5.2	3.2	8.3	3.4	14.4	4.2	20.3
	O1-mol	0.54	0.04	0.58	0.04	0.61	0.04	0.82	0.34	0.74	0.39
	O2-mol	0.39	0.42	0.53	0.57	0.55	0.60	1.63	1.70	1.56	1.67
	O3-mol	1.44	1.57	1.46	1.96	1.54	2.00	2.00	2.86	2.05	2.74
	O4-mol	0.42	-	0.45	-	0.51	-	1.64	-	1.63	-
	O5-mol	0.04	0.08	0.09	0.05	0.03	0.08	0.04	0.14	0.05	0.20
	O6-mol	0.04	0.10	0.07	0.18	0.08	0.20	0.83	0.85	0.84	0.79
	O1-svt	0.09	2.10	0.00	0.95	0.01	1.73	0.01	0.12		
	O2-svt	1.83	1.77	0.80	0.58	1.17	1.16	0.03	0.06		
	O3-svt	1.71	1.20	0.83	0.16	1.10	0.43	0.04	0.03		
	O4-svt	1.72	0.00	0.75	0.00	1.08	0.00	0.03	0.00		
	O5-svt	0.03	0.40	0.00	0.00	0.00	0.13	0.00	0.04		
	O6-svt	2.18	2.12	0.90	0.88	1.67	1.68	0.10	0.06		
MG	O2-O3'	1.3		9.9		4.4		43.6		42.4	
	O6-O3'	1.9		2.7		2.0		1.2		0.2	
	O5-O3'	72.5		76.2		77.6		49.3		46.0	
	O1-O3'	53.3		53.5		57.1		41.5		54.3	
	O2-O6'	0.2		1.8		0.7		1.1		0.3	
	O5-O6'	0.0		0.2		0.1		1.3		0.4	
	O6-O6'	0.0		1.5		0.3		11.8		5.4	
	O1-O6'	0.0		0.1		0.1		0.6		0.3	
	O2-O3	22.6	38.1	24.9	52.7	25.9	54.2	71.1	87.4	63.9	88.6
	O2-O1	5.9	4.5	5.9	3.4	5.0	4.5	8.3	35.6	1.1	40.6
	O2-O5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	O3-O4	37.6	0.0	42.5	0.0	43.2	0.0	81.3	0.0	83.1	0.0
	O4-O6	2.6	0.0	9.3	0.0	12.8	0.0	76.5	0.0	90.5	0.0
	O6-O5	3.9	8.8	6.8	5.3	3.8	7.0	9.9	41.3	4.1	48.9
	O1-mol	0.59	0.05	0.60	0.03	0.62	0.05	0.50	0.36	0.56	0.41
	O2-mol	0.30	0.43	0.43	0.56	0.36	0.59	1.24	1.23	1.08	1.29
	O3-mol	1.37	1.67	1.38	1.95	1.42	1.95	2.13	2.23	2.40	2.31
	O4-mol	0.40	-	0.52	-	0.56	-	1.58	-	1.74	-
	O5-mol	0.04	0.09	0.07	0.05	0.04	0.07	0.11	0.41	0.04	0.49
	O6-mol	0.08	0.09	0.20	0.09	0.19	0.08	0.99	0.56	1.00	0.55
	O1-svt	0.08	2.11	0.00	0.95	0.01	1.69	0.01	0.10		
	O2-svt	1.91	1.79	0.85	0.59	1.44	1.17	0.05	0.04		
	O3-svt	1.75	1.12	0.82	0.14	1.26	0.55	0.05	0.03		
	O4-svt	1.75	0.00	0.76	0.00	1.08	0.00	0.04	0.00		
	O5-svt	0.04	0.40	0.00	0.09	0.01	0.12	0.00	0.02		
	O6-svt	2.20	2.14	0.85	0.93	1.62	1.69	0.08	0.08		

Right portion of the table (compounds GA and MA)

Cmp.	A.	H_2O (1)	H_2O (2)	DMSO (1)	DMSO (2)	CH_3OH (1)	CH_3OH (2)	$CHCl_3$ (1)	$CHCl_3$ (2)	CCl_4 (1)	CCl_4 (2)
GA	O2-O3'	0.0		0.0		0.0		0.0		0.0	
	O6-O3'	0.0		0.0		0.0		0.0		0.0	
	O5-O3'	1.0		1.7		1.6		2.3		2.2	
	O1-O3'	12.5		8.1		12.6		27.1		22.7	
	O2-O6'	3.8		18.8		9.0		60.7		51.9	
	O5-O6'	0.0		0.0		0.0		0.0		0.0	
	O6-O6'	0.0		0.0		0.0		0.0		0.0	
	O1-O6'	0.4		5.7		2.2		5.2		3.5	
	O2-O3	35.1	15.8	41.7	14.8	44.1	18.2	81.2	50.0	83.2	47.6
	O2-O1	3.5	5.9	3.9	8.0	4.6	6.9	12.6	31.3	13.7	33.9
	O2-O5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	O3-O4	38.3	0.0	41.8	0.0	47.7	0.0	84.3	0.0	86.2	0.0
	O4-O6	2.6	0.0	10.4	0.0	11.5	0.0	79.5	0.0	84.5	0.0
	O6-O5	5.0	7.7	12.8	4.7	4.9	8.3	8.0	17.9	6.2	23.5
	O1-mol	0.16	0.06	0.18	0.08	0.19	0.07	0.31	0.07	0.40	0.34
	O2-mol	0.42	0.15	0.64	0.22	0.58	0.22	1.55	1.12	1.49	1.20
	O3-mol	1.44	0.22	1.52	0.24	1.56	0.29	2.06	1.10	2.06	1.11
	O4-mol	0.41	-	0.52	-	0.59	-	1.64	-	1.71	-
	O5-mol	0.05	0.08	0.13	0.05	0.05	0.08	0.08	0.18	0.06	0.24
	O6-mol	0.03	0.12	0.10	0.29	0.12	0.19	0.80	0.84	0.85	0.79
	O1-svt	0.19	2.10	0.00	0.96	0.04	1.64	0.02	0.07		
	O2-svt	1.78	1.89	0.76	0.88	1.16	1.45	0.03	0.06		
	O3-svt	1.72	1.77	0.81	0.85	1.10	1.26	0.03	0.05		
	O4-svt	1.74	0.00	0.73	0.00	1.01	0.00	0.03	0.00		
	O5-svt	0.14	0.44	0.00	0.00	0.04	0.13	0.01	0.02		
	O6-svt	2.24	2.16	0.88	0.81	1.71	1.68	0.09	0.07		
MA	O2-O3'	0.0		0.1		0.1		2.7		2.3	
	O6-O3'	0.0		0.0		0.0		0.0		0.0	
	O5-O3'	1.0		0.9		1.1		2.1		0.8	
	O1-O3'	15.7		25.6		21.5		29.7		16.4	
	O2-O6'	0.5		1.5		0.5		3.6		1.5	
	O5-O6'	0.0		0.0		0.0		0.0		0.0	
	O6-O6'	0.0		0.0		0.0		0.0		0.0	
	O1-O6'	0.1		0.5		0.3		1.5		0.4	
	O2-O3	23.1	15.7	26.0	17.4	28.7	17.5	71.9	56.3	78.9	54.2
	O2-O1	6.9	5.8	19.5	6.0	10.8	6.2	19.7	30.2	35.5	32.1
	O2-O5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	O3-O4	36.1	0.0	41.4	0.0	43.6	0.0	83.1	0.0	85.4	0.0
	O4-O6	3.8	0.0	9.0	0.0	12.6	0.0	86.5	0.0	84.8	0.0
	O6-O5	5.5	7.9	7.6	5.2	5.2	9.2	4.7	38.4	5.7	39.6
	O1-mol	0.23	0.06	0.46	0.06	0.33	0.06	0.51	0.30	0.52	0.32
	O2-mol	0.31	0.22	0.47	0.23	0.40	0.24	0.98	0.87	1.18	0.86
	O3-mol	1.37	0.32	1.36	0.44	1.42	0.40	2.11	0.91	2.11	0.74
	O4-mol	0.40	-	0.50	-	0.56	-	1.70	-	1.70	-
	O5-mol	0.06	0.08	0.08	0.05	0.05	0.09	0.05	0.38	0.06	0.40
	O6-mol	0.09	0.09	0.17	0.07	0.18	0.10	0.91	0.44	0.91	0.42
	O1-svt	0.15	2.11	0.00	0.96	0.02	1.69	0.01	0.08		
	O2-svt	1.91	1.90	0.91	0.85	1.50	1.48	0.06	0.06		
	O3-svt	1.77	1.76	0.76	0.88	1.25	1.29	0.04	0.05		
	O4-svt	1.75	0.00	0.72	0.00	1.08	0.00	0.03	0.00		
	O5-svt	0.17	0.44	0.00	0.00	0.03	0.13	0.01	0.02		
	O6-svt	2.22	2.16	0.88	0.93	1.66	1.72	0.08	0.08		

Cmp.		Series S_p^H				Series S_p^h			
	Reg.	$W_{1.4}^{0.8}$	$W_{1.2}^{0.9}$	$W_{1.0}^{1.0}$	$W_{0.5}^{1.5}$	Reg.	$W_{1.4}^{0.5}$	$W_{0.9}^{0.8}$	$W_{0.6}^{0.7}$	$W_{0.5}^{0.3}$
	A	0.00	0.00	0.00	0.00	A	0.00	0.00	0.00	0.00
	B	0.09	0.12	0.02	0.00	B	0.04	0.02	0.01	0.01
	C	99.91	99.88	99.98	100.00	C	99.96	99.98	99.99	99.99
	C_1	97.67	97.73	98.93	99.78	C_1	98.16	98.31	96.32	92.57
	C_2	1.63	1.53	0.63	0.10	C_2	1.37	1.29	2.38	2.39
	C_3	0.61	0.61	0.42	0.12	C_3	0.43	0.37	1.29	5.03
	A	0.00	0.00	0.00	0.00	A	0.00	0.00	0.00	0.00
	B	0.00	0.00	0.00	0.00	B	0.00	0.00	0.00	0.00
GA	C	100.00	100.00	100.00	100.00	C	100.00	100.00	100.00	100.00
	C_1	89.79	90.55	93.74	96.13	C_1	88.90	90.54	76.16	69.43
	C_2	9.75	9.10	5.83	3.40	C_2	10.76	9.07	23.55	30.26
	C_3	0.45	0.36	0.42	0.48	C_3	0.34	0.39	0.28	0.31
	A	0.00	0.00	0.00	0.00	A	0.00	0.00	0.00	0.00
	B	0.15	0.06	0.05	0.01	B	0.03	0.03	0.01	0.00
MG	C	99.85	99.94	99.95	99.99	C	99.97	99.97	99.99	100.00
	C_1	94.98	96.82	97.74	99.59	C_1	91.43	94.20	71.11	41.27
	C_2	3.85	2.30	1.17	0.23	C_2	1.99	1.48	1.90	0.66
	C_3	1.02	0.82	1.04	0.18	C_3	6.54	4.29	26.98	58.07
	A	0.00	0.00	0.00	0.00	A	0.00	0.00	0.00	0.00
	B	0.00	0.00	0.00	0.00	B	0.00	0.00	0.00	0.01
MA	C	100.00	100.00	100.00	100.00	C	100.00	100.00	100.00	99.99
	C_1	90.25	90.50	93.04	95.57	C_1	90.01	90.60	89.04	89.58
	C_2	9.30	9.07	6.51	4.03	C_2	9.52	9.00	9.32	4.46
	C_3	0.46	0.42	0.45	0.40	C_3	0.47	0.39	1.64	5.95

Table 4.S.4: Populations (in percent) of specific regions (Figure 4.3) in the Ramachandran free-energy maps for the four disaccharides considered (see Fig. 4.1) along the series S_p^H (left) and S_p^h (right) of artificial solvents at 298.15 K and 968 kg·m^{-3}. The results are based on the 60 ns sampling phases of the corresponding LEUS simulations. The populations P_X are calculated by integrating the reweighted probability distributions in the space of the glycosidic dihedral angles ϕ and ψ over the selected regions. The entries follow two of the four series of artificial solvents (Table 4.2) in order of decreasing dielectric permittivity: on the left the series S_p^H, on the right the S_p^h and refer to the four disaccharides considered (GG, GA, MG, MA). The regions considered are A, B, and C as well as subregions C_1, C_2 and C_3. Note that the populations of C_1, C_2 and C_3 sum up to that of C, and the populations of A, B and C sum up to 100%. The data is illustrated graphically in Figure 4.12.

Comp.	Reg.	Series S_h^P $W^{1.1}_{0.9}$	$W^{1.0}_{1.0}$	$W^{0.8}_{1.1}$	$W^{0.7}_{1.2}$	$W^{0.6}_{1.4}$	$W^{0.5}_{1.5}$	Reg.	Series S_h^P $W^{1.5}_{0.4}$	$W^{1.2}_{0.5}$	$W^{1.1}_{0.5}$	$W^{0.8}_{0.7}$	$W^{0.6}_{0.9}$	$W^{0.4}_{1.3}$
GG	A	0.00	0.00	0.00	0.00	0.00	0.00	A	0.00	0.00	0.00	0.00	0.00	0.00
	B	0.01	0.02	0.05	0.07	0.06	0.04	B	0.01	0.01	0.02	0.02	0.02	0.02
	C	99.99	99.98	99.95	99.93	99.94	99.96	C	99.99	99.99	99.98	99.98	99.98	99.98
	C_1	99.07	98.93	98.65	98.18	98.07	98.23	C_1	99.66	98.92	98.49	98.10	97.04	97.09
	C_2	0.56	0.63	0.86	1.23	1.30	1.28	C_2	0.21	0.84	1.09	1.39	2.12	2.07
	C_3	0.35	0.42	0.44	0.52	0.56	0.45	C_3	0.12	0.23	0.41	0.49	0.82	0.82
GA	A	0.00	0.00	0.00	0.00	0.00	0.00	A	0.00	0.00	0.00	0.00	0.00	0.00
	B	0.00	0.00	0.00	0.00	0.00	0.00	B	0.00	0.00	0.00	0.00	0.00	0.00
	C	100.00	100.00	100.00	100.00	100.00	100.00	C	100.00	100.00	100.00	100.00	100.00	100.00
	C_1	94.95	93.74	92.78	91.15	90.84	90.21	C_1	94.19	90.53	85.46	83.15	81.57	81.83
	C_2	4.68	5.83	6.80	8.54	8.73	9.42	C_2	5.40	9.10	14.15	16.51	18.07	17.81
	C_3	0.38	0.42	0.42	0.31	0.43	0.37	C_3	0.41	0.37	0.39	0.33	0.36	0.35
MG	A	0.00	0.00	0.00	0.00	0.00	0.00	A	0.00	0.00	0.00	0.00	0.00	0.00
	B	0.02	0.05	0.08	0.08	0.09	0.05	B	0.01	0.01	0.01	0.02	0.02	0.02
	C	99.98	99.95	99.92	99.92	99.91	99.95	C	99.99	99.99	99.99	99.98	99.98	99.98
	C_1	98.07	97.74	95.14	94.94	93.97	92.63	C_1	99.08	95.44	91.81	85.72	83.85	79.42
	C_2	1.16	1.17	2.24	2.18	2.84	2.39	C_2	0.25	0.80	1.10	1.39	1.85	2.09
	C_3	0.75	1.04	2.55	2.80	3.10	4.93	C_3	0.66	3.75	7.07	12.87	14.28	18.47
MA	A	0.00	0.00	0.00	0.00	0.00	0.00	A	0.00	0.00	0.00	0.00	0.00	0.00
	B	0.00	0.00	0.00	0.00	0.00	0.00	B	0.00	0.00	0.00	0.00	0.00	0.00
	C	100.00	100.00	100.00	100.00	100.00	100.00	C	100.00	100.00	100.00	100.00	100.00	100.00
	C_1	93.72	93.04	89.66	90.41	90.02	90.35	C_1	94.56	92.65	90.64	89.61	88.30	87.23
	C_2	5.84	6.51	9.95	9.24	9.63	9.21	C_2	5.09	6.90	8.79	9.72	10.92	11.87
	C_3	0.44	0.45	0.38	0.35	0.35	0.44	C_3	0.35	0.45	0.56	0.67	0.78	0.89

Table 4.S.5: Populations (in percent) of specific regions (Figure 4.3) in the Ramachandran free-energy maps for the four disaccharides considered (see Fig. 4.1) along the series S_h^P (left) and S_h^P (right) of artificial solvents at 298.15 K and 968 kg·m^{-3}. The results are based on the 60 ns sampling phases of the corresponding LEUS simulations. The populations P_X are calculated by integrating the reweighted probability distributions in the space of the glycosidic dihedral angles ϕ and ψ over the selected regions. The entries follow two of the four series of artificial solvents (Table 4.2) in order of decreasing dielectric permittivity: on the left the series S_h^P, on the right the S_h^P and refer to the four disaccharides considered (GG, GA, MG, MA). The regions considered are A, B, and C as well as the subregions C_1, C_2 and C_3. Note that the populations of C_1, C_2 and C_3 sum up to that of C, and the populations of A, B and C sum up to 100%. The data is illustrated graphically in Figure 4.12.

Cmp.	Reg.	Series S_p^H											
		$W_{1.4}^{0.8}$			$W_{1.2}^{0.9}$			$W_{1.0}^{1.0}$			$W_{0.5}^{1.5}$		
		G_X	$\tilde{G}_X$	$(\tilde{\phi}_X, \tilde{\psi}_X)$	G_X	$\tilde{G}_X$	$(\tilde{\phi}_X, \tilde{\psi}_X)$	G_X	$\tilde{G}_X$	$(\tilde{\phi}_X, \tilde{\psi}_X)$	G_X	$\tilde{G}_X$	$(\tilde{\phi}_X, \tilde{\psi}_X)$
GG	B	17.4	16.2	(284,318)	16.6	16.6	(288,316)	21.1	20.4	(288,318)	25.0	25.2	(286,316)
	C_1	0.0	0.0	(300,120)	0.0	0.0	(296,118)	0.0	0.0	(294,120)	0.0	0.0	(294,116)
	C_2	10.1	-	-	10.3	-	-	12.5	-	-	17.2	-	-
	C_3	12.6	-	-	12.6	-	-	13.5	-	-	16.6	-	-
GA	B	26.7	24.7	(284,310)	26.9	25.3	(292,308)	31.8	27.6	(292,310)	30.8	29.1	(282,312)
	C_1	0.0	0.0	(292,108)	0.0	0.0	(294,104)	0.0	0.0	(292,108)	0.0	0.0	(290,104)
	C_2	5.5	-	-	5.7	-	-	6.9	-	-	8.3	-	-
	C_3	13.1	-	-	13.7	-	-	13.4	-	-	13.2	-	-
MG	B	15.9	14.3	(294,318)	18.2	17.6	(292,318)	18.6	17.9	(294,314)	23.4	22.6	(290,314)
	C_1	0.0	0.0	(300,120)	0.0	0.0	(294,112)	0.0	0.00	(290,114)	0.0	0.0	(296,114)
	C_2	7.9	7.7	(204, 86)	9.3	9.7	(208, 92)	11.0	11.1	(190, 90)	15.1	15.5	(230, 96)
	C_3	11.2	-	-	11.8	-	-	11.3	-	-	15.7	-	-
MA	B	29.0	27.1	(292,312)	27.9	26.4	(288,308)	27.7	26.2	(286,314)	30.2	28.6	(290,312)
	C_1	0.0	0.0	(292,102)	0.0	0.0	(292,106)	0.0	0.0	(296,106)	0.0	0.0	(292,102)
	C_2	5.6	-	-	5.7	-	-	6.6	-	-	7.8	6.7	(230, 74)
	C_3	13.1	-	-	13.3	-	-	13.2	-	-	13.6	-	-

Cmp.	Reg.	Series S_p^h											
		$W_{1.4}^{0.5}$			$W_{0.9}^{0.8}$			$W_{0.6}^{0.7}$			$W_{0.3}^{0.5}$		
		G_X	$\tilde{G}_X$	$(\tilde{\phi}_X, \tilde{\psi}_X)$	G_X	$\tilde{G}_X$	$(\tilde{\phi}_X, \tilde{\psi}_X)$	G_X	$\tilde{G}_X$	$(\tilde{\phi}_X, \tilde{\psi}_X)$	G_X	$\tilde{G}_X$	$(\tilde{\phi}_X, \tilde{\psi}_X)$
GG	B	19.2	19.0	(282,316)	20.7	20.9	(284,314)	22.4	22.8	(264,304)	22.8	23.4	(264,306)
	C_1	0.0	0.0	(288,110)	0.0	0.0	(288,114)	0.0	0.0	(278,108)	0.0	0.0	(278,108)
	C_2	10.6	9.3	(210, 94)	10.7	9.8	(224, 88)	9.2	7.5	(222, 94)	9.1	-	-
	C_3	13.5	13.8	(56,118)	13.8	14.9	(52,120)	10.7	9.0	(46,120)	7.2	5.7	(48,118)
GA	B	27.1	25.5	(280,314)	27.5	26.1	(284,306)	25.6	24.4	(282,310)	26.0	24.5	(282,302)
	C_1	0.0	0.0	(284,100)	0.0	0.0	(286,102)	0.0	1.07	(234, 88)	0.0	0.4	(232, 84)
	C_2	5.2	3.0	(228, 78)	5.7	3.3	(222, 82)	2.9	0.0	(230, 82)	2.1	0.0	(222, 82)
	C_3	13.8	-	-	13.5	-	-	13.9	-	-	13.4	-	-
MG	B	19.6	18.7	(288,316)	20.1	20.0	(290,312)	21.7	21.9	(288,318)	22.8	23.3	(284,316)
	C_1	0.0	0.0	(294,114)	0.0	0.0	(288,112)	0.0	0.0	(292,114)	0.0	0.9	(290,112)
	C_2	9.5	8.5	(186, 90)	10.3	10.2	(192, 90)	9.0	7.6	(188, 90)	10.3	9.4	(184, 84)
	C_3	6.5	7.2	(32,112)	7.7	9.4	(6,104)	2.4	2.2	(30,118)	-0.8	0.0	(32,114)
MA	B	21.2	25.3	(284,312)	27.2	25.5	(286,308)	25.2	23.9	(292,310)	22.4	22.4	(284,304)
	C_1	0.0	0.0	(292,102)	0.0	0.0	(294,102)	0.0	0.0	(302,112)	0.0	0.0	(294,106)
	C_2	5.6	4.8	(196, 76)	5.7	4.7	(196, 78)	5.6	3.6	(194, 80)	7.4	5.9	(188, 78)
	C_3	13.0	-	-	13.5	-	-	9.9	10.6	(2, 82)	6.3	6.1	(6, 76)

Table 4.S.6: Free energies (in kJ mol^{-1}) and positions of the minima (in degrees) of specific regions (Figure 4.3) defined in the Ramachandran free-energy maps for the four disaccharides considered (Figure 4.1) along the series S_p^H (left) and S_p^h (right) of artificial solvents at 298.15 K and 968 kg·m^{-3}. The results are based on the 60 ns sampling phase of the corresponding LEUS simulations. The free energies, calculated either for the state X (G_X) or at the corresponding local minimum ($\tilde{G}_X$ at the position $\tilde{\phi}_X$, $\tilde{\psi}_X$), are reported relative to the corresponding lowest value (in the region C_1). The region A never presents a minimum and is therefore omitted. The minimum in the region C correspond to the lowest of the minima in the subregions C_1, C_2 or C_3. The regions considered are B, C_1, C_2 and C_3. The entries follow the series of artificial solvents (Table 4.2) in order of decreasing dielectric permittivity: on the right the series S_p^H, on the left the S_p^h and refer to the four disaccharides considered (GG, GA, MG, MA). The data is illustrated graphically in Figure 4.12.

Series S_h^P

Cmp.	Reg.	$W_{0.9}^{1.1}$			$W_{1.0}^{1.0}$			$W_{1.1}^{0.8}$			$W_{1.2}^{0.7}$			$W_{1.4}^{0.6}$			$W_{1.5}^{0.5}$		
		G_X	$\tilde{G}_X$	$(\tilde\phi_X,\tilde\psi_X)$	G_X	$\tilde{G}_X$	$(\tilde\phi_X,\tilde\psi_X)$	G_X	$\tilde{G}_X$	$(\tilde\phi_X,\tilde\psi_X)$	G_X	$\tilde{G}_X$	$(\tilde\phi_X,\tilde\psi_X)$	G_X	$\tilde{G}_X$	$(\tilde\phi_X,\tilde\psi_X)$	G_X	$\tilde{G}_X$	$(\tilde\phi_X,\tilde\psi_X)$
GG	B	21.9	21.3	(292,320)	21.1	20.4	(288,318)	18.6	18.1	(288,320)	17.9	18.2	(282,318)	18.2	17.6	(286,316)	19.2	19.3	(286,320)
	C_1	0.0	0.0	(286,110)	0.0	0.0	(294,120)	0.0	0.0	(288,112)	0.0	0.0	(290,114)	0.0	0.0	(286,114)	0.0	0.0	(280,108)
	C_2	12.8	-	-	12.5	-	-	11.8	-	-	10.9	9.99	(226,86)	10.7	9.8	(226,92)	10.8	9.9	(222,92)
	C_3	14.0	-	-	13.5	-	-	13.4	-	-	13.0	-	-	12.8	-	-	13.4	15.1	(46,122)
GA	B	29.8	27.2	(286,306)	31.8	27.6	(292,310)	28.5	26.6	(282,310)	27.7	26.4	(282,306)	28.1	26.9	(280,308)	27.5	26.3	(280,314)
	C_1	0.0	0.0	(296,104)	0.0	0.0	(292,108)	0.0	0.0	(296,104)	0.0	0.0	(290,104)	0.0	0.0	(292,104)	0.0	0.0	(290,106)
	C_2	7.5	-	-	6.9	-	-	6.5	-	-	5.9	-	-	5.8	-	-	5.6	-	-
	C_3	13.7	-	-	13.4	-	-	13.4	-	-	14.1	-	-	13.2	-	-	13.6	-	-
MG	B	21.0	20.3	(288,318)	18.6	17.9	(294,314)	17.7	17.1	(294,318)	17.6	17.1	(282,316)	17.2	16.2	(286,316)	18.6	17.8	(292,312)
	C_1	0.0	0.0	(298,110)	0.0	0.0	(290,114)	0.0	0.0	(290,112)	0.0	0.0	(288,112)	0.0	0.0	(288,112)	0.0	0.0	(290,108)
	C_2	11.0	10.7	(194,86)	11.0	11.1	(190,90)	9.3	9.1	(196,88)	9.4	8.8	(190,92)	8.7	8.1	(196,86)	9.1	8.1	(190,88)
	C_3	12.1	-	-	11.3	-	-	9.0	9.6	(32,102)	8.7	9.0	(8,106)	8.5	9.0	(8,106)	7.3	7.5	(4,108)
MA	B	29.7	27.3	(290,304)	27.7	26.2	(286,314)	28.2	26.6	(284,302)	27.8	26.2	(284,308)	27.9	25.6	(284,308)	27.2	25.5	(290,304)
	C_1	0.0	0.0	(294,102)	0.0	0.0	(296,106)	0.0	0.0	(292,108)	0.0	0.0	(292,100)	0.0	0.0	(298,104)	0.0	0.0	(294,102)
	C_2	6.9	-	-	6.6	-	-	5.4	4.1	(216,72)	5.6	4.5	(218,74)	5.5	4.6	(216,78)	5.7	4.7	(224,76)
	C_3	13.3	-	-	13.2	-	-	13.5	-	-	13.8	-	-	13.8	-	-	13.2	-	-

Series S_h^p

Cmp.	Reg.	$W_{0.4}^{1.5}$			$W_{0.5}^{1.2}$			$W_{0.5}^{1.1}$			$W_{0.7}^{0.8}$			$W_{0.9}^{0.6}$			$W_{1.3}^{0.4}$		
		G_X	$\tilde{G}_X$	$(\tilde\phi_X,\tilde\psi_X)$	G_X	$\tilde{G}_X$	$(\tilde\phi_X,\tilde\psi_X)$	G_X	$\tilde{G}_X$	$(\tilde\phi_X,\tilde\psi_X)$	G_X	$\tilde{G}_X$	$(\tilde\phi_X,\tilde\psi_X)$	G_X	$\tilde{G}_X$	$(\tilde\phi_X,\tilde\psi_X)$	G_X	$\tilde{G}_X$	$(\tilde\phi_X,\tilde\psi_X)$
GG	B	23.7	23.9	(286,320)	22.5	23.1	(290,314)	21.7	22.5	(286,322)	20.9	21.2	(284,314)	21.1	21.3	(284,320)	21.1	22.2	(290,318)
	C_1	0.0	0.0	(290,110)	0.0	0.0	(292,114)	0.0	0.0	(282,114)	0.0	0.0	(278,108)	0.0	0.0	(274,112)	0.0	0.0	(282,112)
	C_2	14.8	14.3	(230,94)	11.8	11.2	(214,90)	11.2	10.1	(228,96)	10.5	8.9	(222,88)	9.5	8.5	(218,92)	9.5	8.9	(226,88)
	C_3	12.4	19.0	(56,118)	15.0	15.6	(50,118)	13.6	13.2	(54,114)	13.2	11.9	(56,116)	11.8	11.0	(54,122)	11.8	11.7	(46,122)
GA	B	28.8	26.4	(282,316)	29.0	27.3	(282,314)	26.8	24.8	(290,310)	27.8	25.7	(278,316)	27.4	25.3	(276,310)	26.9	24.2	(278,310)
	C_1	0.0	0.0	(294,110)	0.0	0.0	(284,98)	0.0	0.0	(286,102)	0.0	0.0	(284,106)	0.0	0.1	(288,102)	0.0	0.0	(294,108)
	C_2	7.1	-	-	5.7	-	-	4.5	-	-	4.0	0.7	(224,80)	3.7	0.0	(230,82)	3.8	0.1	(228,82)
	C_3	13.5	-	-	13.7	-	-	13.4	-	-	13.7	-	-	13.4	13.8	(38,110)	13.5	14.2	(34,110)
MG	B	23.1	23.1	(288,316)	21.9	21.5	(288,314)	21.7	21.5	(286,314)	21.4	21.0	(286,314)	21.0	20.2	(284,322)	21.1	20.8	(286,316)
	C_1	0.0	0.0	(292,110)	0.0	0.0	(292,114)	0.0	0.0	(292,114)	0.0	0.0	(298,110)	0.0	0.0	(286,110)	0.0	0.0	(290,110)
	C_2	14.8	15.4	(198,90)	11.8	11.3	(188,90)	11.0	10.2	(186,86)	10.2	9.0	(188,90)	9.5	7.9	(186,88)	9.0	8.2	(188,90)
	C_3	12.4	19.0	(56,118)	8.0	9.0	(22,108)	6.3	7.6	(24,122)	4.7	5.2	(26,114)	4.4	4.7	(24,116)	3.6	3.9	(32,114)
MA	B	28.3	26.2	(282,312)	28.7	26.2	(286,314)	27.8	26.1	(288,304)	27.5	25.7	(282,310)	25.5	24.1	(284,310)	27.1	25.4	(282,316)
	C_1	0.0	0.0	(296,104)	0.0	0.0	(300,108)	0.0	0.0	(296,100)	0.0	0.0	(294,100)	0.0	0.0	(296,110)	0.0	0.0	(300,108)
	C_2	7.2	6.0	(226,76)	6.4	4.9	(196,78)	5.8	3.7	(194,80)	5.5	3.8	(192,82)	5.2	3.8	(198,78)	4.9	3.3	(190,80)
	C_3	13.9	-	-	13.2	-	-	12.6	13.9	(2,84)	12.1	13.4	(2,94)	11.7	13.3	(4,90)	11.4	11.9	(4,88)

Table 4.S.7: Free energies (in kJ mol⁻¹) and positions of the minima (in degrees) of specific regions (Figure 4.3) defined in the Ramachandran free-energy maps for the four disaccharides considered (Figure 4.1) along the series S_h^P (top) and S_h^p (bottom) of artificial solvents at 298.15 K and 968 kg·m⁻³. The results are based on the 60 ns sampling phase of the corresponding LEUS simulations. The free energies, calculated either for the state X (G_X) or at the corresponding local minimum ($\tilde{G}_X$ at the position $\tilde\phi_X$, $\tilde\psi_X$), are reported relative to the corresponding lowest value (in the region C_1). The region A never presents a minimum and is therefore omitted. The minimum in the region C correspond to the lowest of the minima in the subregions C_1, C_2 or C_3. The regions considered are B, C_1, C_2 and C_3. The entries follow the series of artificial solvents (Table 4.2) in order of decreasing dielectric permittivity: on the top the series S_h^P and on the bottom the series S_h^p, and refer to the four disaccharides considered (GG, GA, MG, MA). The data is illustrated graphically in Figure 4.12.

Sug.	A.	$W^{0.8}_{1.4}$		$W^{0.9}_{1.2}$		$W^{1.0}_{1.0}$		$W^{1.5}_{0.5}$	
	O2-O3'	0.0		0.0		0.0		0.0	
	O6-O3'	1.0		1.1		1.4		1.3	
	O5-O3'	47.0		55.1		65.2		77.8	
	O1-O3'	35.9		42.9		50.1		63.5	
	O2-O6'	0.6		1.0		1.6		0.8	
	O5-O6'	0.0		0.0		0.0		0.0	
	O6-O6'	0.0		0.0		0.0		0.0	
	O1-O6'	0.1		0.2		0.4		0.3	
	O2-O3	24.2	24.5	28.2	30.9	33.4	37.7	48.5	54.1
GG	O2-O1	2.6	3.1	2.7	3.7	3.5	5.2	8.3	8.4
	O2-O5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	O3-O4	28.5	0.00	31.9	0.00	37.4	0.00	48.7	0.00
	O4-O6	0.5	0.00	1.8	0.00	2.0	0.00	3.3	0.00
	O6-O5	2.6	3.6	3.2	5.1	3.6	8.4	8.4	26.0
	O1-svt	0.03	0.52	0.04	0.60	0.10	2.13	0.02	0.64
	O2-svt	0.47	0.48	0.51	0.54	1.86	1.79	0.58	0.58
	O3-svt	0.47	0.48	0.51	0.42	1.74	1.22	0.55	0.55
	O4-svt	0.49	0.00	0.53	0.00	1.76	0.00	0.54	0.00
	O5-svt	0.02	0.11	0.02	0.12	0.03	0.42	0.01	0.13
	O6-svt	0.53	0.53	0.58	0.60	2.20	2.17	0.62	0.65
	O2-O3'	0.0		0.0		0.0		0.0	
	O6-O3'	0.0		0.0		0.0		0.0	
	O5-O3'	0.3		0.7		0.9		2.9	
	O1-O3'	5.9		9.3		12.4		28.2	
	O2-O6'	1.3		2.1		3.1		1.4	
	O5-O6'	0.0		0.0		0.0		0.0	
	O6-O6'	0.0		0.0		0.0		0.0	
	O1-O6'	0.3		0.2		0.8		0.6	
	O2-O3	24.0	9.7	27.1	11.4	33.1	15.5	49.4	25.0
GA	O2-O1	1.6	3.2	2.2	4.3	3.5	5.8	8.7	10.2
	O2-O5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	O3-O4	28.6	0.00	31.3	0.00	35.7	0.00	50.3	0.00
	O4-O6	1.2	0.00	1.7	0.00	3.0	0.00	4.7	0.00
	O6-O5	2.0	4.2	3.4	5.3	4.8	8.5	14.2	26.3
	O1-svt	0.06	0.55	0.06	0.61	0.20	2.10	0.03	0.64
	O2-svt	0.49	0.48	0.52	0.56	1.81	1.89	0.54	0.63
	O3-svt	0.48	0.46	0.55	0.49	1.74	1.78	0.55	0.56
	O4-svt	0.50	0.00	0.53	0.00	1.76	0.00	0.56	0.00
	O5-svt	0.04	0.12	0.05	0.14	0.14	0.45	0.03	0.12
	O6-svt	0.51	0.54	0.60	0.61	2.22	2.18	0.66	0.65

(a)

Sug.	A.	$W_{1.4}^{0.8}$		$W_{1.2}^{0.9}$		$W_{1.0}^{1.0}$		$W_{0.5}^{1.5}$	
	O2-O3'	1.1		0.9		1.2		0.3	
	O6-O3'	1.2		1.2		1.7		1.5	
	O5-O3'	56.1		64.1		71.9		78.4	
	O1-O3'	42.3		47.6		53.2		64.2	
	O2-O6'	0.2		0.3		0.3		0.0	
	O5-O6'	0.0		0.0		0.0		0.0	
	O6-O6'	0.0		0.0		0.0		0.0	
	O1-O6'	0.0		0.0		0.0		0.1	
	O2-O3	14.0	25.9	18.9	29.8	22.3	38.0	36.4	53.2
MG	O2-O1	3.5	2.9	5.1	3.4	6.4	4.8	12.2	8.2
	O2-O5	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00
	O3-O4	29.3	0.00	31.6	0.00	36.6	0.00	47.6	0.00
	O4-O6	0.3	0.00	1.3	0.00	3.0	0.00	4.6	0.00
	O6-O5	1.8	3.2	2.8	4.6	3.7	7.7	9.5	26.7
	O1-svt	0.03	0.54	0.03	0.61	0.09	2.13	0.02	0.59
	O2-svt	0.50	0.50	0.54	0.59	1.92	1.81	0.56	0.58
	O3-svt	0.48	0.47	0.52	0.43	1.78	1.13	0.59	0.60
	O4-svt	0.48	0.00	0.51	0.00	1.77	0.00	0.61	0.00
	O5-svt	0.02	0.09	0.02	0.16	0.04	0.42	0.01	0.12
	O6-svt	0.57	0.53	0.60	0.59	2.20	2.12	0.65	0.65
	O2-O3'	0.0		0.0		0.0		0.0	
	O6-O3'	0.0		0.0		0.0		0.0	
	O5-O3'	0.2		0.7		0.9		2.9	
	O1-O3'	7.8		10.5		15.1		28.5	
	O2-O6'	0.3		0.5		0.5		0.3	
	O5-O6'	0.0		0.0		0.0		0.0	
	O6-O6'	0.0		0.0		0.0		0.0	
	O1-O6'	0.1		0.0		0.2		0.6	
	O2-O3	14.9	9.4	18.8	11.4	22.7	14.9	38.2	25.9
MA	O2-O1	3.6	3.2	4.4	4.2	7.4	6.1	16.0	9.6
	O2-O5	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00
	O3-O4	28.1	0.00	32.1	0.00	36.1	0.00	47.8	0.00
	O4-O6	1.1	0.00	1.9	0.00	2.2	0.00	4.9	0.00
	O6-O5	1.9	3.1	3.7	4.8	5.2	8.6	16.7	27.0
	O1-svt	0.05	0.52	0.07	0.59	0.15	2.11	0.03	0.65
	O2-svt	0.50	0.48	0.56	0.53	1.92	1.89	0.59	0.57
	O3-svt	0.51	0.48	0.52	0.49	1.78	1.78	0.58	0.56
	O4-svt	0.49	0.00	0.52	0.00	1.77	0.00	0.57	0.00
	O5-svt	0.05	0.10	0.07	0.15	0.17	0.44	0.03	0.12
	O6-svt	0.53	0.53	0.62	0.58	2.22	2.18	0.68	0.66

(b)

Table 4.S.8: H-bonding analysis for the four disaccharides considered (GG, GA, MG, MA) based on the 60 ns sampling phase of the corresponding LEUS simulations in the series of pysical solvents at 298.15 K and 1 bar. For each dimer, the first panel reports the occurence f_i of the inter-ring hydrogen bonds. The second panel refers to the the occurence f_i of the intra-ring hydrogen bonds considered separately for the two rings (the first column corresponds to the data for the first ring, the second for the second ring). In parenthesis the percentage of the hydrogen bonds that involve the first atom as a donor is reported. The third and last panel describes the average number n_s of solute-solvent hydrogen bonds for each hydroxyl group of the two rings (the first column corresponds to the data for the first ring, the second for the second ring). The calculation of all H-bonds involved appropriate reweighting of the configurations so as to remove the effect of the biasing potential energy term. The data points follow the series of artificial solvents (Table 4.2) in order of decreasing dielectric permittivity. The data is illustrated graphically in Figure 4.13 and Suppl. Mat. Figures 4.S.5-4.S.7.

Sug.	A.	$W^{0.5}_{1.4}$		$W^{0.8}_{0.9}$		$W^{0.7}_{0.6}$		$W^{0.3}_{0.5}$	
	O2-O3'	0.2		0.1		1.1		4.5	
	O6-O3'	3.7		2.9		3.4		2.0	
	O5-O3'	78.6		77.9		87.9		87.1	
	O1-O3'	55.5		55.7		57.5		57.6	
	O2-O6'	10.3		10.7		42.4		58.1	
	O5-O6'	0.0		0.0		0.2		1.0	
	O6-O6'	0.1		0.1		0.8		3.6	
	O1-O6'	1.7		2.1		8.2		7.4	
	O2-O3	70.0	78.1	77.8	83.9	48.7	59.6	49.0	58.9
	O2-O1	11.2	19.3	15.1	35.6	6.1	7.6	6.5	8.4
	O2-O5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	O3-O4	73.7	0.00	83.4	0.00	50.6	0.00	50.6	0.00
	O4-O6	56.4	0.00	77.6	0.00	12.3	0.00	12.7	0.00
GG	O6-O5	6.2	17.3	4.9	17.7	6.2	12.3	5.8	13.8
	O1-svt	0.03	0.45	0.04	0.52	0.03	0.50	0.02	0.43
	O2-svt	0.40	0.44	0.44	0.49	0.42	0.45	0.39	0.40
	O3-svt	0.43	0.42	0.45	0.45	0.42	0.42	0.40	0.39
	O4-svt	0.43	0.00	0.46	0.00	0.42	0.00	0.40	0.00
	O5-svt	0.02	0.06	0.02	0.09	0.02	0.07	0.02	0.04
	O6-svt	0.45	0.45	0.56	0.52	0.46	0.46	0.42	0.41
	O2-O3'	0.0		0.0		0.0		0.0	
	O6-O3'	0.0		0.0		0.0		0.0	
	O5-O3'	1.2		1.3		1.7		1.6	
	O1-O3'	16.3		17.8		23.5		24.0	
	O2-O6'	18.5		15.5		58.6		73.5	
	O5-O6'	0.0		0.0		0.0		0.0	
	O6-O6'	0.0		0.0		0.1		0.1	
	O1-O6'	2.5		2.8		7.7		2.3	
	O2-O3	72.7	41.8	83.0	47.7	51.1	24.2	50.9	25.5
	O2-O1	10.7	21.4	15.4	32.7	7.4	11.0	7.0	10.5
	O2-O5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	O3-O4	74.7	0.00	85.3	0.00	51.6	0.00	52.1	0.00
	O4-O6	59.5	0.00	80.1	0.00	12.1	0.00	13.9	0.00
GA	O6-O5	10.4	14.5	7.7	15.4	7.8	12.3	8.1	13.6
	O1-svt	0.03	0.44	0.05	0.50	0.04	0.48	0.03	0.42
	O2-svt	0.41	0.43	0.44	0.48	0.43	0.46	0.39	0.40
	O3-svt	0.42	0.41	0.46	0.43	0.44	0.45	0.40	0.40
	O4-svt	0.43	0.00	0.47	0.00	0.43	0.00	0.39	0.00
	O5-svt	0.03	0.06	0.04	0.11	0.04	0.06	0.03	0.04
	O6-svt	0.46	0.45	0.55	0.53	0.49	0.46	0.43	0.40

(a)

Sug.	A.	$W^{0.5}_{1.4}$		$W^{0.8}_{0.9}$		$W^{0.7}_{0.6}$		$W^{0.3}_{0.5}$	
MG	O2-O3'	7.2		4.8		28.1		60.0	
	O6-O3'	2.9		3.3		3.3		0.6	
	O5-O3'	76.9		77.7		63.1		33.8	
	O1-O3'	54.9		55.2		47.9		36.0	
	O2-O6'	1.4		0.9		1.7		0.6	
	O5-O6'	0.2		0.1		1.1		2.8	
	O6-O6'	1.3		0.6		12.1		19.0	
	O1-O6'	0.1		0.2		0.6		0.6	
	O2-O3	59.2	79.9	71.5	88.4	34.0	60.1	35.1	60.4
	O2-O1	9.7	21.6	5.0	37.7	8.5	8.0	8.3	8.8
	O2-O5	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00
	O3-O4	70.6	0.00	81.2	0.00	47.0	0.00	48.1	0.00
	O4-O6	54.3	0.00	73.6	0.00	12.6	0.00	12.6	0.00
	O6-O5	13.3	31.4	12.8	43.4	6.8	13.3	6.4	14.3
	O1-svt	0.03	0.46	0.04	0.53	0.03	0.47	0.02	0.42
	O2-svt	0.43	0.44	0.49	0.49	0.43	0.45	0.39	0.41
	O3-svt	0.42	0.43	0.48	0.43	0.43	0.42	0.41	0.39
	O4-svt	0.41	0.00	0.47	0.00	0.46	0.00	0.40	0.00
	O5-svt	0.02	0.05	0.02	0.10	0.02	0.06	0.01	0.04
	O6-svt	0.46	0.45	0.54	0.54	0.48	0.48	0.41	0.42
MA	O2-O3'	0.0		0.0		0.7		3.6	
	O6-O3'	0.0		0.0		0.0		0.0	
	O5-O3'	1.3		1.6		2.4		1.2	
	O1-O3'	16.2		18.0		27.7		30.0	
	O2-O6'	3.7		2.8		7.0		3.6	
	O5-O6'	0.0		0.0		0.0		0.1	
	O6-O6'	0.0		0.0		0.1		0.1	
	O1-O6'	0.4		0.6		1.6		1.2	
	O2-O3	58.3	45.3	71.3	56.6	35.0	24.9	36.0	25.9
	O2-O1	13.3	20.1	20.0	30.7	8.7	10.4	8.9	10.6
	O2-O5	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00
	O3-O4	72.4	0.00	83.0	0.00	48.7	0.00	49.4	0.00
	O4-O6	67.8	0.00	86.1	0.00	15.1	0.00	15.6	0.00
	O6-O5	7.8	28.8	5.4	38.6	7.3	13.4	9.1	14.2
	O1-svt	0.03	0.44	0.05	0.52	0.03	0.49	0.02	0.41
	O2-svt	0.43	0.43	0.49	0.49	0.45	0.47	0.41	0.41
	O3-svt	0.43	0.41	0.47	0.46	0.44	0.44	0.40	0.41
	O4-svt	0.42	0.00	0.46	0.00	0.43	0.00	0.40	0.00
	O5-svt	0.03	0.06	0.05	0.10	0.03	0.06	0.02	0.04
	O6-svt	0.47	0.44	0.53	0.54	0.46	0.48	0.42	0.41

(b)

Table 4.S.9: H-bonding analysis for the four disaccharides considered (GG, GA, MG, MA) based on the 60 ns sampling phase of the corresponding LEUS simulations in the series of pysical solvents at 298.15 K and 1 bar. For each dimer, the first panel reports the occurence f_i of the inter-ring hydrogen bonds. The second panel refers to the the occurence f_i of the intra-ring hydrogen bonds considered separately for the two rings (the first column corresponds to the data for the first ring, the second for the second ring). In parenthesis the percentage of the hydrogen bonds that involve the first atom as a donor is reported. The third and last panel describes the average number n_s of solute-solvent hydrogen bonds for each hydroxyl group of the two rings (the first column corresponds to the data for the first ring, the second for the second ring). The calculation of all H-bonds involved appropriate reweighting of the configurations so as to remove the effect of the biasing potential energy term. The data points follow the series of artificial solvents (Table 4.2) in order of decreasing dielectric permittivity. The data is illustrated graphically in Figure 4.13 and Suppl. Mat. Figures 4.S.5-4.S.7.

Sug.	A.	$W^{1.1}_{0.9}$		$W^{1.0}_{1.0}$		$W^{0.8}_{1.1}$		$W^{0.7}_{1.2}$		$W^{0.6}_{1.4}$		$W^{0.5}_{1.5}$	
	O2-O3'	0.0		0.0		0.1		0.1		0.1		0.1	
	O6-O3'	1.6		1.4		2.1		2.6		2.6		2.2	
	O5-O3'	68.5		65.2		65.9		68.1		66.9		73.9	
	O1-O3'	50.4		50.1		49.6		51.4		49.2		53.9	
	O2-O6'	1.1		1.6		2.5		3.6		3.5		8.2	
	O5-O6'	0.0		0.0		0.0		0.0		0.0		0.0	
	O6-O6'	0.0		0.0		0.0		0.0		0.0		0.1	
	O1-O6'	0.3		0.4		0.7		0.7		0.8		1.6	
	O2-O3	34.6	40.3	33.4	37.7	37.2	44.4	38.9	46.7	36.5	45.3	43.7	53.1
	O2-O1	4.3	4.8	3.5	5.2	3.8	5.3	4.4	5.4	4.1	4.9	5.8	6.9
	O2-O5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	O3-O4	38.4	0.00	37.4	0.00	39.8	0.00	41.6	0.00	39.8	0.00	46.3	0.00
GG	O4-O6	3.6	0.00	2.0	0.00	3.4	0.00	4.1	0.00	2.8	0.00	9.0	0.00
	O6-O5	4.7	10.5	3.6	8.4	4.5	8.6	4.6	8.4	3.9	8.2	4.8	10.5
	O1-svt	0.03	0.65	0.10	2.13	0.03	0.53	0.03	0.52	0.03	0.48	0.03	0.45
	O2-svt	0.57	0.58	1.86	1.79	0.47	0.50	0.47	0.49	0.45	0.47	0.43	0.44
	O3-svt	0.56	0.41	1.74	1.22	0.49	0.47	0.46	0.46	0.45	0.43	0.44	0.43
	O4-svt	0.55	0.00	1.76	0.00	0.48	0.00	0.45	0.00	0.43	0.00	0.43	0.00
	O5-svt	0.01	0.15	0.03	0.42	0.02	0.10	0.01	0.08	0.01	0.07	0.02	0.06
	O6-svt	0.65	0.64	2.20	2.17	0.53	0.54	0.52	0.51	0.49	0.49	0.45	0.45
	O2-O3'	0.0		0.0		0.0		0.0		0.0		0.0	
	O6-O3'	0.0		0.0		0.0		0.0		0.0		0.0	
	O5-O3'	1.4		0.9		0.7		0.8		0.5		0.8	
	O1-O3'	16.2		12.4		12.3		12.9		12.1		14.8	
	O2-O6'	3.5		3.1		5.5		7.4		6.3		11.4	
	O5-O6'	0.0		0.0		0.0		0.0		0.0		0.0	
	O6-O6'	0.0		0.0		0.0		0.0		0.0		0.0	
	O1-O6'	0.8		0.8		1.2		1.1		1.0		0.9	
	O2-O3	34.3	15.3	33.1	15.5	38.3	17.1	40.0	17.7	38.8	16.8	46.5	21.4
	O2-O1	4.5	6.4	3.5	5.8	4.2	7.3	4.2	7.6	4.2	6.5	5.8	9.0
	O2-O5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	O3-O4	38.3	0.00	35.7	0.00	40.1	0.00	41.5	0.00	39.6	0.00	46.8	0.00
GA	O4-O6	3.1	0.00	3.0	0.00	3.5	0.00	4.1	0.00	5.2	0.00	8.5	0.00
	O6-O5	6.1	10.6	4.8	8.5	4.9	8.7	5.2	8.9	4.5	8.2	5.9	11.0
	O1-svt	0.06	0.69	0.20	2.10	0.05	0.52	0.04	0.50	0.04	0.48	0.03	0.45
	O2-svt	0.56	0.59	1.81	1.89	0.47	0.50	0.46	0.48	0.44	0.46	0.41	0.44
	O3-svt	0.58	0.53	1.74	1.78	0.48	0.46	0.46	0.46	0.44	0.44	0.43	0.42
	O4-svt	0.56	0.00	1.76	0.00	0.48	0.00	0.47	0.00	0.45	0.00	0.43	0.00
	O5-svt	0.05	0.14	0.14	0.45	0.04	0.09	0.04	0.08	0.03	0.07	0.03	0.06
	O6-svt	0.68	0.65	2.22	2.18	0.54	0.54	0.53	0.52	0.50	0.50	0.46	0.46

(a)

Sug.	A.	$W^{1.1}_{0.9}$		$W^{1.0}_{1.0}$		$W^{0.8}_{1.1}$		$W^{0.7}_{1.2}$		$W^{0.6}_{1.4}$		$W^{0.5}_{1.5}$	
MG	O2-O3'	0.9		1.2		2.9		3.1		3.5		5.5	
	O6-O3'	1.0		1.7		1.8		2.3		2.3		2.4	
	O5-O3'	75.2		71.9		68.5		71.8		69.7		76.2	
	O1-O3'	55.3		53.2		50.6		51.1		50.4		54.2	
	O2-O6'	0.1		0.3		0.6		0.7		1.1		1.4	
	O5-O6'	0.0		0.0		0.1		0.1		0.0		0.1	
	O6-O6'	0.0		0.0		0.1		0.2		0.1		0.6	
	O1-O6'	0.1		0.0		0.1		0.1		0.1		0.1	
	O2-O3	24.3	39.9	22.3	38.0	23.3	44.3	25.2	46.5	23.3	44.0	29.7	55.1
	O2-O1	6.6	4.8	6.4	4.8	5.5	5.5	6.3	5.4	5.8	5.3	7.3	6.5
	O2-O5	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00
	O3-O4	37.2	0.00	36.6	0.00	37.7	0.00	37.9	0.00	37.0	0.00	42.8	0.00
	O4-O6	5.0	0.00	3.0	0.00	4.1	0.00	5.2	0.00	4.9	0.00	6.2	0.00
	O6-O5	3.6	10.2	3.7	7.7	4.5	7.8	4.5	8.4	3.7	7.4	5.3	10.3
	O1-svt	0.03	0.66	0.09	2.13	0.04	0.53	0.03	0.51	0.03	0.49	0.02	0.46
	O2-svt	0.65	0.61	1.92	1.81	0.50	0.52	0.47	0.47	0.45	0.47	0.42	0.44
	O3-svt	0.59	0.41	1.78	1.13	0.47	0.46	0.45	0.44	0.45	0.45	0.42	0.42
	O4-svt	0.57	0.00	1.77	0.00	0.47	0.00	0.44	0.00	0.44	0.00	0.42	0.00
	O5-svt	0.01	0.14	0.04	0.42	0.02	0.10	0.02	0.08	0.02	0.07	0.02	0.06
	O6-svt	0.65	0.67	2.20	2.12	0.55	0.52	0.51	0.48	0.48	0.47	0.45	0.44
MA	O2-O3'	0.0		0.0		0.0		0.0		0.0		0.0	
	O6-O3'	0.0		0.0		0.0		0.0		0.0		0.0	
	O5-O3'	1.5		0.9		0.9		0.9		0.7		1.0	
	O1-O3'	18.8		15.1		12.7		12.6		11.4		13.6	
	O2-O6'	0.5		0.5		1.1		1.3		1.1		2.3	
	O5-O6'	0.0		0.0		0.0		0.0		0.0		0.0	
	O6-O6'	0.0		0.0		0.0		0.0		0.0		0.0	
	O1-O6'	0.4		0.2		0.2		0.3		0.2		0.2	
	O2-O3	26.0	15.0	22.7	14.9	24.3	16.6	25.5	17.9	24.6	16.7	31.4	20.9
	O2-O1	8.3	6.0	7.4	6.1	6.0	6.5	6.1	7.0	5.6	6.8	7.2	8.6
	O2-O5	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00
	O3-O4	37.6	0.00	36.1	0.00	38.0	0.00	38.1	0.00	36.3	0.00	43.1	0.00
	O4-O6	3.8	0.00	2.2	0.00	5.3	0.00	6.3	0.00	5.4	0.00	10.0	0.00
	O6-O5	6.8	9.9	5.2	8.6	4.8	8.6	5.4	8.8	4.9	7.6	6.5	10.9
	O1-svt	0.05	0.66	0.15	2.11	0.05	0.52	0.05	0.51	0.04	0.47	0.04	0.44
	O2-svt	0.56	0.58	1.92	1.89	0.48	0.49	0.47	0.49	0.47	0.46	0.43	0.44
	O3-svt	0.57	0.52	1.78	1.78	0.49	0.47	0.47	0.46	0.44	0.45	0.42	0.42
	O4-svt	0.54	0.00	1.77	0.00	0.46	0.00	0.46	0.00	0.44	0.00	0.42	0.00
	O5-svt	0.05	0.17	0.17	0.44	0.05	0.10	0.05	0.10	0.04	0.07	0.03	0.06
	O6-svt	0.65	0.65	2.22	2.18	0.55	0.55	0.52	0.53	0.49	0.49	0.47	0.47

(b)

Table 4.S.10: H-bonding analysis for the four disaccharides considered (GG, GA, MG, MA) based on the 60 ns sampling phase of the corresponding LEUS simulations in the series of pysical solvents at 298.15 K and 1 bar. For each dimer, the first panel reports the occurence f_i of the inter-ring hydrogen bonds. The second panel refers to the the occurence f_i of the intra-ring hydrogen bonds considered separately for the two rings (the first column corresponds to the data for the first ring, the second for the second ring). In parenthesis the percentage of the hydrogen bonds that involve the first atom as a donor is reported. The third and last panel describes the average number n_s of solute-solvent hydrogen bonds for each hydroxyl group of the two rings (the first column corresponds to the data for the first ring, the second for the second ring). The calculation of all H-bonds involved appropriate reweighting of the configurations so as to remove the effect of the biasing potential energy term. The data points follow the series of artificial solvents (Table 4.2) in order of decreasing dielectric permittivity. The data is illustrated graphically in Figure 4.13 and Suppl. Mat. Figures 4.S.5-4.S.7.

Sug.	A.	$W^{1.5}_{0.4}$		$W^{1.2}_{0.5}$		$W^{1.1}_{0.5}$		$W^{0.8}_{0.7}$		$W^{0.6}_{0.9}$		$W^{0.4}_{1.3}$	
GG	O2-O3'	0.0		0.1		0.3		0.3		0.6		0.6	
	O6-O3'	2.3		3.2		4.5		4.1		3.9		4.2	
	O5-O3'	83.0		84.8		85.6		85.2		85.6		85.7	
	O1-O3'	63.5		59.5		58.7		58.0		56.4		56.3	
	O2-O6'	2.5		12.8		18.5		26.6		33.6		34.8	
	O5-O6'	0.0		0.0		0.1		0.1		0.2		0.1	
	O6-O6'	0.0		0.1		0.2		0.2		0.4		0.5	
	O1-O6'	0.8		2.3		4.3		5.5		5.9		5.4	
	O2-O3	62.5	72.8	63.3	71.5	53.9	62.6	60.2	71.3	56.0	67.1	60.4	70.3
	O2-O1	11.4	13.7	11.8	15.4	8.0	11.3	9.2	13.9	8.9	11.7	9.9	13.6
	O2-O5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	O3-O4	64.2	0.00	65.3	0.00	55.8	0.00	61.8	0.00	59.2	0.00	63.2	0.00
	O4-O6	31.8	0.00	40.4	0.00	11.0	0.00	27.9	0.00	17.7	0.00	31.7	0.00
	O6-O5	6.9	16.3	6.9	15.9	8.0	29.2	7.8	22.6	7.3	22.5	7.5	17.9
	O1-svt	0.04	0.69	0.05	0.65	0.07	0.70	0.04	0.54	0.03	0.46	0.03	0.43
	O2-svt	0.53	0.60	0.50	0.58	0.53	0.62	0.45	0.49	0.40	0.42	0.40	0.43
	O3-svt	0.54	0.53	0.53	0.47	0.55	0.40	0.44	0.42	0.43	0.43	0.40	0.39
	O4-svt	0.55	0.00	0.55	0.00	0.56	0.00	0.44	0.00	0.42	0.00	0.40	0.00
	O5-svt	0.01	0.15	0.02	0.14	0.02	0.18	0.02	0.09	0.02	0.06	0.02	0.05
	O6-svt	0.63	0.64	0.63	0.62	0.67	0.66	0.50	0.51	0.46	0.43	0.43	0.42
GA	O2-O3'	0.0		0.0		0.0		0.0		0.0		0.0	
	O6-O3'	0.0		0.0		0.0		0.0		0.0		0.0	
	O5-O3'	3.4		2.2		2.3		1.9		1.7		1.5	
	O1-O3'	26.0		24.2		23.3		21.7		21.6		21.7	
	O2-O6'	5.8		19.2		30.7		39.5		42.2		47.9	
	O5-O6'	0.0		0.0		0.0		0.0		0.0		0.0	
	O6-O6'	0.0		0.0		0.1		0.1		0.1		0.1	
	O1-O6'	1.2		3.5		3.3		5.5		5.0		7.0	
	O2-O3	65.4	36.1	65.6	37.2	55.9	30.5	63.8	35.6	59.1	32.7	64.3	35.1
	O2-O1	9.9	16.7	11.4	18.2	10.4	12.9	10.7	16.4	10.1	13.9	9.8	16.6
	O2-O5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	O3-O4	65.6	0.00	67.0	0.00	56.5	0.00	64.3	0.00	60.1	0.00	64.7	0.00
	O4-O6	37.8	0.00	42.6	0.00	6.6	0.00	31.2	0.00	21.8	0.00	33.7	0.00
	O6-O5	10.8	14.9	10.7	14.2	16.3	28.1	12.7	20.0	13.1	21.5	11.3	15.5
	O1-svt	0.06	0.64	0.07	0.63	0.09	0.67	0.05	0.50	0.03	0.45	0.03	0.43
	O2-svt	0.55	0.58	0.49	0.58	0.49	0.62	0.42	0.48	0.42	0.43	0.41	0.42
	O3-svt	0.55	0.54	0.53	0.50	0.57	0.54	0.44	0.42	0.43	0.43	0.42	0.41
	O4-svt	0.54	0.00	0.53	0.00	0.55	0.00	0.45	0.00	0.42	0.00	0.41	0.00
	O5-svt	0.05	0.15	0.06	0.15	0.09	0.16	0.05	0.10	0.03	0.06	0.03	0.05
	O6-svt	0.67	0.64	0.63	0.63	0.69	0.64	0.50	0.48	0.47	0.46	0.45	0.42

(a)

Sug.	A.	$W^{1.5}_{0.4}$		$W^{1.2}_{0.5}$		$W^{1.1}_{0.5}$		$W^{0.8}_{0.7}$		$W^{0.6}_{0.9}$		$W^{0.4}_{1.3}$	
MG	O2-O3'	0.8		4.2		7.7		13.7		15.2		19.5	
	O6-O3'	2.1		2.7		3.0		3.6		3.0		3.2	
	O5-O3'	82.3		81.6		79.3		73.9		73.0		69.8	
	O1-O3'	64.6		59.8		57.8		54.8		54.0		51.6	
	O2-O6'	0.1		0.6		0.9		1.1		1.6		1.9	
	O5-O6'	0.0		0.1		0.3		0.6		0.5		0.9	
	O6-O6'	0.0		0.5		1.6		3.4		4.9		6.4	
	O1-O6'	0.1		0.2		0.2		0.2		0.3		0.3	
	O2-O3	47.9	72.8	50.1	75.2	44.4	62.0	48.2	71.5	44.9	66.8	47.3	72.7
	O2-O1	9.4	14.4	10.3	15.1	13.9	11.7	10.9	13.5	10.9	11.7	10.0	13.3
	O2-O5	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00
	O3-O4	63.0	0.00	63.5	0.00	55.3	0.00	62.9	0.00	57.5	0.00	62.0	0.00
	O4-O6	40.3	0.00	39.5	0.00	11.7	0.00	40.0	0.00	25.5	0.00	39.4	0.00
	O6-O5	9.8	21.6	10.7	22.8	10.0	29.8	7.8	25.5	8.5	24.1	8.7	23.3
	O1-svt	0.03	0.67	0.05	0.63	0.06	0.71	0.04	0.54	0.03	0.47	0.02	0.43
	O2-svt	0.60	0.61	0.54	0.59	0.58	0.61	0.45	0.49	0.43	0.44	0.40	0.41
	O3-svt	0.61	0.58	0.53	0.46	0.57	0.41	0.47	0.43	0.42	0.41	0.41	0.40
	O4-svt	0.60	0.00	0.52	0.00	0.54	0.00	0.45	0.00	0.43	0.00	0.41	0.00
	O5-svt	0.02	0.14	0.02	0.14	0.03	0.17	0.02	0.09	0.02	0.06	0.02	0.05
	O6-svt	0.65	0.69	0.62	0.63	0.65	0.70	0.51	0.51	0.46	0.46	0.42	0.42
MA	O2-O3'	0.0		0.1		0.1		0.1		0.2		0.3	
	O6-O3'	0.0		0.0		0.0		0.0		0.0		0.0	
	O5-O3'	3.3		3.0		2.3		2.3		2.3		2.1	
	O1-O3'	27.1		24.9		24.5		23.0		22.0		22.5	
	O2-O6'	0.8		2.6		4.9		5.5		7.2		8.8	
	O5-O6'	0.0		0.0		0.0		0.0		0.0		0.0	
	O6-O6'	0.0		0.0		0.1		0.1		0.1		0.1	
	O1-O6'	0.6		0.7		0.9		0.8		0.9		0.9	
	O2-O3	50.0	36.6	50.8	37.8	44.9	30.5	50.3	36.2	46.5	33.6	48.7	35.3
	O2-O1	13.1	16.6	12.6	17.3	16.4	12.5	14.1	15.5	14.3	13.7	12.7	15.7
	O2-O5	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00
	O3-O4	63.1	0.00	65.0	0.00	54.8	0.00	62.1	0.00	57.0	0.00	62.1	0.00
	O4-O6	40.8	0.00	52.2	0.00	12.9	0.00	36.3	0.00	23.0	0.00	44.8	0.00
	O6-O5	10.8	22.3	9.1	22.7	16.5	28.6	12.7	25.3	14.0	24.6	9.8	23.0
	O1-svt	0.05	0.65	0.06	0.62	0.08	0.68	0.05	0.49	0.03	0.45	0.03	0.43
	O2-svt	0.59	0.61	0.52	0.56	0.57	0.61	0.44	0.47	0.43	0.44	0.42	0.42
	O3-svt	0.57	0.57	0.55	0.50	0.59	0.53	0.46	0.42	0.43	0.42	0.42	0.41
	O4-svt	0.58	0.00	0.54	0.00	0.53	0.00	0.45	0.00	0.42	0.00	0.40	0.00
	O5-svt	0.06	0.14	0.07	0.15	0.08	0.18	0.05	0.09	0.03	0.05	0.03	0.05
	O6-svt	0.69	0.67	0.62	0.64	0.66	0.70	0.50	0.51	0.48	0.47	0.44	0.43

(b)

Table 4.S.11: H-bonding analysis for the four disaccharides considered (GG, GA, MG, MA) based on the 60 ns sampling phase of the corresponding LEUS simulations in the series of pysical solvents at 298.15 K and 1 bar. For each dimer, the first panel reports the occurence f_i of the inter-ring hydrogen bonds. The second panel refers to the the occurence f_i of the intra-ring hydrogen bonds considered separately for the two rings (the first column corresponds to the data for the first ring, the second for the second ring). In parenthesis the percentage of the hydrogen bonds that involve the first atom as a donor is reported. The third and last panel describes the average number n_s of solute-solvent hydrogen bonds for each hydroxyl group of the two rings (the first column corresponds to the data for the first ring, the second for the second ring). The calculation of all H-bonds involved appropriate reweighting of the configurations so as to remove the effect of the biasing potential energy term. The data points follow the series of artificial solvents (Table 4.2) in order of decreasing dielectric permittivity. The data is illustrated graphically in Figure 4.13 and Suppl. Mat. Figures 4.S.5-4.S.7.
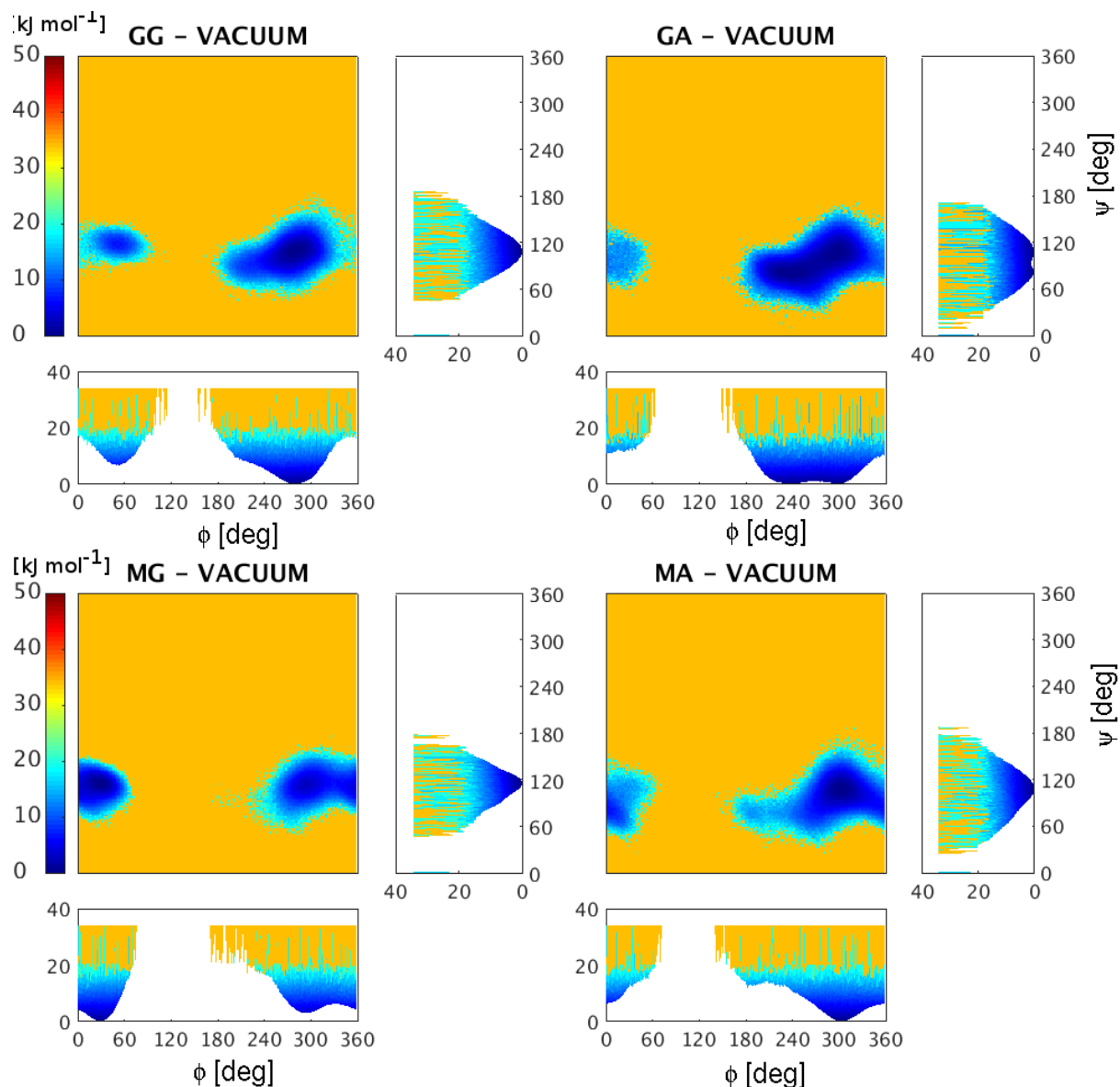
Figure 4.S.1: Free-energy maps $G(\phi, \psi)$ in the space of the glycosidic dihedral angles ϕ and ψ for the four disaccharides considered (Figure 4.1) in vacuum at 298.15 K. The maps are based on 1 μs SD simulations in vacuum. They are anchored to $G = 0$ kJ mol^{-1} at the location of their global minimum, and the value of G at grid points that were never visited during the simulations is arbitrarily set to $G_{max} = 30$ kJ mol^{-1}. The maps are shown in the ϕ,ψ-plane as viewed from the top along with lateral projections in the ϕ,G- and ψ,G-planes.
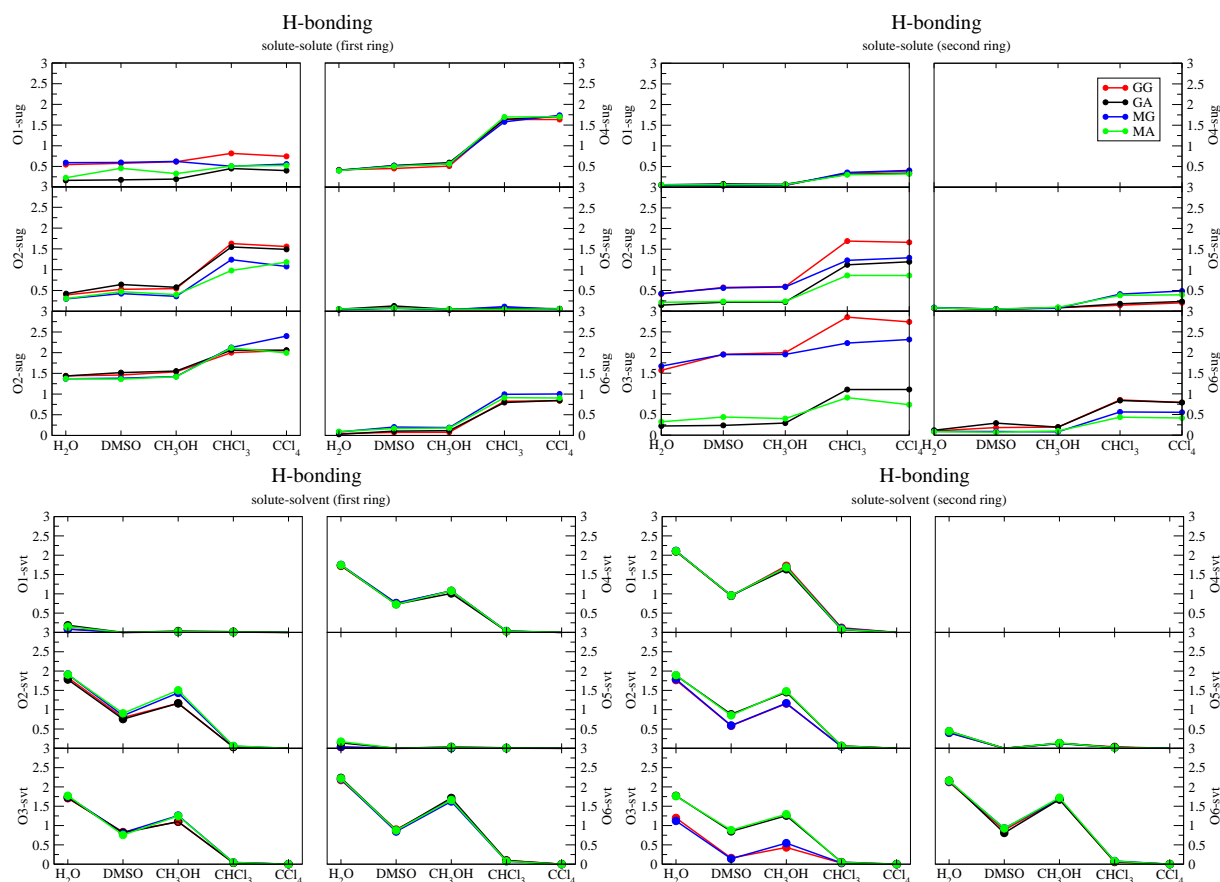
Figure 4.S.2: Numbers of H-bonds involving specific solute atoms for the four disaccharides considered in the physical solvents at 298.15 K and 1 bar. The results are based on the 60 ns sampling phases of the corresponding LEUS simulations. In the top panels, the total average number n_i of solute-solute H-bonds is reported separately for each hydroxyl group of the two rings. In the bottom panels, the average number n_s of solute-solvent H-bonds is reported separately for each hydroxyl group of the two rings. All the hydroxyl groups in the two rings were considered for the analysis. The figure does not distinguish between donors and acceptors. The calculation of H-bonds involved appropriate reweighting of the configurations so as to remove the effect of the biasing potential energy term. The graphs on left side describe H-bonds involving the atoms the first ring, while those on the right side describe the H-bonds involving the atoms of the second ring. The data points follow the series of physical solvents (Table 4.1) in order of decreasing polarity. The different colors refer to the four disaccharides (Figure 4.1), namely GG, MG, GA, MA. The data is reported numerically in Suppl. Mat. Table 4.S.3.
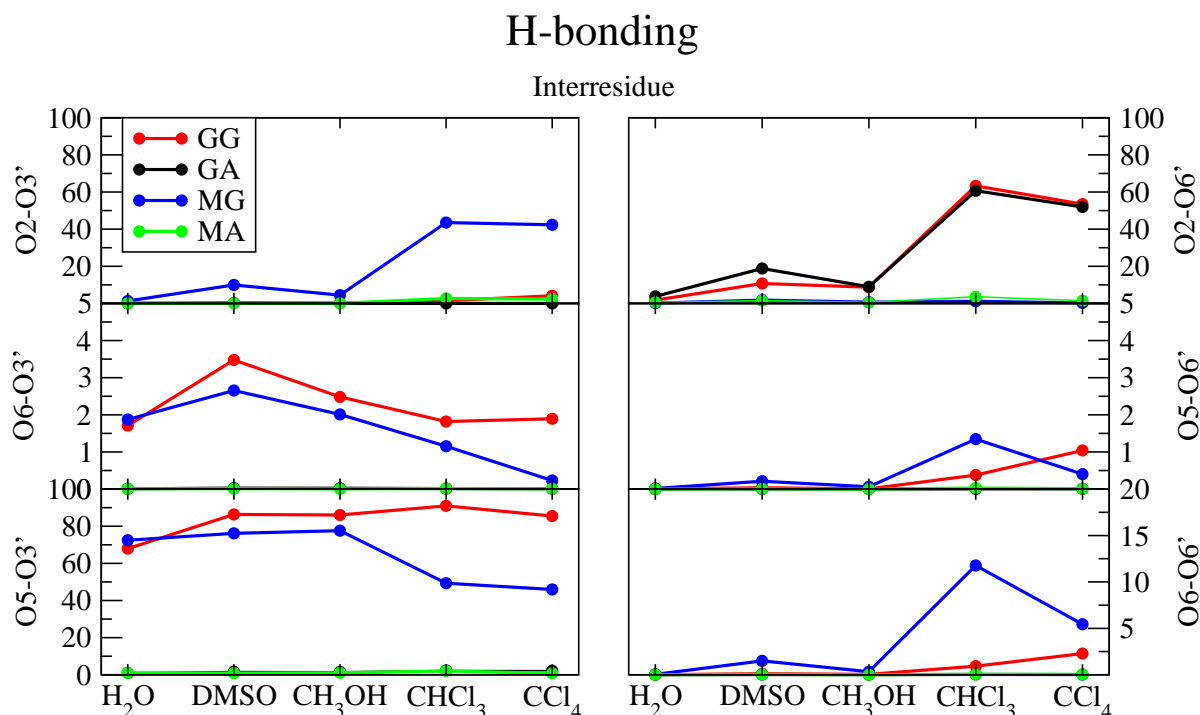
Figure 4.S.3: Occurrences (in percent) of interresidue H-bonds for the four disaccharides considered in the series of physical solvents at 298.15 K and 1 bar. The results are based on the 60 ns sampling phases of the corresponding LEUS simulations. All pairs of hydroxyl groups involving the two rings were considered for the analysis, but only H-bonds with occurrences of at least 0.1 % are reported. The figure does not distinguish between donors and acceptors. The calculation of H-bonds involved appropriate reweighting of the configurations so as to remove the effect of the biasing potential energy term. The panels on the left side describe H-bonds involving the atoms HO'_3 and O'_3 of the second ring, while those on the right side describe the H-bonds involving the atoms HO'_6 and O'_6. The data points follow the series of physical solvents (Table 4.1) in order of decreasing polarity. The different colors refer to the four disaccharides (Figure 4.1), namely GG, MG, GA, MA. The data is reported numerically in Suppl. Mat. Table 4.S.3.
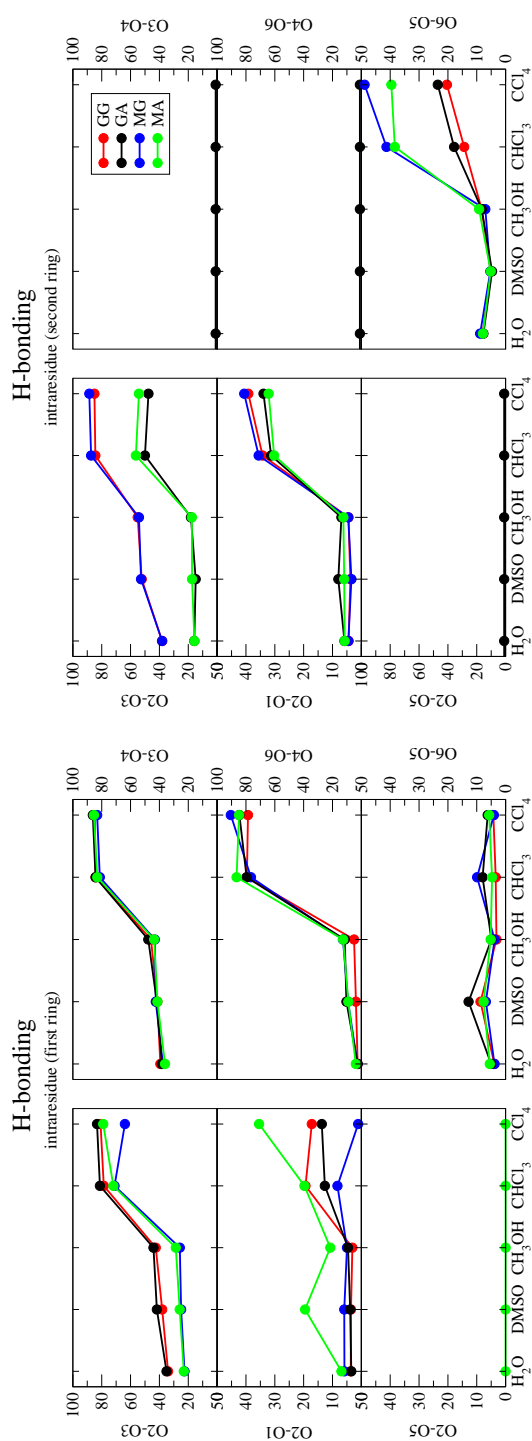
Figure 4.S.4: Occurrences (in percent) of intraresidue H-bonds for the four disaccharides considered in the series of physical solvents at 298.15 K and 1 bar. The results based on the 60 ns sampling phases of the corresponding LEUS simulations. All the hydroxyl groups in the two rings were considered for the analysis, but only H-bonds with occurrences of at least 0.1 % are reported. The figure does not distinguish between donors and acceptors. The calculation of H-bonds involved appropriate reweighting of the configurations so as to remove the effect of the biasing potential energy term. The graphs on left side describe H-bonds involving the atoms the first ring, while those on the right side describe the hydrogen bonds involving the atoms of the second ring. The data points follow the series of physical solvents (Table 4.1) in order of decreasing polarity. The different colors refer to the four disaccharides (Figure 4.1), namely GG, MG, GA, MA. The data is reported numerically in Suppl. Mat. Table 4.S.3.
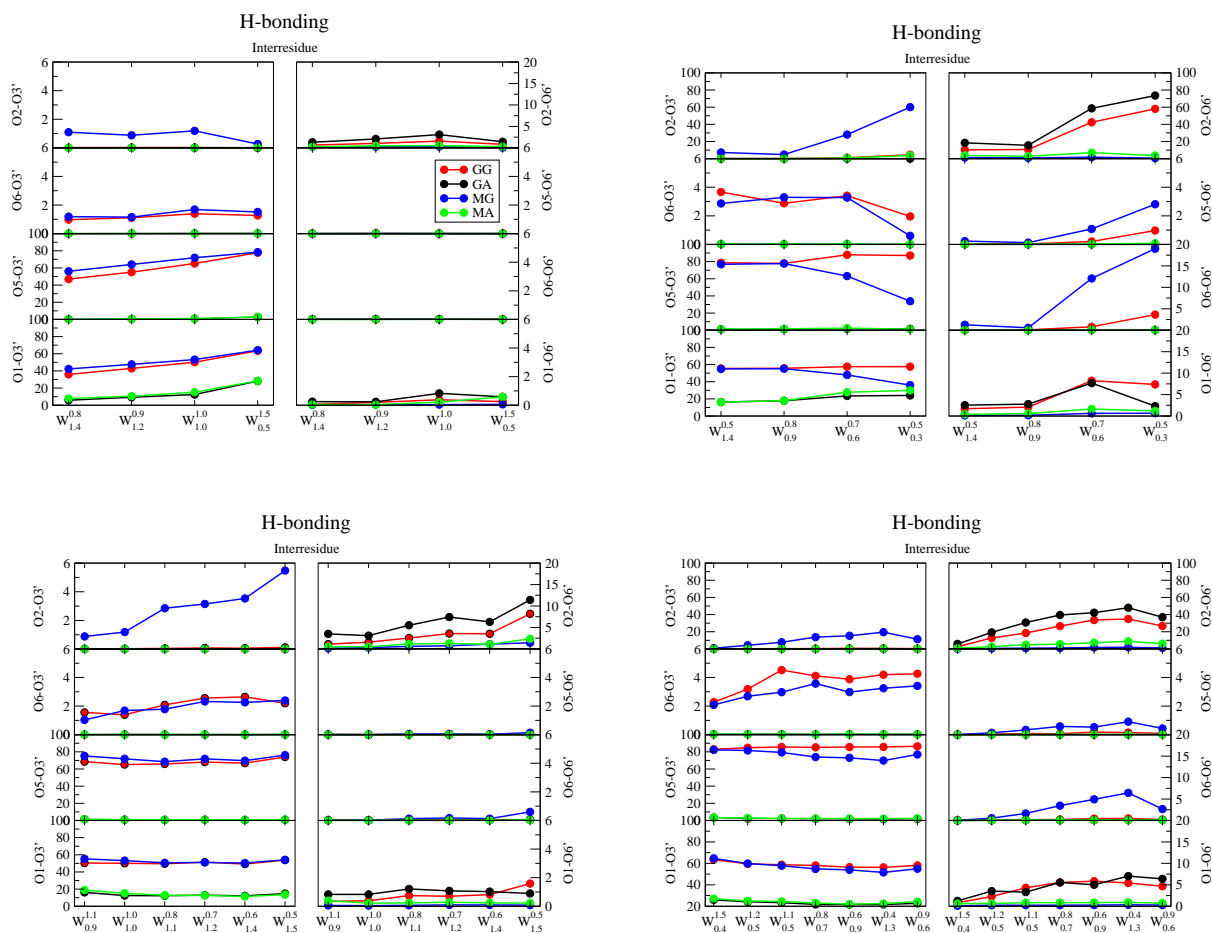
Figure 4.S.5: Occurrences (in percent) of interresidue H-bonds for the four disaccharides considered in the series of artificial solvents at 298.15 K and 968 kg·m^{-3}. The results are based on the 60 ns sampling phases of the corresponding LEUS simulations. All pairs of hydroxyl groups involving the two rings were considered for the analysis, but only H-bonds with occurrences of at least 0.1 % are reported. The figure does not distinguish between donors and acceptors. The calculation of H-bonds involved appropriate reweighting of the configurations so as to remove the effect of the biasing potential energy term. The grphs on the left side describe H-bonds involving the atoms HO$_3'$ and O$_3'$ of the second ring, while those on the right side describe the H-bonds involving the atoms HO$_6'$ and O$_6'$. The data points follow the series of artificial solvents (Table 4.2): in the top graphs the series considered are the S_p^H (on the left) and the S_p^h (on the right) in order of decreasing dielectric permittivity; in the bottom graphs the series considered are the S_h^P (on the left) and the S_h^p (on the right) in order of decreasing H-bonding capacity. The different colors refer to the four disaccharides (Figure 4.1), namely GG, MG, GA, MA. The data is reported numerically in Suppl. Mat. Tables 4.S.8, 4.S.9, 4.S.10 and 4.S.11.
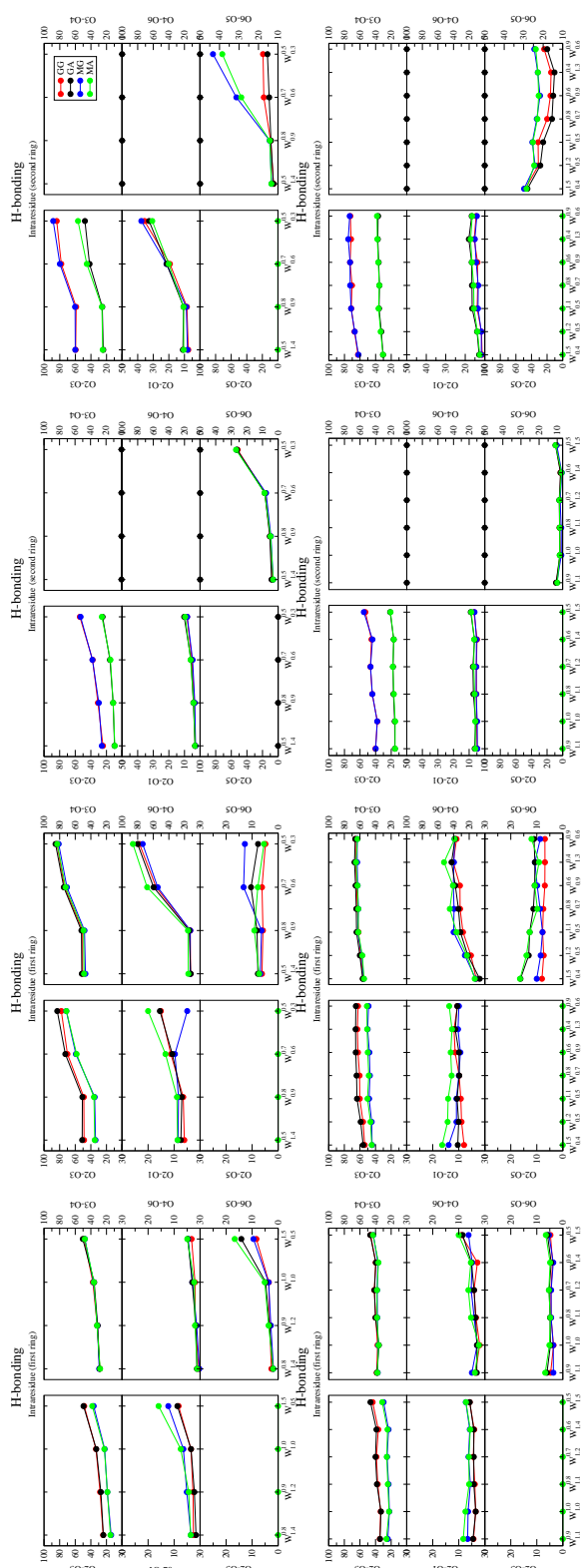
Figure 4.S.6: Occurrences (in percent) of intraresidue H-bonds for the four disaccharides considered in the series of artificial solvents at 298.15 K and 968 kg·m^{-3}. The results based on the 60 ns sampling phases of the corresponding LEUS simulations. All the hydroxyl groups in the two rings were considered for the analysis, but only H-bonds with occurrences of at least 0.1 % are reported. The figure does not distinguish between donors and acceptors. The calculation of H-bonds involved appropriate reweighting of the configurations so as to remove the effect of the biasing potential energy term. The graph on left side describes H-bonds involving the atoms the first ring, while those on the right side describe the hydrogen bonds involving the atoms of the second ring. The data points follow the series of artificial solvents (Table 4.2): in the top graphs the series considered are the S_p^H (on the left) and the S_p^h (on the right) in order of decreasing dielectric permittivity; in the bottom graphs the series considered are the S_h^P (on the left) and the S_h^p (on the right) in order of decreasing H-bonding capacity. The different colors refer to the four disaccharides (Figure 4.1), namely GG, MG, GA, MA. The data is reported numerically in Suppl. Mat. Tables 4.S.8, 4.S.9, 4.S.10 and 4.S.11.

Figure 4.S.7: Numbers of solute-solvent H-bonds involving specific solute atoms for the four disaccharides considered in the artificial solvents at 298.15 K and 968 kg·m^{-3}. The results are based on the 60 ns sampling phases of the corresponding LEUS simulations. The average number n_s of solute-solvent H-bonds is reported separately for the two rings. All the hydroxyl groups in the two rings were considered for the analysis. The figure does not distinguish between donors and acceptors. The calculation of H-bonds involved appropriate reweighting of the configurations so as to remove the effect of the biasing potential energy term. The graphs on left side describe H-bonds involving the atoms the first ring, while those on the right side describe the H-bonds involving the atoms of the second ring. The data points follow the series of artificial solvents (Table 4.2): in the top graphs the series considered are the S_p^H (on the left) and the S_p^h (on the right) in order of decreasing permittivity; in the bottom graphs the series considered are the S_h^P (on the left) and the S_h^p (on the right) in order of decreasing H-bonding capacity. The different colors refer to the four disaccharides (Figure 4.1), namely GG, MG, GA, MA. The data is reported numerically in Suppl. Mat. Tables 4.S.8, 4.S.9, 4.S.10 and 4.S.11.
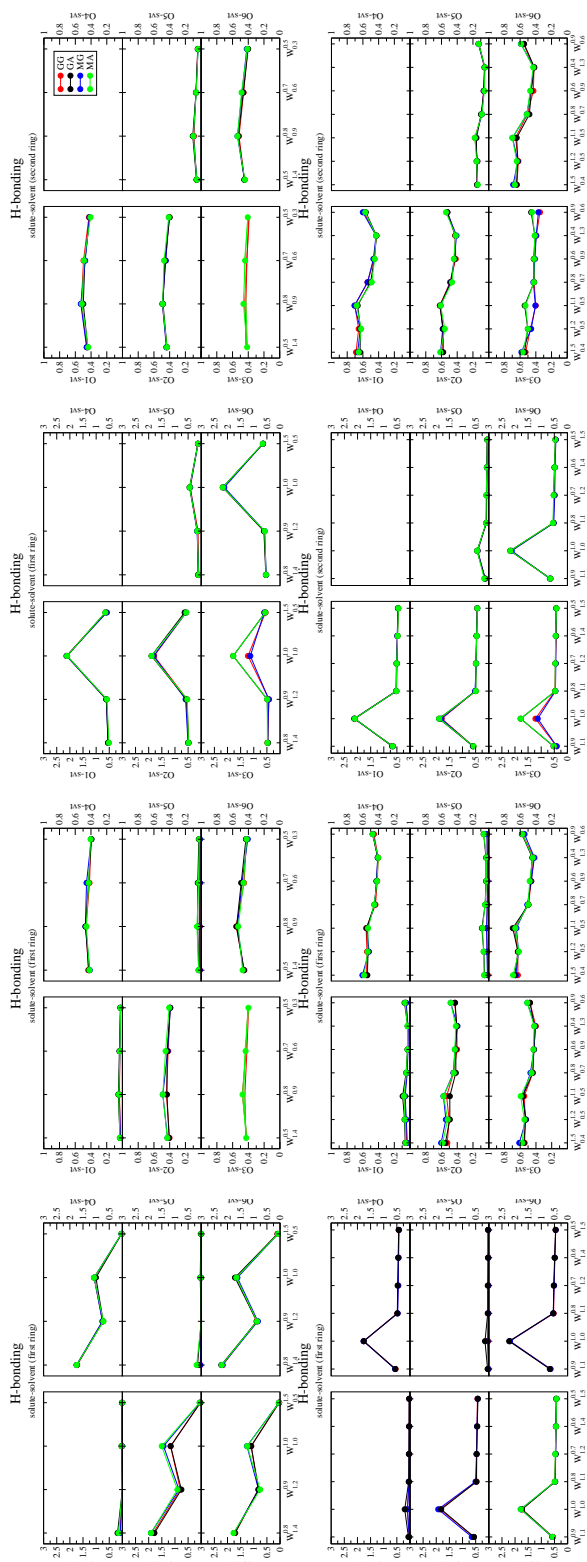
Chapter 5

Outlook

Molecular dynamics simulation is becoming an ever more important tool in the study of molecular systems. With the steady increase in the power of computers, the system dimension and time-scales that are accessible for computer simulations have reached a regime relevant for experimentally studied (bio)chemical processes. The reproduction of physically realistic conditions is not the only aspect of simulations. Artificial and unphysical situations are important alternatives.

In Chapter 2 artificial solvent models derived from the simple point charges water model by systematic modification of the oxygen-hydrogen bond length and of the atomic partial charges were developed. The consequences of the variation of these two parameters on the collective properties of the liquid can give insights into the particular nature of this fundamental solvent. In this chapter the focus was on the basic properties of the liquid and particular attention was given to the dielectric permittivity and hydrogen-bonding capacity, as these properties are strongly connected to the polarity of the solvent. Additional properties might be calculated to obtain a more complete characterization, *e.g.* viscosity, surface tension, heat capacity, thermal expansion coefficient or isothermal compressibility. It would also be interesting to correlate these properties with the dipole moment of the models, as well as with their higher order multipole moments. Moreover, the same approach might be used to systematically vary other parameters related to the geometry of the water molecule (*e.g.* the bond angle).

From an application point of view, these models can help in the study of the solvent

effects on solute molecule, as done in Chapter 3 and Chapter 4. They allow to disentangle the effects of the dielectric permittivity and of the H-bonding capacity of the solvent, while keeping its molecular size and shape constant, as well as water-like dispersive interactions. All kind of solute molecules experiencing conformational changes in non-aqueous environments might be simulated in these artificial solvents, and their behaviour as function of the solvent polarity analyzed.

Carbohydrates are a particularly fascinating type of molecules, and their variety and flexibility make them an endless source of study material. In Chapter 3 and Chapter 4, the importance of intramolecular hydrogen-bonding for these molecules was investigated in physical solvents of different polarity, as well as in the artificial solvents developed in Chapter 2. Hydrogen-bonding is often regarded as a fundamental driving force for the conformational properties of (bio)polymers. In our study, in aqueous environment, it was found to be an opportunistic consequence of the close proximity of two H-bonding groups in a given conformation, and not a factor contributing to the stability of this conformation. This is true for both monosaccharides and disaccharides. On the other hand, with a decrease of the solvent polarity, solvent-exposed hydrogen-bonding progressively evolves from a negligible to a very significant conformational steering force. However, introducing additional degrees of freedom that increase the complexity of the molecules (from monomers to dimers), the importance of intramolecular hydrogen-bonding in low polarity solvents takes a different form. Possible competition between hydrogen-bonds that mutually exclude each others can tune their influence, making them an adverse factor towards specific conformational preferences.

A generalization of this behaviour, especially in case of disaccharides, would be interesting. A more systematic study on different glycosidic linkages might help to understand what is so determinant for the conformational properties of carbohydrates in the aqueous environment. As hydrogen-bonding is generally considered a significant stabilizing factor

for the secondary structure of long carbohydrate chains, it would also be interesting to consider longer polymers. Alternative toy-systems might be used for a better understanding of hydrogen-bond competition. Furthermore, in view of these considerations about carbohydrates, a general rethinking about the role of hydrogen-bonding for (bio)polymers in aqueous environment would be necessary.

Bibliography

[1] N. Bohr, *On the constitution of atoms and molecule*, Philos. Mag. 26 (1913), pp. 1-24.

[2] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, *Equation of state calculations by fast computing machines*, J. Chem. Phys. 21 (1953), pp. 1087-1092.

[3] M.P. Allen and D.J. Tildesley, *Computer simulation of liquids*, Oxford University Press, New York, USA (1987).

[4] W.F. van Gunsteren and H.J.C. Berendsen, *Computer simulation of molecular dynamics: Methodology, applications and perspectives in chemistry*, Angew. Chem. Int. Ed. 29 (1990), pp. 992-1023.

[5] W.F. van Gunsteren, D. Bakowies, R. Baron, I Chandrasekhar, M. Christen, X. Daura, P. Gee, D.P. Geerke, A. Glättli, P.H. Hünenberger, M.A. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N.F.A. van der Vegt and H.B. Yu, *Biomolecular modelling: goals, problems, perspectives*, Angew. Chem. Int. Ed. 45 (2006), pp. 4064-4092.

[6] H.J.C. Berendsen, *Simulating the physical world*, Cambridge University Press, Cambridge, UK (2007).

[7] C.Y. Wang and X. Zhang, *Multiscale modeling and related hybrid approaches*, Curr. Opin. Solid State Mat. Sci. 10 (2006), pp. 2-14.

[8] M. Praprotnik, L. Delle Site and K. Kremer, *Multiscale simulation of soft matter: From scale bridging to adaptive resolution*, Ann. Rev. Phys. Chem. 59 (2008), pp. 545-571.

[9] P.H. Hünenberger and W.F. van Gunsteren Empirical classical interaction functions for molecular simulations In: *Computer simulation of biomolecular systems, theoretical and experimental applications Volume 3.*; W.F. van Gunsteren, P.K. Weiner and A.J. Wilkinson, Eds.; Kluwer/Escom Science Publishers, Dordrecht, The Netherlands (1997); pp 3-82..

[10] D. Frenkel and B. Smit, *Understanding molecular simulation Edition 2.*, Academic Press, San Diego, USA (2002).

[11] W.F. van Gunsteren, T. Huber and A.E. Torda, *Biomolecular modelling: overview of types of methods to search and sample conformational space*, AIP Conf. Proc. 330 (1995), pp. 253-268.

[12] A.F. Voter, F. Monatlenti and T.C. Germann, *Extending the time scale in atomistic simulations of materials*, Annu. Rev. Mater. Res. 32 (2002), pp. 321-46.

[13] M. Christen and W.F. van Gunsteren, *On searching in, sampling of, and dynamically moving through conformational space of biomolecular systems: A review*, J. Comput. Chem. 29 (2008), pp. 157-166.

[14] H.S. Hansen and P.H. Hünenberger, *Using the local elevation method to construct optimized umbrella sampling potentials: Calculation of the relative free energies and interconversion barriers of glucopyranose ring conformers in water*, J. Comput. Chem. 31 (2010), pp. 1-23.

[15] H.S. Hansen and P.H. Hünenberger, *Ball-and-stick local elevation umbrella sampling: molecular simulations involving enhanced sampling within conformational or alchemical subspaces of low internal dimensionalities, minimal irrelevant volume and problem-adapted geometries*, J. Chem. Theory Comput. 6 (2010), pp. 2622-2646.

[16] P.H. Hünenberger, *Thermostat algorithms for molecular dynamics simulations*, Adv. Polym. Sci. 173 (2005), pp. 105-149.

[17] W.F. van Gunsteren and M. Karplus, *Effect of constraints, solvent and crystal environment on protein dynamics*, Nature 293 (1981), pp. 677-678.

[18] J. Warwicker and H.C. Watson, *Calculation of the electric potential in the active site cleft due to α-helix dipoles*, J. Mol. Biol. 157 (1982), pp. 671-679.

[19] M.K. Gilson and B. Honig, *Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis*, Proteins: Struct. Funct. Genet. 4 (1988), pp. 7-18.

[20] S.C. Harvey, *Treatment of electrostatic effects in macromolecular modeling*, Proteins: Struct. Funct. Genet. 5 (1989), pp. 78-92.

[21] M.E. Davis and J.A. McCammon, *Electrostatics in biomolecular structure and dynamics*, Chem. Rev. 90 (1990), pp. 509-521.

[22] J.D. Madura, M.E. Davis, M.K. Gilson, R.C. Wade, B.A. Luty and J.A. McCammon Biological applications of electrostatic calculations and Brownian dynamics simulations In: *Reviews in computational chemistry Volume 4.*; K.B. Lipkowitz and D.B. Boyd, Eds.; VCH Publishers Inc., New York, USA (1994); pp 229-267.

[23] B. Honig and A. Nicholls, *Classical electrostatics in biology and chemistry*, Science 268 (1995), pp. 1144-1149.

[24] A. Warshel and A. Papazyan, *Electrostatic effects in macromolecules: Fundamental concepts and practical modeling*, Curr. Opin. Struct. Biol. 8 (1998), pp. 211-217.

[25] B. Roux and T. Simonson, *Implicit solvent models*, Biophys. Chem. 78 (1999), pp. 1-20.

[26] N.A. Baker, *Improving implicit solvent simulations: a Poisson-centric view*, Curr. Opin. Struct. Biol. 15 (2005), pp. 137-143.

[27] J.H. Chen, C.L. Brooks and J. Khandogin, *Recent advances in implicit solvent-based methods for biomolecular simulations*, Curr. Opin. Struct. Biol. 18 (2008), pp. 140-148.

[28] M. Levitt and A. Warshel, *Computer simulation of protein folding*, Nature 253 (1975), pp. 694-698.

[29] B. Smit, P.A.J. Hilbers, K. Esselink, L.A.M. Rupert, N.M. van Os and A.G. Schlijper, *Computer simulations of a water/oil interface in the presence of micelles*, Nature 348 (1990), pp. 624-625.

[30] M.L. Huertas, S. Navarro, M.C.L. Martinez and J.G. de la Torre, *Simulation of the conformation and dynamics of a double-helical model for DNA*, Biophys. J. 73 (1997), pp. 3142-3153.

[31] R.L.C. Akkermans and W.J. Briels, *Coarse-grained dynamics of one chain in a polymer melt*, J. Chem. Phys. 113 (2000), pp. 6409-6422.

[32] J. Baschnagel, K. Binder, P. Doruker, A.A. Gusev, O. Hahn, K. Kremer, W.L. Mattice, F. Müller-Plathe, M. Murat, W. Paul, S. Santos, U.W. Suter and W. Tries, *Bridging the gap between atomistic and coarse-grained models of polymers: status and perspectives*, Adv. Polym. Sci. 152 (2000), pp. 41-156.

[33] K. Drukker, G. Wu and G.C. Schatz, *Model simulations of DNA denaturation dynamic*, J. Chem. Phys. 114 (2001), pp. 579-590.

[34] F. Müller-Plathe, *Coarse-graining in polymer simulations: From the atomistic to the mesoscopic scale and back*, Chem. Phys. Chem. 3 (2002), pp. 754-769.

[35] S.J. Marrink, A.H. de Vries and A.E. Mark, *Coarse grained model for semiquantitative lipid simulations*, J. Phys. Chem B 108 (2004), pp. 750-760.

[36] V. Molinero and W.A. Goddard III, *M3B: A coarse-grained force field for molecular simulations of malto-oligosaccharides and their water mixtures*, J. Phys. Chem. B 108 (2004), pp. 1414-1427.

[37] S.O. Nielsen, C.F. Lopez, G. Srinivas and M.L. Klein, *Coarse grain models and the computer simulation of soft materials*, J. Phys.: Condens. Matter 16 (2004), pp. R481-R512.

[38] A. van den Noort, W.K. den Otter and W.J. Briels, *Coarse graining of slow variables in dynamic simulations of soft matter*, Europhys. Lett. 80 (2007), pp. 28003/1-28003/5.

[39] G.A. Voth, *Coarse-graining of condensed phase and biomolecular systems*, Taylor & Francis, Inc., New York, USA (2008).

[40] A. Warshel and M. Levitt, *Theoretical studies of enzymic reactions - dielectric, electrostatic and steric stabilization of carbonium-ion in reaction of lysozyme*, J. Mol. Biol. 103 (1976), pp. 227-249.

[41] W.L. Jorgensen and J. Tirado-Rives, *The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin*, J. Am. Chem. Soc. 110 (1988), pp. 1657-1666.

[42] W.L. Jorgensen, D.S. Maxwell and J. Tirado-Rives, *Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids*, J. Am. Chem. Soc. 118 (1996), pp. 11225-11236.

[43] P.K. Weiner and P.A. Kollman, *AMBER - Assisted model building with energy refinement - a general program for modeling molecules and their interactions*, J. Comput. Chem. 2 (1981), pp. 287-303.

[44] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell and P.A. Kollman, *A second generation force field for the simulation of proteins, nucleic acids and organic molecules*, J. Am. Chem. Soc. 117 (1995), pp. 5179-5197.

[45] J.W. Ponder and D.A. Case, *Force fields for protein simulations*, Adv. Prot. Chem. 66 (2003), pp. 27-85.

[46] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan and M. Karplus, *CHARMM: A program for macromolecular energy, minimization and dynamics calculations*, J. Comput. Chem. 4 (1983), pp. 187-217.

[47] A.D. MacKerell Jr., D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *All-atom empirical potential for molecular modeling and dynamics studies of proteins*, J. Phys. Chem. B 102 (1998), pp. 3586-3616.

[48] A.D. MacKerell, N. Banavali and N. Foloppe, *Development and current status of the charmm force field for nucleic acids*, Biopolymers 56 (2000), pp. 257-265.

[49] W.F. van Gunsteren and M. Karplus, *Effect of constraints on the dynamics of macromolecules*, Macromolecules 15 (1982), pp. 1528-1544.

[50] N. Schmid, A.P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A.E. Mark and W.F. van Gunsteren, *Definition and testing of the GROMOS force-field versions 54A7 and 54B7*, Eur. Biophys. J. 40 (2011), pp. 843-856.

[51] A.C.T. van Duin, S. Dasgupta, F. Loarant and W.A.III Goddard, *ReaxFF: A reactive force field for hydrocarbons*, J. Phys.Chem. A 105 (2001), pp. 9396-9409.

[52] D.W.M. Hofmann, L. Kuleshova and B. D'Aguanno, *A new reactive potential for the molecular dynamics simulation of liquid wate*, Chem. Phys. Lett. 448 (2007), pp. 138-143.

[53] J.-P. Ryckaert, G. Ciccotti and H.J.C. Berendsen, *Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes*, J. Comput. Phys. 23 (1977), pp. 327-341.

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

[55] M. Fixman, *Classical statistical mechanics of constraints: A theorem and application to polymers*, Proc. Natl. Acad. Sci. USA 71 (1974), pp. 3050-3053.

[56] E. Helfand, *Flexible vs. rigid constraints in statistical mechanics*, J. Chem. Phys. 71 (1979), pp. 5000-5007.

[57] B.M. Axilrod and E. Teller, *Interaction of the van der Waals type between three atoms*, J. Chem. Phys. 11 (1943), pp. 299-300.

[58] S.M. Foiles, M.I. Baskes and M.S. Daw, *Embedded-atoms-method functions for the fcc metals Cu, Ag, Au, Ni, Pd, Pt and their alloys*, Phys. Rev. B 33 (1986), pp. 7983-7991.

[59] F. Cleri and V. Rosato, *Tight-binding potentials for transition metals and alloys*, Phys. Rev. B 48 (1986), pp. 22-33.

[60] J.A. Barker and R.O. Watts, *Monte Carlo studies of the dielectric properties of water-like models*, Mol. Phys. 26 (1973), pp. 789-792.

[61] F.J Vesely, *N-particle dynamics of polarizable Stockmayer-type molecules*, J. Comp. Phys. 24 (1977), pp. 361-371.

[62] F.H. Stillinger and C.W. David, *Polarization model for water and its ionic dissociation products*, J. Chem. Phys. 69 (1978), pp. 1473-1484.

[63] P. Barnes, J.L. Finney, J.D. Nicholas and J.E. Quinn, *Cooperative effects in simulated water*, Nature 282 (1979), pp. 459-464.

[64] A. Warshel, *Calculations of chemical processes in solutions*, J. Phys. Chem. 83 (1979), pp. 1640-1652.

[65] I.G. Tironi, R. Sperb, P.E. Smith and W.F. van Gunsteren, *A generalized reaction field method for molecular dynamics simulations*, J. Chem. Phys. 102 (1995), pp. 5451-5459.

[66] P.P. Ewald, *Die Berechnung optischer und elektrostatischer Gitterpotentiale*, Ann. Phys. 369 (1921), pp. 253-287.

[67] T. Darden, L. Perera, L. Li and L. Pedersen, *New tricks for modelers from the crystallography toolkit: The particle mesh Ewald algorithm and its use in nucleic acid simulations*, Structure 7 (1999), pp. R55-R60.

[68] J.W. Eastwood, R.W. Hockney and D.N. Lawrence, *P3MDP - The three-dimensional periodic particle-particle/particle-mesh program*, Comput. Phys. Commun. 19 (1980), pp. 215-261.

[69] H.J.C. Berendsen, W.F. van Gunsteren, H.R.J Zwinderman and R.G. Geurtsen, *Simulations of proteins in water*, Ann. New York Acad. Sci. 482 (1986), pp. 269-285.

[70] M. Rigby, E.B. Smith, W.A. Wakeham and G.C. Maitland, *The forces between molecules*, Clarendon Press, Oxford, UK (1986).

[71] J.E. Jones, *On the determination of molecular fields. - I. From the variation of the viscosity of a gas with temperature*, Proc. R. Soc. London Ser. A 106 (1924), pp. 441-462.

[72] J.E. Jones, *On the determination of molecular fields. - II. From the equation of state of a gas*, Proc. R. Soc. London Ser. A 106 (1924), pp. 463-477.

[73] J.E. Lennard-Jones, *The equation of state of gases and critical phenomena*, Physica 4 (1937), pp. 941-956.

[74] A. Warshel and S. Lifson, *Consistent force field calculations. II. Crystal ttructures, sublimation energies, molecular and lattice vibrations, molecular conformations, and enthalpies of alkane*, J. Chem. Phys. 53 (1970), pp. 582-594.

[75] J.C. Slater, *The normal state of helium*, Phys. Rev. 32 (1928), pp. 349-360.

[76] R.A. Buckingham, *The classical equation of state of gaseous helium, neon and argon*, Proc. Roy. Soc. Lond. A 168 (1938), pp. 264-283.

[77] P.M. Morse, *Diatomic molecules according to the wave mechanics. II. Vibrational levels*, Phys. Rev. 34 (1929), pp. 57-64.

[78] M. C. Lawrence and G. N. Robertson, *Estimating the proton potential in kdp from infrared and crystallographic data*, Ferroelectrics (1981), pp. 34:1.

[79] P. Minary, G.J. Martyna and M.E. Tuckerman, *Algorithms and novel applications based on the isokinetic ensemble. II. Ab initio molecular dynamics*, J. Chem. Phys. 118 (2003), pp. 2527-2538.

[80] S.I. Sandler and J.K. Wheatley, *Intermolecular potential parameter combining rules for Lennard-Lones 6-12 potential*, Chem. Phys. Lett. 10 (1971), pp. 375-378.

[81] W.F. van Gunsteren, *The GROMOS software for biomolecular simulation. Available at:* http://www.gromos.net (05/05/2011).

[82] A.P. Eichenberger, J.R. Allison, J. Dolenc, D.P. Geerke, B.A.C. Horta, K. Meier, C. Oostenbrink, N. Schmid, D. Steiner, D. Wang and W.F. van Gunsteren, *The GROMOS++ software for the analysis of biomolecular simulation trajectories*, J. Chem. Theory Comput. 7 (2011), pp. 3379-3390.

[83] A.-P. E. Kunz, J.R. Allison, D.P. Geerke, B.A.C. Horta, P.H. Hünenberger, S. Riniker, N. Schmid and W.F. van Gunsteren, *New functionalities in the GROMOS biomolecular simulation software*, J. Comput. Chem. 33 (2012), pp. 340-353.

[84] N. Schmid, C.D. Christ, M. Christen, A.P. Eichenberger and W.F. van Gunsteren, *Architecture, implementation and parallelisation of the GROMOS software for biomolecular simulation*, Comput. Phys. Commun. 183 (2012), pp. 890-903.

[85] W.F. van Gunsteren, R. Kaptein and E.R.P. Zuiderweg Use of molecular dynamics computer simulations when determining protein structure by 2D NMR In: *Proceedings NATO/CECAM workshop on nucleic acid conformation and dynamics*; W.K. Olson, Ed.; Orsay, CECAM, France. (1984); pp 79-92.

[86] A.E. Torda, R.M. Scheek and W.F. van Gunsteren, *Time-dependent distance restraints in molecular dynamics simulations*, Chem. Phys. Lett. 157 (1989), pp. 289-294.

[87] A.E. Torda, R.M. Brunne, T. Huber, H. Kessler and W.F. van Gunsteren, *Structure refinement using time-averaged J-coupling constant restraints*, J. Biomol. NMR 3 (1993), pp. 55-66.

[88] C.A. Schiffer, R. Huber, K. Wütrich and W.F. van Gunsteren, *Simultaneous refinement of the structure of BPTI against NMR data measured in solution and X-ray diffraction data measured in single crystals*, J. Mol. Biol. 241 (1994), pp. 588-599.

[89] C.A. Schiffer, P. Gros and W.F. van Gunsteren, *Time-averaging crystallographic refinement: Possibilities and limitations using α-cyclodextrin as a test system*, Acta Crystallogr. D 51 (1995), pp. 85-92.

[90] T. Huber, A.E. Torda and W.F. van Gunsteren, *Local elevation: A method for improving the searching properties of molecular dynamics simulation*, J. Comput.-Aided Mol. Des. 8 (1994), pp. 695-708.

[91] T.P. Straatsma and H.J.C. Berendsen, *Free energy of ionic hydration: Analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations*, J. Chem. Phys. 89 (1988), pp. 5876-5886.

[92] X. Kong and C.L. Brooks III, *λ-dynamics: A new approach to free energy calculations*, J. Chem. Phys. 105 (1996), pp. 2414-2423.

[93] T. Schlick, *New approaches to potential energy minimization and molecular dynamics algorithms*, Computers Chem. 15 (1991), pp. 251-260.

[94] S. Kirkpatrick, C.D. Gelatt Jr. and M.P. Vecchi, *Optimization by simulated annealing*, Science 220 (1983), pp. 671-680.

[95] V. Cerny, *Thermodynamical approach to the traveling salesman problem: An efficient simulation algorith*, JOTA 45 (1985), pp. 41-51.

[96] R.C. van Schaik, H.J.C. Berendsen, A.E. Torda and W.F. van Gunsteren, *A structure refinement method based on molecular dynamics in four spatial dimensions*, J. Mol. Biol. 234 (1993), pp. 751-762.

[97] A. Amadei, A.B.M. Linssen and H.J.C. Berendsen, *Essential dynamics of proteins*, Proteins: Struct. Funct. Genet. 17 (1993), pp. 412-425.

[98] R.H. Swendsen and J.-S. Wang, *Replica Monte Carlo simulation of spin-glasses*, Phys. Rev. Lett. 57 (1986), pp. 2607-2609.

[99] K. Hukushima and K. Nemoto, *Exchange Monte Carlo method and application to spin glass simulations*, J. Phys. Soc. Jpn. 65 (1996), pp. 1604-1608.

[100] S. Yun-yu, W. Lu and W.F. van Gunsteren, *On the approximation of solvent effects on the conformation and dynamics of cyclosporin A by stochastic dynamics simulation techniques*, Mol. Simul. 1 (1988), pp. 369-383.

[101] W.F. van Gunsteren, S.R. Billeter, A.A. Eising, P.H. Hünenberger, P. Krüger, A.E. Mark, W.R.P. Scott and I.G. Tironi, *Biomolecular simulation: The GROMOS96 manual and user guide*, Verlag der Fachvereine, Zürich, Switzerland (1996).

[102] L. Verlet, *Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules*, Phys. Rev. 159 (1967), pp. 98-103.

[103] R.W. Hockney, *The potential calculation and some applications*, Methods Comput. Phys. 9 (1970), pp. 135-211.

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

[105] G.M. Torrie and J.P. Valleau, *Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling*, J. Comput. Phys. 23 (1977), pp. 187-199.

[106] H. Tsujishita, I. Moriguchi and S. Hirono, *Potential-scaled molecular dynamics and potential annealing. Effective conformational search techniques for biomolecules*, J. Phys. Chem. 97 (1993), pp. 4416-4420.

[107] G. Jacucci and A. Rahman The possibility of using a larger timestep in MD studies of water In: *Report on Workshop Methods in Molecular Dynamics - Long Timescale events, Orsa*; ; C.E.C.A.M. (1974); pp 32-40.

[108] C.H. Bennett, *Mass tensor molecular dynamics*, J. Comput. Phys. 19 (1975), pp. 267-279.

[109] T. Huber and W.F. van Gunsteren, *SWARM-MD: Searching conformational space by cooperative molecular dynamics*, J. Phys. Chem. A 102 (1998), pp. 5937-5943.

[110] M.A. Kastenholz and P.H. Hünenberger, *Influence of artificial periodicity and ionic strength in molecular dynamics simulations of charged biomolecules employing lattice-sum methods*, J. Phys. Chem. B 108 (2004), pp. 774-788.

[111] M.M. Reif, V. Kräutler, M.A. Kastenholz, X. Daura and P.H. Hünenberger, *Explicit-solvent molecular dynamics simulations of a reversibly-folding β-heptapeptide in methanol: Influence of the treatment of long-range electrostatic interactions*, J. Phys. Chem. B 113 (2009), pp. 3112-3128.

[112] W.G. Hoover, A.J.C. Ladd and B. Moran, *High-strain-rate plastic flow studied via nonequilibrium molecular dynamics*, Phys. Rev. Lett. 48 (1982), pp. 1818-1820.

[113] D.J. Evans, *Computer "experiment" for nonlinear thermodynamics of Couette flow*, J. Chem. Phys. 78 (1983), pp. 3297-3302.

[114] L.V. Woodcock, *Isothermal molecular dynamics calculations for liquid salts*, Chem. Phys. Lett. 10 (1971), pp. 257-261.

[115] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. di Nola and J.R. Haak, *Molecular dynamics with coupling to an external bath*, J. Chem. Phys. 81 (1984), pp. 3684-3690.

[116] S. Nosé, *A molecular dynamics method for simulations in the canonical ensemble*, Mol. Phys. 52 (1984), pp. 255-268.

[117] S. Nosé, *A unified formulation of the constant temperature molecular dynamics methods*, J. Chem. Phys. 81 (1984), pp. 511-519.

[118] W.G. Hoover, *Canonical dynamics: Equilibrium phase-space distributions*, Phys. Rev. A 31 (1985), pp. 1695-1697.

[119] B. Mao, G.M. Maggiora and K.C. Chou, *Mass-weighted molecular dynamics simulation of cyclic polypeptides*, Biopolymers 31 (1991), pp. 1077-1086.

[120] D.J. Evans and G.P. Morriss, *Isothermal-isobaric molecular dynamics*, Chem. Phys. 77 (1983), pp. 63-66.

[121] D.J. Evans and G.P. Morriss, *The isothermal/isobaric molecular dynamics ensemble*, Phys. Lett. 98A (1983), pp. 433-436.

[122] F.F. Abraham, *Computer simulations of surfaces, interfaces, and physisorbed films*, J. Vac. Sci. Technol. B 2 (1983), pp. 534-549.

[123] W.G. Hoover, *Constant-pressure equations of motion*, Phys. Rev. A 34 (1986), pp. 2499-2500.

[124] G.J. Martyna, D.J. Tobias and M.L. Klein, *Constant pressure molecular dynamics algorithms*, J. Chem. Phys. 101 (1994), pp. 4177-4189.

[125] A. Imberty and S. Pérez, *A revisit to the three-dimensional structure of B-type starch*, Biopolymers 25 (1988), pp. 1205-1221.

[126] L. Pauling and R.B. Corey, *Atomic coordinates and structure factors for two helical configurations of polypeptide chains*, Proc. Natl. Acad. Sci. USA 37 (1951), pp. 235-240.

[127] J.D. Watson and F.H.C. Crick, *Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid*, Nature 171 (1953), pp. 737-738.

[128] W. Zhang, H. Zhao, I. Carmichael and A. Serianni, *An NMR investigation of putative interresidue H-bonding in methyl α-cellobioside in solution*, Carbohydr. Res. 344 (2009), pp. 1582-1587.

[129] B. Guillot, *A reappraisal of what we have learnt during three decades of computer simulations on water*, J. Mol. Liq. 101 (2002), pp. 219-260.

[130] M.W. Mahoney and W.L. Jorgensen, *A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions*, J. Chem. Phys. 112 (2000), pp. 8910-8922.

[131] I. Gladich, P. Shepson, I. Szleifer and M. Carignano, *Halide and sodium ion parameters for modeling aqueous solutions in TIP5P-Ew water*, Chem. Phys. Lett. 489 (2010), pp. 113-117.

[132] D. Wang, M.L. Ámundadóttir, W.F. van Gunsteren and P.H. Hünenberger, *Intramolecular hydrogen-bonding in aqueous carbohydrates as a cause or consequence of conformational preferences: A molecular dynamics study of cellobiose stereoisomers*, Eur. Biophys. J. 42 (2013), pp. 521-537.

[133] J.P.M. Postma, *MD of H2O, a molecular dynamics study of water*, Thesis, University of Groningen (1985).

[134] E.C. Fuchs, *Can a century old experiment reveal hidden properties of water*, Water 2 (2010), pp. 381-410.

[135] P. Ball, *Water: water - an enduring mystery*, Nature 452 (2008), pp. 291-292.

[136] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren and J. Hermans Interaction models for water in relation to protein hydration In: *Intermolecular Forces*; B. Pullman, Ed.; Reidel, Dordrecht, The Netherlands (1981); pp 331-342.

[137] W.L. Jorgensen and J.D. Madura, *Solvation and conformation of methanol in water*, J. Am. Chem. Soc. 105 (1983), pp. 1407-1413.

[138] A. Glättli, X. Daura and W.F. van Gunsteren, *Derivation of an improved simple point charge model for liquid water: SPC/A and SPC/L*, J. Chem. Phys. 116 (2002), pp. 9811-9828.

[139] A. Glättli, X. Daura, D. Seebach and W.F. van Gunsteren, *Can one derive the conformational preference of a β-peptide from its CD spectrum*, J. Am. Chem. Soc. 124 (2002), pp. 12972-12978.

[140] H. Nada and J.P.J.M. van der Eerden, *An intermolecular potential model for the simulation of ice water and water near the melting point: a six-site model of H_2*, J. Chem. Phys. 118 (2003), pp. 7401-7413.

[141] Fischer M-L., Chandra J.T., Brooks A., Ichiye B.R., , *A temperature of maximum density in soft sticky dipole water*, Chem. Phys. Lett. 376 (2003), pp. 646-652.

[142] L.-P. Wang, T. Head-Gordon, J.W. Ponder, P. Ren, J.D. Chodera, P.K. Eastman and Pande V.S. Martinez T.J., *Systematic improvement of a classical molecular model of wate*, J. Phys. Chem. B 117 (2003), pp. 9956-9972.

[143] S. Izadi, R. Anandakrishnan and A.V. Onufriev, *Building water models: A different approach*, J. Phys. Chem. Lett. 5 (2014), pp. 3863â3871.

[144] H. Yu and W.F. van Gunsteren, *Charge-on-spring polarizable water models revisited: From water clusters to liquid water to ice*, J. Chem. Phys. 121 (2004), pp. 9549-9564.

[145] A.-P. E. Kunz and W.F. van Gunsteren, *Development of a non-linear classical polarization model for liquid water and aqueous solutions: COS/D*, J. Phys. Chem. A 113 (2009), pp. 11570-11579.

[146] S.J. Bachmann and W.F. van Gunsteren, *An improved simple polarisable water model for use in biomolecular simulation*, J. Chem. Phys. 141 (2014), pp. 22D515/1-22D515/15.

[147] G.S. Kell, *Precise representation of volume properties of water at one atmosphere*, J. Chem. Eng. Data 12 (1967), pp. 66-69.

[148] P.E. Smith and W.F. van Gunsteren, *Consistent dielectric properties of the simple point charge and extended simple point charge water models at 277 and 300 K*, J. Chem. Phys. 100 (1994), pp. 3169-3174.

[149] D.P. Geerke, C. Oostenbrink, N.F.A. van der Vegt and W.F. van Gunsteren, *An effective force field for molecular dynamics simulations of dimethyl sulfoxide and dimethyl sulfoxide-water mixtures*, J. Phys. Chem. B 108 (2004), pp. 1436-1445.

[150] M. Neumann, *Dipole moment fluctuation formulas in computer simulations of polar systems*, Mol. Phys. 50 (1983), pp. 841-858.

[151] B. Halle and H. Wennerström, *Interpretation of magnetic resonance data from water nuclei in heterogenous systems*, J. Chem. Phys. 75 (1981), pp. 1928.

[152] M. Neumann, *The dielectric constant of water. Computer simulations with the MCY potential*, J. Chem. Phys. 82 (1985), pp. 5663-5672.

[153] S. Riniker, C.D. Christ, H.S. Hansen, P.H. Hünenberger, C. Oostenbrink, D. Steiner and W.F. van Gunsteren, *Calculation of relative free energies for ligand-protein binding, solvation and conformational transitions using the GROMOS biomolecular simulation software*, J. Phys. Chem. B 115 (2011), pp. 13570-13577.

[154] S. L. Carnie and G. N. Patey, *Fluids of polarizable hard spheres with dipoles and tetrahedral quadrupoles Integral equation results with application to liquid wate*, Mol. Phys. 47 (1982), pp. 1129-1151.

[155] P. G. Kusalik and G. N. Patey, *On the molecular theory of aqueous electrolyte solutions. I. The solution of the RHNC approximation for models at finite concentratio*, J. Chem. Phys. 88 (1988), pp. 7715.

[156] T Ichiye and M. L. Tan, *Soft sticky dipole-quadrupole-octupole potential energy function for liquid water: an approximate expansion*, J. Phys. Chem. 124 (2006), pp. 134504.

[157] J.L.F. Abascal, R.G. Fernández, L.G. MacDowell, E. Sanz and C. Vega, *Ice: A fruitful source of information about liquid water*, J. Mol. Liq. 136 (2007), pp. 214-220.

[158] J.L.F. Abascal and C. Vega, *Dipole-quadrupole force ratios determine the ability of potential models to describe the phase diagram of water*, Phys. Rev. Lett. 98 (2007), pp. 237801/1-237801/4.

[159] J.L.F. Abascal and C. Vega, *The water forcefield: Importance of dipolar and quadrupolar interactions*, J. Phys. Chem. C 111 (2007), pp. 15811-15822.

[160] J.L.F. Abascal and C. Vega, *The melting point of hexagonal ice (Ih) is strongly dependent on the quadrupole of the water models*, Phys. Chem. Chem. Phys. 9 (2007), pp. 2775-2778.

[161] J.A. Te and T. Ichiye, *Understanding structural effects of multipole moments on aqueous solvation of ions using the soft-sticky dipole-quadrupole-octupole water model*, Chem. Phys. Lett. 499 (2010), pp. 219-225.

[162] S. Niu, M.-L. Tan and T. Ichiye, *The large quadrupole of water molecules*, J. Chem. Phys. 134 (2011), pp. 134501/1-134501/11.

[163] C. Kramer, A. Spinn and K. R. Liedl, *Charge anisotropy: where the atomic multipoles matter mos*, J. Chem. Theory Comput. 10 (2014), pp. 4488-4496.

[164] J. A. Hayward and J. R. Reimers, *Unit cells for the simulation of hexagonal ic*, J. Chem. Phys. 106 (1997), pp. 1518.

[165] A. L. Shostak, W. L. Ebenstein and J. S. Muenter, *The dipole moment of water. I Dipole moments and hyperfine properties of H2O and HDO in the ground and excited vibrational states*, J. Chem. Phys. 94 (1991), pp. 5875.

[166] D.R. Lide, *CRC Handbook of Chemistry and Physics Edition 88 (Internet version).*, CRC Press, Boca Raton, Florida (2008).

[167] K. Krynicki, C. D. Green and D. W. Sawyer, *Pressure and temperature dependence of self-diffusion in wate*, Faraday Discuss. Chem. Soc. 66 (1978), pp. 199.

[168] R. Ludwig, *NMR relaxation studies in water-alcohol mixtures: the water-rich regio*, Chem. Phys. 195 (1995), pp. 329.

[169] K.J. Naidoo and J.Y. Chen, *The role of water in the design of glycosidic linkage flexibility*, Mol. Phys. 101 (2003), pp. 2687-2694.

[170] V. Kräutler, M. Müller and P.H. Hünenberger, *Conformation, dynamics, solvation and relative stabilities of selected β-hexopyranoses in water: a molecular dynamics study with the GROMOS 45A4 force field*, Carbohydr. Res. 342 (2007), pp. 2097-2124.

[171] L. Perić-Hassler, H.S. Hansen, R. Baron and P.H. Hünenberger, *Conformational properties of glucose-based disaccharides using molecular dynamics simulations with local elevation umbrella sampling*, Carbohydr. Res. 345 (2010), pp. 1781-1801.

[172] J. Reuben, *Effects of solvent and hydroxyl deuteration on the carbon-13 NMR spectrum of D-idose: Spectral assignments, tautomeric compositions, and conformational equilibria*, J. Am. Chem. Soc. 107 (1985), pp. 5867-5870.

[173] P. Dais and A.S. Perlin, *Intramolecular hydrogen-bonding and solvation contributions to the relative stability of the β-furanose form of D-fructose in dimethyl sulfoxide*, Carbohydr. Res. 169 (1987), pp. 159-169.

[174] B. Bernet and A. Vasella, *^{1}H-NMR analysis of intra- and intermolecular H-bonds of alcohols in DMSO: Chemical shift of hydroxy groups and aspects of conformational analysis of selected monosaccharides, inositols and ginkgolides*, Helv. Chim. Acta 83 (2000), pp. 995-1021.

[175] A. Roën, J.I. Padrón and J.T. Vásquez, *Hydroxymethyl rotamer populations in disaccharides*, J. Org. Chem. 68 (2002), pp. 4615-4630.

[176] V.N. Gritsan, V.P. Panov and V.G. Kachur, *Theoretical analysis of hydrogen bonds in carbohydrate crystals*, Carbohydr. Res. 112 (1983), pp. 11-21.

[177] P. Çarçabal, R.A. Jockusch, I. Hünig, L.C. Snoek, R.T. Kroemer, B.G. Davis, D.P. Gamblin, I. Compagnin, J. Oomens and J.P. Simons, *Hydrogen bonding and cooperativity in isolated and hydrated sugars: Mannose, galactose, glucose and lactose*, J. Am. Chem. Soc. 127 (2005), pp. 11414-11425.

[178] T.A. Bryce, A.A. McKinnon, E.R. Morris, D.A. Rees and D. Thom, *Chain conformations in the sol-gel transitions for polysaccharide systems, and their characterisation by spectroscopic methods*, Faraday Discuss. 57 (1975), pp. 221-229.

[179] R.J. Abraham, E.J. Chambers and W.A. Thomas, *Conformational analysis. Part 19. Conformational analysis of 6-deoxy-6-fluoro-D-glucose (6DFG) in solution*, Magn. Reson. Chem. 30 (1992), pp. S60-S65.

[180] Y. Nishida, H. Ohrui and H. Meguro, *^{1}H-NMR studies of (6R)- and (6S)-deuterated hexoses: assignment of the preferred rotamers about C5-C6 bond of D-glucose and D-galactose derivatives in solution*, Tetrahedron Lett. 25 (1984), pp. 1575-1578.

[181] Y. Nishida, H. Hori, H. Ohrui and H. Meguro, *[1]-H NMR analyses of rotameric distribution of C5-C6 bonds of D-glucopyranose in solution*, J. Carbohydr. Chem. 7 (1988), pp. 239-250.

[182] R.J. Abraham, E.J. Chambers and W.A. Thomas, *Conformational analysis. Part 20. Conformational analysis of 4-deoxy-4-fluoro-D-glucose and 6-deoxy-6-fluoro-D-galactose in solution*, Magn. Reson. Chem. 32 (1994), pp. 248-254.

[183] M.-C. Brochier-Salon and C. Morin, *Conformational analysis of 6-deoxy-6-iodo-D-glucose in aqueous solution*, Magn. Reson. Chem. 38 (2000), pp. 1041-1042.

[184] C. Thibaudeau, R. Stenutz, B. Hertz, T. Klepach, S. Zhao, Q. Wu, I. Carmichael and A.S. Serianni, *Correlated C-C and C-O bond conformations in saccharide hydroxymethyl groups: Parametrization and application of redundant ^1H-^1H, ^{13}C-^1H, and $^{13}HC^{13}C$ NMR J-couplings*, J. Am. Chem. Soc. 126 (2004), pp. 15668-15685.

[185] T. Suzuki, H. Kawashima and T. Sota, *Conformational properties of and a reorientation triggered by sugar-water vibrational resonance in the hydroxymethyl group in hydrated β-glucopyranose*, J. Phys. Chem. B 110 (2006), pp. 2405-2418.

[186] K. Bock and H. Thøgersen, *Nuclear Magnetic Resonance Spectroscopy in the Study of Mono- and Oligosaccharides*, Ann. Rep. NMR Spect. 13 (1983), pp. 1-57.

[187] H. Ohrui, Y. Nishida, H. Higuchi, H. Hori and H. Meguro, *The preferred rotamer about the C_5-C_6 bond of D-galactopyranoses and the stereochemistry of dehydrogenation by D-galactose oxidase*, Can. J. Chem. 65 (1987), pp. 1145-1153.

[188] N.K. de Vries and H.M. Buck, *Different rotamer populations around the C5-C6 bond for α and β-galactopyranosides through the combined interaction of the gauche and anomeric effects: A 300-MHz ^1H-NMR and MNDO study*, Carbohydr. Res. 165 (1987), pp. 1-16.

[189] K. Bock and J.O. Duus, *A conformational study of hydroxymethyl groups in carbohydrates investigated by ^1H-NMR spectroscopy*, J. Carbohydr. Chem. 13 (1994), pp. 513-543.

[190] I. Tvaroška, F.R. Taravel, J.P. Utille and J.P. Carver, *Quantum mechanical and NMR spectroscopy studies on the conformations of the hydroxymethyl groups in aldohexosides*, Carbohydr. Res. 337 (2002), pp. 353-367.

[191] R. Stenutz, I. Carmichael, G. Widmalm and A.S. Serianni, *Hydroxymethyl group conformation in saccharides: Structural dependencies of the $^2J_{HH}$, $^3J_{HH}$, and $^1J_{CH}$ coupling constants*, J. Org. Chem. 67 (2002), pp. 949-958.

[192] S. Wolfe, *The gauche effect. Some stereochemical consequences of adjacent electron pairs and polar bonds*, Acc. Chem. Res. 5 (1970), pp. 102-111.

[193] S. Wolfe, L.M. Tel, W.J. Haines, M.A. Robb and I.G. Csizmadia, *The gauche effect. A study of localized molecular orbitals and excited-state geometries in FCH_2OH*, J. Am. Chem. Soc. 95 (1973), pp. 4863-4870.

[194] N.D. Epiotis, S. Sarkanen, D. Bjorkquist, L. Bjorkquist and R. Yates, *Open shell interactions, nonbonded attraction, and aromaticity. Implications for regiochemistry*, J. Am. Chem. Soc. 96 (1974), pp. 4075-4084.

[195] A. Abe and J.E. Mark, *Conformational energies and the random-coil dimensions and dipole moments of the polyoxides $CH_3[(CH_2)_yO]_xCH_3$*, J. Am. Chem. Soc. 98 (1976), pp. 6468-6476.

[196] O.A. Subbotin and N.M. Sergeyev, *Application of carbon-13 nuclear magnetic resonance spectrometry to the study of isomer and conformer ratios of dichlorocyclohexanes in their mixtures*, Anal. Chem. 48 (1976), pp. 545-546.

[197] N.S. Zefirov, L.G. Gurvich, A.S. Shashkov, M.Z. Krimer and E.A. Vorobeva, *Stereochemical studies - XX. Conformations of 1,2-trans-disubstituted cyclohexanes*, Tetrahedron 32 (1976), pp. 1211-1219.

[198] N.S. Zefirov, V.V. Samoshin, O.A. Subbotin and V.I. Baranekov, *The gauche effect. On the nature of the interaction between electronegative substituents in trans-1,2-disubstituted cyclohexanes*, Tetrahedron 34 (1978), pp. 2953-2959.

[199] S. Inagaki, K. Iwase and Y. Mori, *A new theory for the anti-periplanar effect*, Chem. Lett. 3 (1986), pp. 417-420.

[200] N.K. de Vries and H.M. Buck, *Solvent dependence of the rotamer population around the interglycosidic C(5)-C(6) bond of $(1 \rightarrow 6)$-β-linked digalactopyranosides*, Rec. Trav. Chim. Pays-Bas 106 (1987), pp. 453-460.

[201] A. Abe and K. Inomata, *Gas phase NMR of 1,2-dimethoxyethane*, J. Mol. Struct. 245 (1991), pp. 399-402.

[202] R.J. Abraham, E.J. Chambers and W.A. Thomas, *Conformational analysis of 6-deoxy-6-fluoro-D-glucopyranose, 6-deoxy-6-fluoro-D-galactopyranose, and 4-deoxy-4-fluoro-D-glucopyranose by ^1H-NMR. spectroscopy*, Carbohydr. Res. 226 (1992), pp. 1-5.

[203] E. Juaristi and S. Antúnez, *Conformational analysis of 5-substituted 1,3-dioxanes. 6. Study of the attractive gauche effect in O-C-C-O segments*, Tetrahedron 48 (1992), pp. 5941-5950.

[204] G.D. Rockwell and T.B. Grindley, *Effect of solvation on the rotation of hydroxymethyl groups in carbohydrates*, J. Am. Chem. Soc. 120 (1998), pp. 100953-100963.

[205] V. Babin, C. Roland and C. Sagui, *Adaptively biased molecular dynamics for free energy calculations*, J. Chem. Phys. 128 (2008), pp. 134101/1-134101/7.

[206] P.F.J. Fuchs, H.S. Hansen, P.H. Hünenberger and B.A.C. Horta, *A GROMOS parameter set for vicinal diether functions: properties polyethyleneoxide and polyethyleneglycol*, J. Chem. Theory Comput. 8 (2012), pp. 3943-3963.

[207] R.U. Lemieux and J.T. Brewer, *Conformational preferences for solvated hydroxymethyl groups in hexopyranose structures*, Adv. Chem. 117 (1973), pp. 121-146.

[208] D.G. Streefkerk and A.M. Stephen, *P.M.R. studies on fully methylated aldohexopyranosides and their 6-deoxy analogues using lanthanide-shift reagents*, Carbohydr. Res. 49 (1976), pp. 13-25.

[209] C. Beeson, N. Pham, G. Shipps Jr. and T.A. Dix, *A comprehensive description of the free energy of an intramolecular hydrogen bond as a function of solvation: NMR study*, J. Am. Chem. Soc. 115 (1993), pp. 6803-6812.

[210] E.Q. Morales, J.I. Padrón, M. Trujillo and J.T. Vásquez, *CD and ^1H NMR study of the rotational population dependence of the hydroxymethyl group in β glucopyranosides on the aglycon and its absolute configuration*, J. Org. Chem. 60 (1995), pp. 2537-2548.

[211] J. Gonzalez-Outeirino, K.N. Kirschner, S. Thobhani and R.J. Woods, *Reconciling solvent effects on rotamer populations in carbohydrates - A joint MD and NMR analysis*, Can. J. Chem. 84 (2006), pp. 569-579.

[212] P. Charisiadis, V.G. Kontogianni, C.G. Tsiafoulis, A.G. Tzakos, M. Siskos and I.P. Gerothanassis, *^1H-NMR as a structural and analytical tool of intra- and intermolecular hydrogen bonds of phenol-containing natural products and model compounds*, Molecules 19 (2011), pp. 13643-13682.

[213] G.T. Giuffredi, V. Gouverneur and B. Bernet, *Intramolecular OH$\cdots$FC hydrogen bonding in fluorinated carbohydrates: CHF is a better hydrogen bond acceptor than CF_2*, Angew. Chem. Int. Ed. Engl. 52 (2013), pp. 10524â10528.

[214] C. Mayato, R.L. Dorta and J.T. Vázquez, *Experimental evidence on the hydroxymethyl group conformation in alkyl β-D-mannopyranosides*, Tetrahedron: Asymmetry 15 (2004), pp. 2385-2397.

[215] S.E. Barrows, F.J. Dulles, C.J. Cramer, A.D. French and D.G. Truhlar, *Relative stability of alternative chair forms and hydroxymethyl conformations of β-D-glucopyranose*, Carbohydr. Res. 276 (1995), pp. 219-251.

[216] S.B. Engelsen, C. Hervé du Penhoat and S. Pérez, *Molecular relaxation of sucrose in aqueous solution: How a nanosecond molecular dynamics simulation helps to reconcile NMR data*, J. Phys. Chem. 99 (1995), pp. 13334-13350.

[217] G.I. Csonka, K. Éliás and I.G. Csizmadia, *Relative stability of the $^1C_4$ and $^4C_1$ chair forms of β-D-glucose: a density functional study*, Chem. Phys. Lett. 257 (1996), pp. 49-60.

[218] K.A. Jebber, K. Zhang, C.J. Cassady and A. Chung-Phillips, *Ab initio and experimental studies on the protonation of glucose in the gas phase*, J. Am. Chem. Soc. 118 (1996), pp. 10515-10524.

[219] G.I. Csonka, I. Kolossváry, P. Császár, K. Éliás and I.G. Csizmadia, *The conformational space of selected aldo-pyranohexoses*, J. Mol. Struct. 395 (1997), pp. 29-40.

[220] W. Damm, A. Frontera, J. Tirado-Rives and W. Jorgensen, *OPLS all-atom force field for carbohydrates*, J. Comput. Chem. 18 (1997), pp. 1955-1970.

[221] B.D. Wladkowski, S.A. Chenoweth, K.E. Jones and J.W. Brown, *Exocyclic hydroxymethyl rotational conformers of β- and α-D-glucopyranose in the gas phase and aqueous solution*, J. Phys. Chem. A 102 (1998), pp. 5086-5092.

[222] J.-H. Lii, B. Ma and N.L. Allinger, *Importance of selecting proper basis set in quantum mechanical studies of potential energy surfaces of carbohydrates*, J. Comput. Chem. 20 (1999), pp. 1593-1603.

[223] F.O. Talbot and J.P. Simons, *Sugars in the gas phase: the spectroscopy and structure of jet-cooled phenyl β-D-glucopyranoside*, Phys. Chem. Chem. Phys. 4 (2002), pp. 3562-3565.

[224] R.A. Jockusch, F.O. Talbot and J.P. Simons, *Sugars in the gas phase. Part 2: The spectroscopy and structure of jet-cooled phenyl β-D-galactopyranoside*, Phys. Chem. Chem. Phys. 5 (2003), pp. 1502-1507.

[225] M. Appell, G. Strati, J.L. Willett, , *B3LYP/6-311++G** study of α- and β-D-glucopyranose and 1,5-anhydro-D-glucitol: $^4C_1$ and $^1C_4$ chairs, $^{3,O}B$ and $B_{3,O}$ boats, and skew-boat conformations*, Carbohydr. Res. 339 (2004), pp. 537-551.

[226] J.C. Corchado, M.L. Sánchez and M.A. Aguilar, *Theoretical study of the relative stability of rotational conformers of α and β-D-glucopyranose in gas phase and aqueous solution*, J. Am. Chem. Soc. 126 (2004), pp. 7311-7319.

[227] J.P. Simons, R.A. Jockusch, P. Çarçabal, I. Hünig, R.T. Kroemer, N.A. McLeod and L.C. Snoek, *Sugars in gas phase. Spectroscopy, conformation, hydration, co-operativity and selectivity*, Int. Rev. Phys. Chem. 24 (2005), pp. 489-531.

[228] N.A. Macleod, C. Johannessen, L. Hecht, L.D. Barron and J.P. Simons, *From the gas phase to aqueous solution: Vibrational spectroscopy, Raman optical activity and conformational structure of carbohydrates*, Int. J. Mass Spectrom. 253 (2006), pp. 193-200.

[229] F.A. Momany, M. Appell, J.L. Willett, U. Schnupf and W.B. Bosma, *DFT study of α- and β-D-galactopyranose at the B3LYP/6-311++G** level of theory*, Carbohydr. Res. 341 (2006), pp. 525-537.

[230] Y.K. Sturdy, C.-K. Skylaris and D.C. Clary, *Torsional anharmonicity in the conformational analysis of β-D-galactose*, J. Phys. Chem. B 110 (2006), pp. 3485-3492.

[231] S. Karabulut and J. Leszczynski, *Anomeric and rotameric preferences of glucopyranose in vacuo, water and organic solvents*, J. Mol. Model. 19 (2013), pp. 3637-3645.

[232] D.C. Solha, T.M. Barbosa, R.V. Viesser, R. Rittner and C.F. Tormena, *Experimental and theoretical studies of intramolecular hydrogen bonding in 3-hydroxytetrahydropyran: Beyond AIM analysis*, J. Phys. Chem. A 118 (2014), pp. 2794-2800.

[233] L.M.J. Kroon-Batenburg, J. Kroon and M.G. Northolt, *Theoretical studies on β(1→4) glucose oligomers as models for native and regenerated cellulose fibers*, Papier 44 (1990), pp. 640-647.

[234] P. Klewinghaus, B.P. van Eijck, M.L.C.E. Kouwijzer and J. Kroon, *Molecular dynamics study of conformational equilibria in aqueous D-glucose and D-galactose*, J. Mol. Struct. 395 (1997), pp. 289-295.

[235] K.N. Kirschner and R.J. Woods, *Solvent interactions determine carbohydrate conformation*, Proc. Natl. Acad. Sci. USA 98 (2001), pp. 10541-10545.

[236] M. Kuttel, J.W. Brady and K.J. Naidoo, *Carbohydrate solution simulations: Producing a force field with experimentally consistent primary alcohol rotational frequencies and populations*, J. Comput. Chem. 23 (2002), pp. 1236-1243.

[237] V. Spiwok and I. Tvaroška, *Metadynamics modelling of the solvent effect on primary hydroxyl rotamer equilibria in hexopyranosides*, Carbohyd. Res. 344 (2009), pp. 1575-1581.

[238] M.D. Battistel, R. Pendrill, G. Widmalm and D. Freedberg, *Direct evidence for hydrogen bonding in glycans: A combined NMR and molecular dynamics study*, J. Phys. Chem. B 117 (2013), pp. 4860â4869.

[239] M. Karplus, *Contact electron-spin coupling of nuclear magnetic moments*, J. Chem. Phys. 30 (1959), pp. 11-15.

[240] M. Karplus, *Vicinal proton coupling in nuclear magnetic resonance*, J. Am. Chem. Soc. 85 (1963), pp. 2870-2871.

[241] C.A.G. Haasnoot, F.A.A.M. de Leeuw and C. Altona, *The relationship between proton-proton NMR coupling constants and substituent electronegativities. I. An empirical generalization of the Karplus equation*, Tetrahedron 36 (1980), pp. 2783-2792.

[242] M. Tafazzoli and M. Ghiasi, *New Karplus equations for ^2JHH, $^3J_{HH}$, $^2J_{CH}$, $^3J_{CH}$, $^3J_{COCH}$, $^3J_{CSCH}$, and $^3J_{CCCH}$ in some aldohexopyranoside derivatives as determined using NMR spectroscopy and density functional theory calculation*, Carbohydr. Res. 342 (2007), pp. 2086-2096.

[243] D. Bedrov, M. Pekny and G.D. Smith, *Quantum-chemistry-based force field for 1,2-dimethoxyethane and poly(ethylene oxide) in aqueous solution*, J. Phys. Chem. B 102 (1998), pp. 996-1001.

[244] C. Oostenbrink, A. Villa, A.E. Mark and W.F. van Gunsteren, *A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6*, J. Comput. Chem. 25 (2004), pp. 1656-1676.

[245] H.S. Hansen and P.H. Hünenberger, *A reoptimized GROMOS force field for hexopyranose-based carbohydrates accounting for the relative free energies of ring conformers, anomers, epimers, hydroxymethyl rotamers and glycosidic linkage conformers*, J. Comput. Chem. 32 (2011), pp. 998-1032.

[246] W. Plazinski, A. Lonardi and P.H. Hünenberger, *Revision of the GROMOS 56A6$_{CARBO}$ force field: Improving the description of ring-conformational equilibria in hexopyranose-based carbohydrates chains*, J. Comput. Chem. 37 (2016), pp. 354-365.

[247] M.M. Reif, P.H. Hünenberger and C. Oostenbrink, *New interaction parameters for charged amino acid side chains in the GROMOS force field*, J. Chem. Theory Comput. 8 (2012), pp. 3705-3723.

[248] R.D. Lins and P.H. Hünenberger, *A new GROMOS force field for hexopyranose-based carbohydrates*, J. Comput. Chem. 26 (2005), pp. 1400-1412.

[249] J.S. Rowlinson, *The lattice energy of ice and the second virial coefficient of water vapour*, Trans. Faraday Soc. 47 (1951), pp. 120-129.

[250] C.M. Hansen, *Hansen Solubility Parameters: A User's Handbook Edition 2nd Edition.*, CRC Press, Boca Raton, FL (2007).

[251] J.A. Riddick, W.B. Bunger and T.K. Sakano, *Organic solvents, physical properties and methods of purification*, John Wiley & Sons, New York, USA (1986).

[252] R. Walser, A.E. Mark, W.F. van Gunsteren, M. Lauterbach and G. Wipff, *The effect of force-field parameters on properties of liquids: Parametrization of a simple three-site model for methanol*, J. Chem. Phys. 112 (2000), pp. 10450-10459.

[253] R.C. Weast, *Handbook of chemistry and physics Volume XXX.*, The Chemical Rubber Co, USA (1972).

[254] I.G. Tironi and W.F. van Gunsteren, *A molecular dynamics study of chloroform*, Mol. Phys. 83 (1994), pp. 381-403.

[255] Z. Lin, A.-P. Kunz and W.F. van Gunsteren, *A one-site polarizable model for liquid chloroform: COS/C*, Mol. Phys. 108 (2010), pp. 1749-1757.

[256] I.G. Tironi, P. Fontana and W.F. van Gunsteren, *A molecular dynamics simulation study of liquid carbon tetrachloride*, Mol. Sim. 18 (1996), pp. 1-11.

[257] R.E. Reeves, *The shape of pyranoside rings*, J. Am. Chem. Soc. 72 (1950), pp. 1499-1506.

[258] S.J. Angyal, V.A. Pickles and R. Ahluwahlia, *The interaction between an axial methyl an an axial hydroxyl group in pyranoses*, Carbohydr. Res. 1 (1966), pp. 365-370.

[259] R.U. Lemieux and J.D. Stevens, *The proton magnetic resonance spectra and tautomeric equilibria of aldoses in deuterium oxide*, Can. J. Chem. 44 (1966), pp. 249-262.

[260] S.J. Angyal, *Conformational analysis in carbohydrate chemistry. I. Conformational free energies. The conformation and α:β ratios of aldopyranoses in aqueous solution*, Aust. J. Chem. 21 (1968), pp. 2737-2746.

[261] S.J. Angyal, *The composition and conformation of sugars in solution*, Angew. Chem. Int. Ed. 8 (1969), pp. 157-166.

[262] J. Augé and S. David, *Hexopyranose sugars conformation revisited*, Tetrahedron 40 (1984), pp. 2101-2106.

[263] F. Franks, P.J. Lillford and G. Robinson, *Isomeric equilibria of monosaccharides in solution. Influence of solvent and temperature*, J. Chem. Soc. Farad. Trans. I 85 (1989), pp. 2417-2426.

[264] M. Lingenheil, R. Denschlag, R. Reichold and P. Tavan, *The "hot-solvent/cold-solute" problem revisited*, J. Chem. Theory Comput. 4 (2008), pp. 1293-1306.

[265] P. Dais and R.H. Marcheussault, *^{13}C nuclear magnetic resonance relaxation of amylose and dynamic behavior of the hydroxymethyl group*, Macromolecules 24 (1991), pp. 4611-4614.

[266] P.J. Hajduk, D.A. Horita and L.E. Lerner, *Picosecond dynamics of simple monosaccharides as probed by NMR and molecular dynamics simulations*, J. Am. Chem. Soc. 115 (1993), pp. 9196-9201.

[267] L. Poppe, *Modelling carbohydrate conformation from NMR data: Maximum entropy rotameric distribution about the C5-C6 bond in gentiobiose*, J. Am. Chem. Soc. 115 (1993), pp. 8241-8426.

[268] J. Stenger, M.K. Cowman, F. Eggers, E.M. Eyring, U. Kaatze and S. Petrucci, *Molecular dynamics and kinetics of monosaccharides in solution. A broadband ultrasonic relaxation study*, J. Phys. Chem. B 104 (2000), pp. 4782-4790.

[269] R. Behrends and U. Kaatze, *Molecular dynamics and conformational kinetics of mono- and disaccharides in aquous solution*, Chem. Phys. Chem. 6 (2005), pp. 1133-1145.

[270] H.S. Hansen, X. Daura and P.H. Hünenberger, *Enhanced conformational sampling in molecular dynamics simulations of solvated peptides: fragment-based local elevation umbrella sampling*, J. Chem. Theory Comput. 6 (2010), pp. 2598-2621.

[271] H.M. Pickett and H.L. Strauss, *Conformational structure, energy and inversion rates of cyclohexane and some related oxanes*, J. Am. Chem. Soc. 92 (1970), pp. 7281-7290.

[272] C. Altona, R. Francke, R. de Haan, J.H. Ippel, G.J. Daalmans, A.J.A. Hoekzema and J. van Wijk, *Empirical group electronegativities for vicinal NMR proton-proton couplings along a C-C Bond : Solvent effects and reparameterization of the Haasnoot equation*, Magn. Reson. Chem. 32 (1994), pp. 670-678.

[273] M.L. Huggins, *Bond Energies and Polaritie*, J. Am. Chem. Soc. 75 (1953), pp. 4123-4126.

[274] M. Christen, A.-P.E. Kunz and W.F. van Gunsteren, *Sampling of rare events using hidden restraints*, J. Phys. Chem. B 110 (2006), pp. 8488-8498.

[275] E. Autieri, M. Sega, F. Pederiva and G. Guella, *Puckering free energy of pyranoses: a NMR and metadynamics-umbrella sampling investigation*, J. Chem. Phys. 133 (2010), pp. 095104/1-095104/14.

[276] L. Pol-Fachin, V.H. Rusu, H. Verli and R.D. Lins, *GROMOS 53A6$_{GLYC}$, an improved force field for hexopyranose-based carbohydrates*, J. Chem. Theory Comput. 8 (2012), pp. 4681-4690.

[277] P.L. Polavarapu and C.S. Ewig, *Ab initio computed molecular structures and energies of the conformers of glucose*, J. Comput. Chem. 13 (1985), pp. 1255-1261.

[278] C.J. Cramer and D.G. Truhlar, *Quantum chemical conformational analysis of glucose in aqueous solution*, J. Am. Chem. Soc. 115 (1993), pp. 5745-5753.

[279] S.E. Barrows, J.W. Storer, C.J. Cramer, A.D. French and D.G. Truhlar, *Factors controlling relative stability of anomers and hydroxymethyl conformers of glucopyranose*, J. Comput. Chem. 19 (1998), pp. 1111-1129.

[280] J.M. Silla, R.A. Cormanich, R. Rittner and M.P. Freitas, *Does intramolecular hydrogen bond play a key role in the stereochemistry of α- and β-D-glucose*, Carbohydr. Res. 396 (2014), pp. 9-13.

[281] J.S. Lomas, ^1H *NMR spectra of butane-1,4-diol and other 1,4-diols: DFT calculation of shifts and coupling constants*, Magn. Reson. Chem. 52 (2014), pp. 87-97.

[282] J.G. Mathieson and B.E. Conway, *Partial molal compressibilities of salts in aqueous solution and assignment of ionic contributions*, J. Solut. Chem. 3 (1974), pp. 455-477.

[283] I. Tvaroška and J. Gajdoš, *Angular dependence of vicinal carbon-proton coupling constants for conformational studies of the hydroxymethyl group in carbohydrates*, Carbohydr. Res. 271 (1995), pp. 151-162.

[284] J. Stangret and T. Gampe, *Ionic hydration behavior derived from infrared spectra in HDO*, J. Phys. Chem. A 106 (2002), pp. 5393-5402.

[285] D. Kony, W. Damm, S. Stoll and W.F. van Gunsteren, *An improved OPLS-AA force field for carbohydrates*, J. Comput. Chem. 23 (2002), pp. 1416-1429.

[286] G.H. Ko and W.H. Fink, *Rapidly converging lattice sums for nonelectrostatic interactions*, J. Comput. Chem. 23 (2002), pp. 477-483.

[287] S.A.H. Spieser, J.A. van Kuik, L.M.J. Kroon-Batenburg and J. Kroon, *Improved carbohydrate force field for GROMOS: ring and hydroxymethyl group conformations and* exo-*anomeric effect*, Carbohydr. Res. 322 (1999), pp. 264-273.

[288] A. de Bruyn and M. Anteunis, *Conformation of the C-5/C-6 fragment of hexopyranoses*, Carbohydr. Res. 47 (1976), pp. 311-314.

[289] N. Desrosiers and J.E. Desnoyers, *Enthalpies, heat capacities, and volumes of transfer of tetrabutylammonium ion from water to aqueous mixed-solvents from point of view of scaled-particle theory*, Can. J. Chem. 54 (1976), pp. 3800-3808.

[290] A. de Bruyn, M. Anteunis and P. Kováč, *Rotameric behaviour of methoxy groups in some aldopyranose*, Collection Czechoslov. Chem. Commun. 42 (1977), pp. 357-3068.

[291] S.J. Perkins, L. Johnson and D.C. Phillips, *High resolution ^1H- and ^{13}C-NMR spectra of D-glucopyranose, 2-acetamido-1-deoxy-D-glucopyranose, and related compounds in aqueous medi*, Carbohydr. Re. 59 (1977), pp. 19-34.

[292] M-O. Portman and G. Birch, *Sweet taste and solution properties of α,α-trehalos*, J. Sci. Food Agric. 69 (1995), pp. 275-281.

[293] M.U. Roslund, P. Tähtinen, M. Niemitz and R. Sjöholm, *Complete assignments of the ^1H and ^{13}C chemical shifts and $J_{H,H}$ coupling constants in NMR spectra of D-glucopyranose and all D-glucopyranosyl-D-glucopyranosides*, Carbohydr. Res. 343 (2008), pp. 101-112.

[294] K.A. Dill, *Determinant forces in protein folding*, Biochemistry 29 (1990), pp. 7133-7155.

[295] B. Honig and A.-S. Yang, *Free energy balance in protein folding*, Adv. Prot. Chem. 46 (1995), pp. 27-58.

[296] A.J. Doig and M.J.E. Sternberg, *Side-chain conformational entropy in protein folding*, Protein Sci. 4 (1995), pp. 2247-2251.

[297] A.-S. Yang and B. Honig, *Free energy determinants of secondary structure formation: I. alpha-helices*, J. Mol. Biol. 252 (1995), pp. 351-365.

[298] A.-S. Yang and B. Honig, *Free energy determinants of secondary structure formation: II. antiparallel beta-sheets*, J. Mol. Biol. 252 (1995), pp. 366-376.

[299] A.-S. Yang and B. Honig, *Free energy determinants of secondary structure formation: III. beta-turns and their role in protein folding*, J. Mol. Biol. 259 (1996), pp. 873-882.

[300] K.A. Dill and S. Bromberg, *Molecular Driving Forces, Statistical Thermodynamics in Chemistry and Biology*, Garland Science, New York, USA (2003).

[301] R.L. Baldwin, *Energetics of protein folding*, J. Mol. Biol. 371 (2007), pp. 283-301.

[302] K.H. Gardner and J. Blackwell, *The structure of native cellulose*, Biopolymers 13 (1974), pp. 1975-2001.

[303] K.H. Gardner and J. Blackwell, *The hydrogen bonding in native cellulose*, Biochim. Biophys. Acta 343 (1974), pp. 232-237.

[304] F.J. Kolpak and J. Blackwell, *Determination of structure of cellulose II*, Macromolecules 9 (1976), pp. 273-278.

[305] A. Sarko and H.-C.H. Wu, *The crystal structures of A-, B- and C-polymorphs of amylose and starch*, Starch 30 (1978), pp. 73-78.

[306] S.S.C. Chu and G.A. Jeffrey, *The refinement of the crystal structures of β-D-glucose and cellobiose*, Acta Crystallogr. B 24 (1968), pp. 830-838.

[307] J.T. Ham and D.G. Williams, *The crystal and molecular structure of methyl β-cellobioside-methano*, Acta Crystallogr. B 26 (1970), pp. 1373-1383.

[308] G.M. Lipkind, A.S. Shashkov and N.K. Kochetkov, *Nuclear Overhauser effect and conformational states of cellobiose in aqueous solution*, Carbohydr. Res. 141 (1985), pp. 191-197.

[309] H. Sugiyama, K. Hisamichi, T. Usui, K. Sakai and J. Ishiyama, *A study of the conformation of β-1,4-linked glucose oligomers, cellobiose to cellohexaose, in solution*, J. Mol. Struct. 556 (2000), pp. 173-177.

[310] H. Sugiyama, T. Nitta, M. Horri, K. Motohashi, K. Sakai, T. Usui, K. Hisamichi and J. Ishiyama, *The conformation of α-(1→4)-linked glucose oligomers from maltose to maltoheptaose and short-chain amylose in solution*, Carb. Res. 325 (2000), pp. 177-182.

[311] F. Corzana, M.S. Motawia, C. Hervé du Penhoat, S. Perez, S.M. Tschampel, R.J. Woods and S.B. Engelsen, *A hydration study of (1→4) and (1→6) linked α-glucans by comparative 10 ns molecular dynamics simulations and 500-MHz NMR*, J. Comput. Chem. 25 (2004), pp. 573-586.

[312] K. Matsuo and K. Gekko, *Vacuum-ultraviolet circular dichroism study of saccharides by synchrotron radiation spectrophotometry*, Carbohydr. Res. 339 (2004), pp. 591-597.

[313] U. Olsson, A.S. Serianni and R. Stenutz, *Conformational analysis of β-glycosidic linkages in ^{13}C-labeled glucobiosides using inter-residue scalar coupling constant*, J. Phys. Chem. B 112 (2008), pp. 4447-4453.

[314] T.L. Beck, *The influence of water interfacial potentials on ion hydration in bulk water and near interfaces*, Chem. Phys. Lett. 561-562 (2013), pp. 1-13.

[315] G.M. Lipkind, V.E. Verovsky and N.K. Kochetkov, *Conformational states of cellobiose and maltose in solution: a comparison of calculated and experimental data*, Carbohydr. Res. 133 (1984), pp. 1-13.

[316] A.D. French, A-M. Kelterer, G.P. Johnson, M.K. Dowd and C.J. Cramer, *HF/6-31G* energy surfaces for disaccharide analogs*, J. Comput. Chem. 22 (2001), pp. 65-78.

[317] W.B. Bosma, M. Appell, J.L. Willett and F.A. Momany, *Stepwise hydration of cellobiose by DFT methods. 1. Conformational and structural changes brought about by the addition of one to four water molecules*, J. Mol. Struct. THEOCHEM 776 (2006), pp. 1-19.

[318] W.B. Bosma, M. Appell, J.L. Willett and F.A. Momany, *Stepwise hydration of cellobiose by DFT methods. 2. Energy contributions to the relative stabilities of cellobiose·$(H_2O)_{1-4}$ complexes*, J. Mol. Struct. THEOCHEM 776 (2006), pp. 21-31.

[319] C.S. Pereira, D. Kony, R. Baron, M. Müller, W.F. van Gunsteren and P.H. Hünenberger, *Conformational and dynamical properties of disaccharides in water: A molecular dynamics study*, Biophys. J. 90 (2006), pp. 4337-4344.

[320] C.A. Stortz and A.D. French, *Disaccharide conformational maps: adiabaticity in analogues with variable ring shapes*, Mol. Simul. 34 (2008), pp. 373-389.

[321] A.D. French, G.P. Johnson, C.J. Cramer and G.I. Csonka, *Conformational analysis of cellobiose by electronic structure theories*, Carbohydr. Res. 350 (2012), pp. 68-76.

[322] R.A. Jacobson, J.A. Wunderlich and W.N. Lipscomb, *The crystal and molecular structure of cellobiose*, Acta Crystallogr. 14 (1961), pp. 598-607.

[323] B.R. Leeflang, J.F.G. Vliegenthart, L.M.J. Kroon-Batenburg, B.P. van Eijck and J. Kroon, *A H-NMR and MD study of intramolecular hydrogen bonds in methyl β-cellobioside*, Carbohydr. Res. 230 (1992), pp. 41-61.

[324] R.B. Best, G.E. Jackson and K.J. Naidoo, *Molecular dynamics and NMR study of the $\alpha(1\rightarrow4)$ and $\alpha(1\rightarrow6)$ glycosidic linkages: Maltose and isomaltose*, J. Phys. Chem. B 105 (2001), pp. 4742-4751.

[325] R. Polacek, J. Stenger and U. Kaatze, *Chair-chair conformational flexibility, pseudorotation, and exocyclic group isomerization of monosaccharides in water*, J. Chem. Phys. 116 (2002), pp. 2973-2982.

[326] R. Hagen and U. Kaatze, *Conformational kinetics of disaccharides in aqueous solutions*, J. Chem. Phys. 120 (2004), pp. 9656-9664.

[327] N.W.H. Cheetham, P. Dasgupta and G.E. Ball, *NMR and modelling studies of disaccharide conformation*, Carbohydr. Res. 338 (2003), pp. 955-962.

[328] C.A. Stortz, G.P. Johnson, A.D. French and G.I. Csonka, *Comparison of different force fields for the study of disaccharides*, Carbohyd. Res. 344 (2009), pp. 2217-2228.

[329] W. Plazinski and M. Drach, *The influence of the hexopyranose ring geometry on the conformation of glycosidic linkages investigated using molecular dynamics simulations*, Carbohydr. Res. 415 (2015), pp. 17â27.

Curriculum Vitae

Personal Data

Name	Alice Lonardi
Birth	July 24, 1983, Verona (Italy)
Nationality	Italian

Education

2011-2016	Doctor of Science, ETH Zürich, Switzerland
2005-2008	Master of Science, Università di Trento, Trento, Italy
2002-2006	Bachelor of Science, Università di Trento, Trento, Italy
1997-2002	Dipolma classico, Liceo Classico Don Nicola Mazza, Verona, Italy