

Alma Mater Studiorum – Università di Bologna
in cotutela con L'École Nationale Supérieure de Chimie de Montpellier (ENSCM),
France

DOTTORATO DI RICERCA IN

CHIMICA

Ciclo XXXIV

Settore Concorsuale: 03/C2

Settore Scientifico Disciplinare: CHIM/04

**PHYSICAL CHEMISTRY OF THE INTERACTIONS BETWEEN
MULTICOMPONENT SOLVENTS AND BIOMASS**

Presentata da: Sonia Milena Aguilera Segura

Coordinatore Dottorato

Prof. Domenica Tonelli

Supervisore

**Dr. Tzonka Mineva
Prof. Stefania Albonetti**

**Co-Supervisore
Prof. Francesco Di Renzo**

Esame finale anno 2019

THESE POUR OBTENIR LE GRADE DE DOCTEUR DE L'ECOLE NATIONALE SUPERIEURE DE CHIMIE DE MONTPELLIER

En Chimie et Physico-Chimie des Matériaux

École doctorale 459 – Sciences Chimiques Balard

Unité de recherche – Institut Charles Gerhardt de Montpellier (ICGM) – UMR5253

En partenariat international avec l'Université de Bologne, ITALIE

Chimie-physique des interactions entre solvants multicomposants et biomasse

Présentée par **Sonia Milena AGUILERA SEGURA**

Le 27 novembre 2019

Sous la direction de **Tzonka MINEVA,**
Francesco DI RENZO et **Stefania ALBONETTI**

Devant le jury composé de

Mme Emilia SICILIA, professeure associée, Università della Calabria
M. Yoshiharu NISHIYAMA, DR CNRS HDR, CERMAV, Grenoble
Mme Barbara BONELLI, professeure associé, Politecnico di Torino
M. Marco GARAVELLI, professeur, Università di Bologna
M. Jean-Michel GUENET, DR CNRS, Université de Strasbourg
M. Eric CLOT, DR CNRS, Université de Montpellier
M. Ivan RIVALTA, professeure associé, Università di Bologna
Mme Stefania ALBONETTI, professeure associée, Università di Bologna
M. Francesco DI RENZO, DR CNRS ENSCM, Montpellier
Mme Tzonka MINEVA, DR CNRS, ENSCM, ICGM, Montpellier

Rapportrice
Rapporteur
Examinatrice
Examineur
Examineur
Examineur
Membre invité
Co-directrice de thèse
Co-directeur de thèse
Directrice de thèse

THESIS TO OBTAIN THE GRADE OF DOCTOR OF PHILOSOPHY OF THE ÉCOLE NATIONALE SUPÉRIEURE DE CHIMIE DE MONTPELLIER AND THE UNIVERSITÀ DI BOLOGNA

In Chimie et Physico-Chimie des Matériaux- École doctorale 459 Chimie Balard

In Chimica Industriale- Scuola Dottorale di Chimica

In the framework of the Erasmus Mundus Joint Doctorate SINCEM (Sustainable Industrial Chemistry)

Physical chemistry of the interactions between multicomponent solvents and biomass

Presented by **Sonia Milena AGUILERA SEGURA**
The 27th of November 2019

Under the supervision of **Tzonka MINEVA,**
Francesco DI RENZO, and **Stefania ALBONETTI**

In front of the commission

Mme Emilia SICILIA, professore associato, Università della Calabria
M. Yoshiharu NISHIYAMA, DR CNRS HDR, CERMAV, Grenoble
Mme Barbara BONELLI, professore associato, Politecnico di Torino
M. Marco GARAVELLI, professore ordinario, Università di Bologna
M. Jean-Michel GUENET, DR CNRS, Université de Strasbourg
M. Eric CLOT, DR CNRS, Université de Montpellier
M. Ivan RIVALTA, professore associato, Università di Bologna
Mme Stefania ALBONETTI, professore associato, Università di Bologna
M. Francesco DI RENZO, DR CNRS ENSCM, Montpellier
Mme Tzonka MINEVA, DR CNRS, ENSCM, ICGM, Montpellier

Reviewer
Reviewer
Examiner
Examiner
Examiner
Examiner
Invited member
Thesis Co-director
Thesis Co-director
Thesis Director



Abstract

This research project has been carried out in the Sustainable Industrial Chemistry (SINCHEM) Erasmus Mundus Joint Doctorate. The research activity has been conducted both at the *Ecole Nationale Supérieure de Chimie de Montpellier (ENSCM)*, as home university, under the supervision of Dr. Tzonka Mineva and Dr. Francesco Di Renzo, and at the *University of Bologna*, as host university, under the supervision of Prof. Stefania Albonetti in collaboration with Prof. Marco Garavelli and Prof. Ivan Rivalta. The objectives of this thesis focus on atomistic understanding of the interactions between lignocellulosic biomass and mixed solvents, studied with molecular dynamic (MD) simulations and quantum chemical (QM) approaches. The work is motivated by the experimental research carried out in our team (ICGM-MACS) on the fractionation of wood with mixed water/organic solvents. The observed unusual swelling of wood in a series of mixed aqueous solvents led to ***the hypothesis that synergic interactions of solvent components with biomass play an essential role in the cohesion of composite biomaterials***, a subject of high fundamental and technological relevance.

The project has been divided into the followings parts i) study of structures, intermolecular interactions, and polarity in terms of chemical hardness, of binary water-organic solvents (ethanol, isopropanol, and acetonitrile) by MD ii) study of the solvent effect on conformations, role of inter- and intramolecular hydrogen bonds(HB) in the considered solvents and dynamics of lignocellulose compounds (cellulose, lignin, and xylan) by MD, QM, and QM/MM methods; iii) study of the solvent effect on the interactions between cellulose-lignin and xylose-lignin complexes by MD; and iv) Ab-initio spectroscopic features of lignocellulose compounds: the case of UV-VIS spectra of lignin dimer conformers.

First, a systematic comparative study of structural and thermodynamic properties of the selected pure solvents (water, ethanol, acetonitrile, and isopropanol) and their binary aqueous mixtures (25, 50, and 75wt%) was carried out with MD simulations. Also, a simple method for the averaged, dynamic, total hardness computations is suggested. The computational methods and detailed results of this work can be found in ref.[1]. The obtained thermodynamic properties are in a good agreement with the experimental and previous theoretical studies, demonstrating the excellent performance of the TIP4P (water) and CHARMM36 (organic solvents) potentials for these types of binary mixtures. The analysis of the average numbers of HBs with the liquid composition shows that alcohol molecules tend to substitute water molecules, allowing compensating for the loss of H-bonds in the water solvent domains. Acetonitrile is not an H-bond donor, and it is inferred that the presence of water and the formation of H-bonds helps it to reduce the strong dipole-dipole interactions while preserving the microheterogeneity of the solvent mixture. The averaged hardness, similar to the other dynamic and thermodynamic properties, has a nonlinear profile with the solvent compositions. As a general trend, a softening of the water solvent by introducing the organic cosolvents is obtained with the only

exception for the isopropanol-water mixture at $X = 0.75$. The proposed method could be applicable within the HSAB principle to rationalise the behaviour of complex solute systems in mixed solvents.

Next, Molecular Dynamics (MD) simulations were carried out in order to study the structure and dynamics of interactions between the solvents and lignocellulose model compounds (cellulose, lignin, and xylose). A thorough description of the computational details and results can be found in refs. [2]. These were characterised by solvent-accessible surface areas (SASA), radial distribution functions (RDF), torsion angles, H-bonding and H-bond lifetimes, and intermolecular energies. Also, we analysed the atom polarisation of cellulose chains and guaiacyl tetramers employing density functional theory (DFT) calculations, which was found to correlated with the observed preferred solvent – lignocelluloses interactions gathered from MD simulations. Furthermore, we analysed the stability of guaiacyl dimers by QM/MM simulations using the Born-Oppenheimer Molecular dynamics (BOMD) simulations for lignin, coupled to MD simulation of the solvent molecules (BOMD/MD). This allowed to infer that the stability of the lignin dimers in the solvents follows the order $0.75 \text{ EtOH} > \text{Water} > \text{EtOH} > 0.75 \text{ ACN} > \text{ACN}$. The analysis of the BOMD/MD results together with the conclusions from MD simulations led to establish that solvent interactions can stabilise the lignin structure by gaining energy due to H-bonds formation with the solvent molecules. This compensates the loss of intramolecular H-bonds in the presence of solvents. In the case of 75% ethanol, the lignin-solvent interactions are, therefore, maximised and stabilised extended lignin conformation. The stacked lignin conformer in water, however, gains stability by reducing the surface-accessible area in order to minimise the hydrophobic interactions with water. Despite the relatively small lignin models, our results are in a qualitative agreement with recent studies that show an increase of lignin solubility in water-ethanol mixtures[3]. The SASA distributions analysis shows an increase of the SASA of cellulose and lignin in the binary mixtures.

The torsion angle analysis of the xylose glycosidic bond has shown that xylose adopts a left-handed threefold conformation in water and diluted mixtures, whereas a two-fold conformation is favoured in pure and concentrated organic solvents. The analysis of the β -O-4 torsion angle in lignin shows a dihedral conformation distribution in pure water and diluted ethanol which corresponds to a stacked conformation. In contrast, a more extended T-stacked conformation is found in pure and concentrated organic solvents. Furthermore, analysis of the site-to-site RDFs shows a preferential orientation of solvent molecules toward specific oxygens of the glucose monomer. This preferential solvent- $\text{O}_{\text{cellulose}}$ binding varies with the nature of the organic solvent and the water content. For example, in water-acetonitrile solvent, the integration of the RDFs revealed a solvent phase separation at the cellulose interface, promoting the existence of water-rich and acetonitrile-rich cellulose interfaces. Structure analysis indicates that chain deformation obeys to the selective presence of solvents in the different interfaces of cellulose. This deformation is a response of conformational transitions of the hydroxymethyl chain due to solvation. *Consequently, intermolecular hydrogen bonds, as well as their dynamics, vary as a function of co-solvent concentration.* The atomic charges (computed with DFT-Mulliken

population scheme) of cellulose and lignin oxygens in the presence of water are considerably affected by the solvent, whereas the effect is significantly reduced in the presence of acetonitrile.

The effect of binary water-organic solvents on the adhesion of lignocellulose components was studied with molecular dynamics simulations, considering ethanol and acetonitrile as organic co-solvents in binary water-organic solvent mixtures. The lignin-xylan and lignin-cellulose binding is significantly disrupted in the mixed solvents, which could be explained with preferential interactions between water or organic solvent components and the individual polymers. In pure water, the hydrophobic nature of lignin favours the lignin affinity toward cellulose surface. Thus lignin-cellulose interactions are enhanced, and long-lasting H-bonds between lignin and cellulose are observed. On the contrary, xylose-lignin interactions are rapidly interrupted upon hydration of the xylose chain. The RDF analysis and interaction energies reveal that water-organic solvent mixtures act together to disrupt more effectively the cellulose-lignin interactions. However, our analyses show that the presence of the organic phase can enhance or slow down the detachment between xylose and lignin. This suggests that a detailed description of the intermolecular interactions between lignocellulose models and solvent molecules can provide better insight into the effect of water-organic mixtures on the xylan-lignin and lignin-cellulose adhesion.

The aromatic nature of lignin makes it a great candidate for UV-VIS spectra studies, as each of its aromatic rings provides native local probes for tracking lignin dynamics in solution. Two-dimensional (2D) UV electronic spectroscopy (2DES) is a recently developed technique, which can disentangle signals arising from different intermolecular interactions that remain hidden in the 1D pump-probe spectra. 2DUV electronic spectroscopy can target the π - π^* transitions of aromatic residues, and it can be employed for tracking all sorts of localised, delocalised, energy and charge transfer photoinduced phenomena; however, their interpretation proves to be challenging. Alternatively, computational simulation of 2DES can be used to separate the different peaks into their contributing specific electronic transitions, thus providing a tool to recognise, understand and interpret the observed primary fingerprints [4].

In this work, we studied the electronic structure and conformations in water of a guaiacyl lignin dimer using 2D electronic spectroscopy in the UV-VIS region, using the SOS//QM/MM, with QM = CASSCF and RASSCF. First, vertical excitation energies and 2D-UV/VIS excitation spectra were obtained for all lignin model monomers (guaiacol, creosol, homoveratrole) in gas phase, including all π and lone-pair electrons and orbitals in the active space, and they were further compared against anisole and previously studied benzene and phenol, to aid in excited state assignments. We found an excellent agreement between the 2DEF obtained by both QM schemes for the monomers. Besides, we studied the 2D electronic spectra of stacked and unstacked lignin conformers in solution taken from MD simulations of the lignin dimer in water reported

above. The main result is that charge transfer (CT) states become bright upon stacking, allowing to distinguish between two conformers with different degrees of stacking.

The research work over the three years SINCHEM project led to the following most significant conclusions:

(i) The structures of mixed solvents are disturbed in comparison to their monocomponent counterparts and the mechanism for solvent structure disruption varies according to the solvent nature.

(ii) A method to compute the dynamic hardness values, as estimates of solvents reactivity, is proposed. The dynamic hardness profile has a non-linear evolution with the liquid compositions, similarly to the thermodynamic properties of these non-ideal binary solvents. The solvent mixtures with predominant hydrogen bonding interactions (alcohol-water) are chemically harder (tending to preferably react with hard solutes) than those with predominant dipole-dipole interactions (acetonitrile-water).

(iii) Non-linear effects have also been observed in the solvation of lignin, which reaches a maximum for 50% ethanol and 75% acetonitrile aqueous solution.

(iv) The preferential interaction sites between lignocellulose components and the pure or binary solvent molecules were established in detail. These preferential interactions were evidenced to induce solvent-dependent conformational changes of the lignocellulose compounds. We could, therefore, explain the substantial decrease in cellulose-lignin interactions in the presence of organic components in water. Even more, we could distinguish between ethanol and acetonitrile effects. Our results suggest that the preferential interactions of water and acetonitrile for different sites in both lignin and cellulose results in a more effective disruption of the lignin-cellulose interactions in water-acetonitrile solvents than in ethanol-water mixtures, and in the pure solvents. On the contrary, the addition of the organic phase in the water solvent slows down the xylose-lignin separation.

(v) QM/MM models and methodology has been established to compute the stabilisation of lignocelluloses compounds in solvents, as illustrated in this work on lignin in water, ethanol, acetonitrile and their 75 wt% binary organic-aqueous mixtures, and lignin optical properties.

(vi) Our results provide evidence that the use of the SOS//QM/MM approach employed here is suitable for studying different degrees of aggregation of lignin dimers by 2DUV-vis spectra.

1. Aguilera-Segura, S.M., F. Di Renzo, and T. Mineva, *Structures, intermolecular interactions, and chemical hardness of binary water-organic solvents: a molecular dynamics study*. *J Mol Model*, 2018. **24**: p. 292.

2. Aguilera-Segura, S.M., et al., *Synergistic sorption of mixed solvents in wood cell walls: experimental and theoretical approach*. Macromolecular symposia, 2019. **386**(1): p. 1900022.
3. Goldmann, W.M., et al., *Solubility and fractionation of Indulin AT kraft lignin in ethanol-water media*. Separation and Purification Technology, 2019. **209**: p. 826-832.
4. Rivalta, I., et al., *Ab initio simulations of two-dimensional electronic spectra: The SOS//QM/MM approach*. International Journal of Quantum Chemistry, 2014. **114**(2): p. 85-93.

Abstract

Questo progetto di ricerca è stato condotto nell'ambito del Dottorato congiunto Erasmus Mundus di Chimica Industriale Sostenibile (SINCHEM). L'attività di ricerca si è svolta sia presso l'Ecole Nationale Supérieure de Chimie de Montpellier (ENSCM), come università di provenienza, sotto la supervisione della Dott.ssa Tzonka Mineva e del Dott. Francesco Di Renzo, sia presso l'Università di Bologna, come università ospitante, sotto la supervisione della Prof. Stefania Albonetti in collaborazione con il Prof. Marco Garavelli e il Prof. Ivan Rivalta. Gli obiettivi di questa tesi si concentrano sulla comprensione atomistica delle interazioni tra biomassa lignocellulosica e solventi misti, studiata con simulazioni di dinamica molecolare (MD) e approcci di chimica quantistica (QM). Il lavoro è motivato dalla ricerca sperimentale condotta nel nostro team (ICGM-MACS) sul frazionamento del legno con acqua mista / solventi organici. Il gonfiarsi insolito del legno osservato in una serie di solventi misti acquosi ha portato a formulare l'ipotesi che le interazioni sinergiche dei componenti del solvente con la biomassa svolgano un ruolo essenziale nella coesione dei biomateriali compositi, argomento di elevata rilevanza fondamentale e tecnologica.

Il progetto è stato suddiviso nelle seguenti parti i) studio di strutture, interazioni intermolecolari e polarità in termini di durezza chimica, in solventi binari organico-acqua (etanolo, isopropanolo e acetonitrile) mediante MD ii) studio dell'effetto del solvente su conformazioni, del ruolo dei legami inter- e intra-molecolari dell'idrogeno (HB) nei solventi considerati e della dinamica dei composti lignocellulosici (cellulosa, lignina e xilano) mediante metodi MD, QM e QM / MM; iii) studio dell'effetto del solvente sulle interazioni tra i complessi di cellulosa-lignina e xilosio-lignina mediante MD; e iv) caratteristiche spettroscopiche ab-initio dei composti della lignocellulosa: il caso degli spettri UV-VIS dei conformeri dimeri di lignina.

In primo luogo, uno studio comparativo sistematico delle proprietà strutturali e termodinamiche dei solventi puri selezionati (acqua, etanolo, acetonitrile e isopropanolo) e le loro miscele acquose binarie (25, 50 e 75% in peso) è stato condotto con simulazioni MD. Inoltre, viene suggerito un metodo semplice per i calcoli della durezza media, dinamica e totale. I metodi computazionali ed i risultati dettagliati di questo lavoro sono riportati in ref. [1]. Le proprietà termodinamiche ottenute sono in buon accordo con gli studi sperimentali e teorici precedenti, dimostrando le eccellenti prestazioni dei potenziali del modello TIP4P (acqua) e CHARMM36 (solventi organici) per questi tipi di miscele binarie. L'analisi dei numeri medi di HB con la composizione liquida mostra che le molecole di alcol tendono a sostituire le molecole di acqua, consentendo di compensare la perdita di legami H nei domini dei solventi d'acqua. L'acetonitrile non è un donatore di legami H e si deduce come la presenza di acqua e la formazione di legami H aiuti a ridurre le forti interazioni dipolo-dipolo preservando la microeterogeneità della miscela di solventi. La durezza media, simile alle altre proprietà dinamiche e termodinamiche, ha un profilo non lineare con le composizioni del solvente. Come tendenza generale, si ottiene un addolcimento del solvente in acqua introducendo i cosolventi organici, con la

sola eccezione della miscela isopropanolo-acqua a $X = 0,75$. Il metodo proposto potrebbe essere applicabile nell'ambito del principio HSAB per razionalizzare il comportamento di sistemi complessi di soluti in solventi misti.

Successivamente, sono state eseguite simulazioni di Molecular Dynamics (MD) al fine di studiare la struttura e la dinamica delle interazioni tra i solventi ed il modello dei composti di lignocellulosa (cellulosa, lignina e xilosio). Una descrizione completa dei dettagli computazionali e dei risultati può essere trovata nei riferimenti [2]. Questi erano caratterizzati da aree di superficie accessibili ai solventi (SASA), funzioni di distribuzione radiale (RDF), angoli di torsione, legami H e durata dei legami H, ed energie intermolecolari. In aggiunta, abbiamo analizzato la polarizzazione dell'atomo delle catene di cellulosa e dei tetrameri di guaiacile mediante calcoli con la teoria funzionale della densità (DFT), che sono risultati correlare con le interazioni preferenziali osservate tra solvente e lignocellulose ottenute da simulazioni MD. Inoltre, abbiamo analizzato la stabilità dei dimeri guaiacilici mediante simulazioni QM / MM utilizzando le simulazioni di dinamica molecolare di Born-Oppenheimer (BOMD) per la lignina, insieme alla simulazione MD delle molecole di solvente (BOMD / MD). Ciò ha permesso di dedurre che la stabilità dei dimeri di lignina nei solventi segue l'ordine $0,75 \text{ EtOH} > \text{Acqua} > \text{EtOH} > 0,75 \text{ ACN} > \text{ACN}$. L'analisi dei risultati di BOMD / MD, insieme alle conclusioni delle simulazioni MD, ha portato a stabilire che le interazioni dei solventi possono stabilizzare la struttura della lignina acquisendo energia grazie alla formazione di legami H con le molecole di solvente. Ciò compensa la perdita di legami H intramolecolari in presenza di solventi. Nel caso dell'etanolo al 75%, le interazioni lignina-solvente sono pertanto massimizzate e stabilizzano la conformazione estesa della lignina. Il conformatore impilato di lignina in acqua, tuttavia, guadagna stabilità riducendo l'area accessibile in superficie al fine di ridurre al minimo le interazioni idrofobe con l'acqua. Nonostante i modelli relativamente piccoli di lignina, i nostri risultati sono in accordo qualitativo con studi recenti che mostrano un aumento della solubilità della lignina in miscele di acqua-etanolo [3]. L'analisi delle distribuzioni SASA mostra un aumento del SASA di cellulosa e lignina nel binario miscela.

L'analisi dell'angolo di torsione del legame glicosidico dello xilosio ha mostrato che lo xilosio adotta una conformazione triplice sinistrorsa in acqua e miscela diluite, mentre una conformazione duplice è favorita nei solventi organici puri e concentrati. L'analisi dell'angolo di torsione β -O-4 nella lignina mostra una distribuzione di conformazione diedrica in acqua pura e etanolo diluito che corrisponde a una conformazione impilata. Al contrario, una conformazione impilata a T più estesa si trova in solventi organici puri e concentrati. Inoltre, l'analisi delle RDF per ogni sito mostra un orientamento preferenziale delle molecole di solvente verso specifici atomi di ossigeno del monomero di glucosio. Questo legame preferenziale solvente- $O_{\text{cellulosa}}$ varia in base alla natura del solvente organico ed al contenuto di acqua. Ad esempio, nel solvente acqua-acetonitrile l'integrazione delle RDF ha rivelato una separazione di fase del solvente all'interfaccia della cellulosa, promuovendo l'esistenza di interfacce di cellulosa ricche di acqua e di acetonitrile. L'analisi della

struttura indica che la deformazione della catena obbedisce alla presenza selettiva di solventi nelle diverse interfacce della cellulosa. Questa deformazione è una risposta delle transizioni conformazionali della catena idrossimetilica dovuta alla solvatazione. *Di conseguenza, i legami intermolecolari dell'idrogeno e le loro dinamiche variano in funzione della concentrazione di co-solvente.* Le cariche atomiche (calcolate con lo schema di popolazione DFT-Mulliken) degli atomi di ossigeno di cellulosa e lignina in presenza di acqua sono notevolmente influenzate dal solvente, mentre l'effetto è significativamente ridotto in presenza di acetonitrile.

L'effetto dei solventi binari organici-acquosi sull'adesione dei componenti lignocellulosici è stato studiato con simulazioni di dinamica molecolare, considerando etanolo e acetonitrile come co-solventi organici in miscele binarie di solventi organici-acquosi. Il legame lignina-xilano e lignina-cellulosa è significativamente interrotto nei solventi misti, il che potrebbe essere spiegato con interazioni preferenziali tra l'acqua o i componenti di solventi organici e i singoli polimeri. In acqua pura, la natura idrofobica della lignina favorisce l'affinità della lignina con la superficie della cellulosa, migliorando così le interazioni lignina-cellulosa, e si osservano legami H di lunga durata tra lignina e cellulosa. Al contrario, le interazioni xilosio-lignina vengono rapidamente interrotte all'idratazione della catena del xilosio. L'analisi RDF e le energie di interazione rivelano che le miscele di solvente organico-acquoso agiscono insieme per interrompere più efficacemente le interazioni cellulosa-lignina. Tuttavia, le nostre analisi mostrano che la presenza della fase organica può migliorare o rallentare il distacco tra xilosio e lignina. Ciò suggerisce che una descrizione dettagliata delle interazioni intermolecolari tra i modelli di lignocellulosa e le molecole di solvente può fornire una migliore comprensione dell'effetto delle miscele idro-organiche sull'adesione di xilan-lignina e lignina-cellulosa.

La natura aromatica della lignina la rende un ottimo candidato per gli studi sugli spettri UV-VIS, poiché ciascuno dei suoi anelli aromatici fornisce sonde locali native per il monitoraggio della dinamica della lignina in soluzione. La spettroscopia elettronica bidimensionale UV (2DES) è una tecnica, recentemente sviluppata, che può isolare i segnali derivanti da diverse interazioni intermolecolari che rimangono nascoste negli spettri pompa-sonda 1D. La spettroscopia elettronica 2DUV può indirizzare le transizioni π - π^* dei residui aromatici e può essere impiegata per tracciare tutti i tipi di fenomeni fotoindotti localizzati, delocalizzati, di trasferimento di energia e carica; tuttavia, la loro interpretazione si rivela complicata. In alternativa, la simulazione computazionale del 2DES può essere utilizzata per separare i diversi picchi nelle transizioni elettroniche specifiche che contribuiscono, fornendo così uno strumento per riconoscere, comprendere e interpretare le impronte digitali primarie osservate [4].

In questo lavoro, abbiamo studiato la struttura elettronica e le conformazioni in acqua di un dimero di lignina guaiacilica mediante spettroscopia elettronica 2D nella regione UV-VIS, usando SOS // QM / MM, con QM = CASSCF e RASSCF. Innanzitutto, sono state ottenute energie di eccitazione verticale e spettri di eccitazione 2D-UV / VIS per tutti i monomeri del modello di lignina (guaiacolo, creosolo, omoveratrolo) in fase gassosa,

inclusi tutti gli elettroni e gli orbitali π e coppie solitarie nello spazio attivo, e sono stati ulteriormente confrontati con l'anisolo e il benzene ed il fenolo precedentemente studiati, per essere d'aiuto nelle assegnazioni di stati eccitati. Abbiamo trovato un eccellente accordo tra i 2DEF ottenuti da entrambi gli schemi QM per i monomeri. Inoltre, abbiamo studiato gli spettri elettronici 2D di conformatori di lignina impilati e non impilati in soluzione ottenuti da simulazioni MD del dimero di lignina in acqua, sopra riportate. Il risultato principale è che gli stati di trasferimento di carica (CT) diventano luminosi al momento dell'impilamento, consentendo di distinguere tra due conformatori con diversi gradi di impilamento.

Il lavoro di ricerca nel triennio del progetto SINCHEM ha portato alle seguenti conclusioni più significative:

(i) Le strutture dei solventi misti sono disturbate in confronto alle loro controparti monocomponenti, ed il meccanismo per la distruzione della struttura del solvente varia a seconda della natura del solvente.

(ii) Viene proposto un metodo per calcolare i valori di durezza dinamica, come stime della reattività dei solventi. Il profilo di durezza dinamica ha un'evoluzione non lineare con le composizioni liquide, analogamente alle proprietà termodinamiche di questi solventi binari non ideali. Le miscele di solventi con predominanti interazioni di legami idrogeno (alcol-acqua) sono chimicamente più dure (avendo la tendenza a reagire preferibilmente con soluti duri) rispetto a quelle con interazioni dipolo-dipolo predominanti (acetone-acqua).

(iii) Effetti non lineari sono stati osservati anche nella solvatazione della lignina, che raggiunge un massimo per soluzione acquosa di etanolo al 50% e acetone al 75%.

(iv) I siti di interazione preferenziale tra i componenti della lignocellulosa e le molecole pure o binarie di solvente sono stati stabiliti in dettaglio. Queste interazioni preferenziali sono state rilevate indurre cambiamenti conformazionali dei composti lignocellulosici dipendenti dal solvente. Potremmo quindi spiegare la sostanziale riduzione delle interazioni cellulosa-lignina in presenza di componenti organici nell'acqua. Ancor di più, potremmo distinguere tra effetti di etanolo e acetone. I nostri risultati suggeriscono che le interazioni preferenziali di acqua e acetone per siti diversi sia nella lignina che nella cellulosa determinano un'interruzione più efficace delle interazioni lignina-cellulosa nei solventi acqua-acetone rispetto alle miscele etanolo-acqua e nei solventi puri. Al contrario, l'aggiunta della fase organica nel solvente acquoso rallenta la separazione xilosio-lignina.

(v) Sono stati stabiliti modelli e metodologie QM / MM per calcolare la stabilizzazione dei composti lignocellulosici in solventi, come illustrato in questo lavoro sulla lignina in acqua, etanolo, acetone e le loro miscele binarie idro-organiche al 75% in peso, e le proprietà ottiche della lignina .

(vi) I nostri risultati dimostrano che l'uso dell'approccio SOS // QM / MM qui impiegato è adatto per studiare diversi gradi di aggregazione di dimeri di lignina mediante spettro 2DUV-vis.

1. Aguilera-Segura, S.M., F. Di Renzo, and T. Mineva, *Structures, intermolecular interactions, and chemical hardness of binary water–organic solvents: a molecular dynamics study*. J Mol Model, 2018. **24**: p. 292.
2. Aguilera-Segura, S.M., et al., *Synergistic sorption of mixed solvents in wood cell walls: experimental and theoretical approach*. Macromolecular symposia, 2019. **386**(1): p. 1900022.
3. Goldmann, W.M., et al., *Solubility and fractionation of Indulin AT kraft lignin in ethanol-water media*. Separation and Purification Technology, 2019. **209**: p. 826-832.
4. Rivalta, I., et al., *Ab initio simulations of two-dimensional electronic spectra: The SOS//QM/MM approach*. International Journal of Quantum Chemistry, 2014. **114**(2): p. 85-93.

Résumé

Ce projet de recherche a été réalisé dans le cadre du Doctorat en cotutelle Erasmus Mundus en Chimie Industrielle Durable (SINCHEM). L'activité de recherche a été menée à la fois à l'École Nationale Supérieure de Chimie de Montpellier (ENSCM), en tant qu'université d'origine, sous la supervision du Dr. Tzonka Mineva et du Dr. Francesco Di Renzo, et à l'Université de Bologne, en tant qu'université hôte, sous la supervision du Prof. Stefania Albonetti en collaboration avec le Prof. Marco Garavelli et le Prof. Ivan Rivalta. Les objectifs de cette thèse se concentrent sur la compréhension atomistique des interactions entre la biomasse lignocellulosique et les solvants mixtes, étudiés avec des simulations dynamiques moléculaires (MD) et des approches chimiques quantiques (QM). Le projet est motivé par la recherche expérimentale menée au sein de notre équipe (ICGM-MACS) sur le fractionnement du bois avec des solvants mixtes eau/organique. Le gonflement inhabituel du bois observé dans une série de solvants aqueux mélangés a conduit à *l'hypothèse que les interactions synergiques des composants solvants avec la biomasse jouent un rôle essentiel dans la cohésion des biomatériaux composites*, un sujet de grande importance fondamentale et technologique.

Le projet est composé des quatre parties suivantes: i) l'étude des structures, des interactions intermoléculaires et de la polarité en termes de dureté chimique, des solvants binaires eau-organique (éthanol, isopropanol et acétonitrile) par MD; ii) l'étude de l'effet des solvants sur les conformations, du rôle des liaisons inter- et intramoléculaires de l'hydrogène (HB) dans les solvants considérés et de la dynamique des composés lignocellulosiques (cellulose, lignine et xylane) par les méthodes MD, QM et QM/MM; iii) l'étude de l'effet du solvant sur les interactions entre la cellulose-lignine et les complexes xylose-lignine par MD; et finalement, iv) l'étude des caractéristiques spectroscopiques ab-initio des composés lignocellulosiques : cas des spectres UV-VIS des conformateurs de lignine dimer.

La première étape est la réalisation d'une étude orientée vers la comparaison systématique des propriétés structurales et thermodynamiques des solvants purs sélectionnés (eau, éthanol, acétonitrile et isopropanol) et de leurs mélanges aqueux binaires (25, 50 et 75 % en poids) au moyen de simulations MD. Pour y parvenir, une méthode simple pour les calculs de dureté moyenne, dynamique et totale est suggérée. Les méthodes de calcul et les résultats détaillés de ces travaux se trouvent dans la réf. [1]. Les propriétés thermodynamiques obtenues sont en accord avec les études théoriques expérimentales précédentes, démontrant l'excellente performance des champs de force TIP4P (eau) et CHARMM36 (solvants organiques) pour ce type de mélanges binaires. De même, l'analyse des nombres moyens de HB de composition liquide montre que les molécules d'alcool ont tendance à remplacer les molécules d'eau, ce qui permet de compenser la perte de liaisons H dans les domaines des solvants d'eau. L'acétonitrile n'est pas un donneur de liaisons H, on en déduit que la présence d'eau et la formation de liaisons H contribuent à la réduction des fortes interactions

dipôle-dipole tout en préservant la microhétérogénéité du mélange solvant. En outre, la dureté moyenne, similaire aux autres propriétés dynamiques et thermodynamiques, a un profil non linéaire avec les compositions de solvant. Comme tendance générale, on obtient un adoucissement du solvant d'eau en introduisant les cosolvants organiques, à exception du mélange isopropanol-eau à $X = 0,75$. La méthode proposée pourrait être applicable dans le cadre du principe HSAB pour rationaliser le comportement des systèmes à soluté complexe dans les solvants mixtes.

Ensuite, des simulations de dynamique moléculaire (MD) ont été effectuées afin d'étudier la structure et la dynamique des interactions entre les solvants et les composés du modèle lignocellulosique (cellulose, lignine et xylose). Une description approfondie des détails computationnelles et des résultats est disponible dans les références. [2]. Les interactions entre les solvants et les composés ont été caractérisés par des zones de surface accessibles aux solvants (SASA), des fonctions de distribution radiale (RDF), des angles de torsion, des durées de vie des liaisons H et H, et aussi par des énergies intermoléculaires. Également, et s'appuyant sur la théorie fonctionnelle de la densité (DFT), on a analysé la polarisation des atomes des chaînes de cellulose et des tétramères de guaiacyle, d'où on a constaté qu'ils étaient corrélés avec le solvant préféré observé – interactions lignocellulosiques recueillies à partir des simulations MD. De plus, on a analysé la stabilité des dimères de guaiacyle par simulations QM/MM à l'aide des simulations de dynamique moléculaire Born-Oppenheimer (BOMD) pour la lignine, couplées à la simulation MD des molécules de solvant (BOMD/MD). Cela a permis de déduire que la stabilité des dimères de lignine dans les solvants suit l'ordre $0,75 \text{ EtOH} > \text{Eau} > \text{EtOH} > 0,75 \text{ ACN} > \text{ACN}$. D'un autre côté, l'analyse des résultats BOMD/MD ainsi que les conclusions des simulations MD, ont permis d'établir que les interactions des solvants peuvent stabiliser la structure de la lignine en gagnant de l'énergie grâce à la formation de liaisons H avec les molécules de solvant, ce qui compense la perte de liaisons H intramoléculaires en présence de solvants. Dans le cas de l'éthanol à 75wt%, les interactions lignine-solvant sont donc maximisées et stabilisent la conformation lignine étendue. Néanmoins, la lignine empilée respecte dans l'eau gagne la stabilité en réduisant la surface accessible, ce qui permet la minimisation des interactions hydrophobes. Malgré les modèles de lignine relativement petits, les résultats obtenus concordent qualitativement avec les études récentes qui montrent une augmentation de la solubilité de la lignine dans les mélanges eau-éthanol [3]. De plus, l'analyse des distributions SASA montre une augmentation du SASA de cellulose et de lignine dans les mélanges binaires.

La troisième étape est concentrée sur l'analyse de l'angle de torsion de la liaison glycosidique xylose a montré que la xylose adopte une conformation triple gauchère dans l'eau et les mélanges dilués, tandis qu'une conformation double est privilégiée dans les solvants organiques purs et concentrés. L'analyse de l'angle de torsion $\beta\text{-O-4}$ en lignine montre une distribution de conformation dièdre en eau pure et éthanol dilué qui correspond à une conformation empilée. En revanche, une conformation T-empilée plus étendue se trouve dans les solvants organiques purs et concentrés. De plus, l'analyse des fonction de distribution radiale (FDR) de

site à site montre une orientation préférentielle des molécules de solvants vers des oxygènes spécifiques du monomère de glucose. Cette liaison préférentielle solvant- $O_{\text{cellulose}}$ varie selon la nature du solvant organique et la contenu en eau. Par exemple, dans le solvant eau-acétonitrile, l'intégration des FDR a mis en évidence la séparation de phase solvant à l'interface cellulose. L'analyse de structure montre que la déformation de la chaîne de la cellulose solvatée par transition conformationnelle de la chaîne hydroxyméthyle obéit à la présence sélective de solvants dans les différentes interfaces de la cellulose. *Par conséquent, les liaisons intermoléculaires d'hydrogène ainsi que leur dynamique varient en fonction de la concentration du co-solvant.* Les charges atomiques (calculées avec le schéma de population *DFT-Mulliken*) de cellulose et de lignine oxygènes en présence d'eau sont considérablement affectées par le solvant, tandis que l'effet est significativement réduit en présence d'acétonitrile.

L'effet des solvants hydro organiques binaires sur l'adhérence des composants lignocellulosiques a été étudié à travers des simulations de dynamique moléculaire, en considérant l'éthanol et l'acétonitrile comme des co-solvants organiques dans des mélanges binaires eau-solvants organiques. La liaison lignine-xylose et lignine-cellulose est considérablement perturbée dans les solvants mélangés, ce qui pourrait s'expliquer par des interactions préférentielles entre l'eau ou les composants organiques du solvant et les polymères individuels. Cependant, dans l'eau pure, la nature hydrophobe de la lignine favorise l'affinité de la lignine à la surface de la cellulose, ce qui améliore les interactions lignine-cellulose et permet d'observer des liaisons H durables entre la lignine et la cellulose. Autrement, les interactions xylose-lignine sont rapidement interrompues lors de l'hydratation de la chaîne xylose. Non seulement l'analyse du RDF mais aussi les énergies d'interaction révèlent que les mélanges de solvants eau-organique agissent ensemble pour perturber plus efficacement les interactions cellulose-lignine. Pourtant, nos analyses montrent que la présence de la phase organique peut augmenter ou ralentir le détachement entre la xylose et la lignine., ce qui suggère qu'une description détaillée des interactions intermoléculaires entre les modèles de lignocellulose et les molécules de solvant pourrait fournir un meilleur aperçu de l'effet des mélanges eau-organique sur l'adhérence xylo-lignine et lignine-cellulose.

La nature aromatique de la lignine en fait un excellent candidat pour les études spectrales UV-VIS, chacun de ses cycles aromatiques fournit des sondes locales natives pour suivre la dynamique de la lignine en solution. La spectroscopie électronique UV (2DES) bidimensionnelle (2D) est une technique récente permettant de démêler les signaux résultants des différentes interactions intermoléculaires qui restent cachées dans les spectres de la pompe-sonde 1D. La spectroscopie électronique 2DUV peut non seulement cibler les transitions $\pi-\pi^*$ des résidus aromatiques, mais aussi peut être utilisée pour suivre les variétés de phénomènes photo induits localisés, délocalisés, de transfert d'énergie et de charge. Toutefois, leur interprétation s'avère difficile. Par ailleurs, la simulation computationnelle de 2DES peut être utilisée pour séparer les différents pics

des transitions électroniques spécifiques, fournissant ainsi un outil qui permet la reconnaissance, compréhension et interprétions des empreintes digitales primaires observées [4].

Dans ce travail, nous avons étudié la structure électronique et les conformations dans l'eau d'un dimer lignine guaiacyl en utilisant la spectroscopie électronique 2D dans la région UV-VIS, et le SOS//QM/MM, avec QM = CASSCF et RASSCF. D'un côté, des énergies d'excitation verticale et des spectres d'excitation 2D-UV/VIS ont été obtenus pour tous les monomères de lignine (guaiacol, créosol, homovératrole) en phase gazeuse, y compris tous les électrons π et les électrons monopieces et orbitaux dans l'espace actif. Puis, ils ont été comparés à l'anisole et au benzène et phénol, précédemment analysées, dans le but de contribuer aux d'état excité. Nous avons trouvé un excellent accord entre le 2DEF obtenu par les deux systèmes QM pour les monomères. En outre, nous avons étudié les spectres électroniques 2D des aligneurs empilés et non plaqués en solution tirés des simulations MD du dimère de lignine dans l'eau, signalées ci-dessus. Le résultat principal décrit que les états de transfert de charge (CT) deviennent brillants lors de l'empilage, permettant de distinguer entre deux conformes avec des degrés d'empilage différents.

Les travaux de recherche menés au cours des trois années du projet SINCHEM ont abouti aux conclusions suivantes:

- (i) Les structures des solvants mixtes sont perturbées par rapport à leurs contreparties monocomponiques et le mécanisme de perturbation de la structure des solvants varie selon la nature du solvant.
- (ii) Une méthode de calcul des valeurs de dureté dynamique, sous forme d'estimations de la réactivité des solvants, est proposée. Le profil de dureté dynamique et les propriétés thermodynamiques des solvants binaires non idéaux évoluent de façon non linéaire avec les compositions liquides. Les mélanges de solvants avec des interactions prédominantes de liaison hydrogène (alcool-eau) sont chimiquement plus dure (ayant tendance à réagir avec des solutés durs) que ceux avec des interactions dipôle-dipole prédominantes (acétonitrile-eau).
- (iii) Des effets non linéaires ont également été observés dans la solvation de la lignine, qui atteint un maximum de 50wt% d'éthanol et de 75wt% d'acétonitrile en solution aqueuse.
- (iv) Les sites d'interaction préférentiels entre les composants lignocellulosiques et les molécules à solvant pur ou binaire ont été établis en détail. Il a été démontré que ces interactions préférentielles induisaient des changements conformationnels dépendant des solvants des composés lignocellulosiques. Cela pourrait donc expliquer la diminution substantielle des interactions cellulose-lignine en présence de composants organiques dans l'eau. De plus, nous pourrions dissocier les effets de l'éthanol de ceux de l'acétonitrile. Les résultats suggèrent que les interactions préférentielles de l'eau et de l'acétonitrile pour différents sites dans la lignine et

la cellulose entraînent une perturbation plus efficace des interactions lignine-cellulose dans l'eausolvants acétonitrile que dans les mélanges éthanol-eau et les solvants purs. Au contraire, l'ajout de la phase organique dans le solvant à eau ralentit la séparation xylose-lignine.

(v) Des modèles et une méthodologie QM/MM ont été établis pour calculer la stabilisation des composés lignocellulosiques dans les solvants, comme le montrent ces travaux sur l'apprentissage dans l'eau, l'éthanol, l'acétonitrile et leur 75wt% les mélanges aqueux et les propriétés optiques de la lignine.

(vi) Nos résultats montrent que l'utilisation de l'approche SOS//QM/MM utilisée ici est adaptée à l'étude de différents degrés d'agrégation des dimères de lignine par les spectres de 2DUV-vi.

1. Aguilera-Segura, S.M., F. Di Renzo, and T. Mineva, *Structures, intermolecular interactions, and chemical hardness of binary water-organic solvents: a molecular dynamics study*. J Mol Model, 2018. **24**: p. 292.
2. Aguilera-Segura, S.M., et al., *Synergistic sorption of mixed solvents in wood cell walls: experimental and theoretical approach*. Macromolecular symposia, 2019. **386**(1): p. 1900022.
3. Goldmann, W.M., et al., *Solubility and fractionation of Indulin AT kraft lignin in ethanol-water media*. Separation and Purification Technology, 2019. **209**: p. 826-832.
4. Rivalta, I., et al., *Ab initio simulations of two-dimensional electronic spectra: The SOS//QM/MM approach*. International Journal of Quantum Chemistry, 2014. **114**(2): p. 85-93.

Acknowledgements

I would like to thank my supervisors, Dr. Tzonka Mineva and Prof. Francesco di Renzo, for the patient guidance, encouragement and advice they have provided throughout my time as their student. I have been fortunate to have supervisors who cared so much about my work, and who were always available to answer my questions, discuss my results, and share my enthusiasm for my research. In particular, I would like to thank Dr Tzonka Mineva for her dedication and constant feedback in writing this thesis.

I thank Dr. Marco Garavelli for welcoming me in his group and for the intense and rewarding experience. To Dr Ivan Rivalta for all his knowledge and patience while I was learning so many new things. To all the members of Garavelli's group, for everything I learnt from them, and for embracing my love for lasagna and tortellini in brodo.

Being part of the SINCEM program has been the most challenging and fulfilling experience of my life. I thank Prof. Francesco di Renzo for making it possible, and for his endless kindness. Also, I am thankful to Prof. Stefania Albonetti for making possible my exchange with the University of Bologna in the framework of my SINCEM program. For being an amazing "mom" to our SINCEM family. To my SINCEM family, for the gratifying experience, both cultural and scientific, and for the drinks, meals, and moments that we shared together.

To my family, particularly my parents, sister, cousin, and my husband for their encouragement, the proofreading of meaningless pages full of scary terms, and for helping me keep things in perspective. I must express my gratitude to Paolo, my husband, for his constant support and patience through all the ups and downs of my research and mobilities. Completing this work would have been even more difficult if not for the support and the friendships I gained in the last three years. I am grateful with the *theoreticians*, the *mahboul*, *mes coquins*, *les filles*, the member of the MACS team, and my Colombian friends, for all the fun we have had in the last three years and for keeping me sane.

Finally, I would like to thank the ENSCM, not only for providing the funding which allowed me to undertake this research, but also for giving me the opportunity to attend conferences and meet so many interesting people. Also, to the administrative and technical staff at MACS and Bologna for their support.

Table of Contents

Abstract	5
Abstract	10
Résumé	15
Acknowledgements	20
Table of Contents	21
General Introduction.....	26
Project Description	28
References	29
Chapter 1. State of the art.....	37
Abstract	38
1.1 Wood structure and constituents	38
1.1.1 Cellulose	40
1.1.2 Hemicellulose	41
1.1.3 Lignin	42
1.2 Interactions of solvents with lignocellulose biomass.....	43
1.2.1 Pretreatment and fractionation processes	43
1.2.2 Swelling: first step towards polymer dissolution.....	44
1.2 Computational studies of lignocellulose	45
1.3.1 Cellulose	46
1.3.2 Hemicellulose	52
1.3.3 Lignin	53
1.3.4 Lignocellulose interactions	54
References	55
Chapter 2. Theoretical background	66
Abstract	67
1. Fundamentals of Statistical Mechanics.....	67

2.	Molecular Dynamics Simulations	69
2.2	Initialization	70
2.2	Force Calculation	70
2.3	Integrating the Equations of motion	71
2.4	Commonly used software for MD simulations.....	72
2.5	Force Field	72
3.	Quantum Chemistry (QC) background	77
3.1	Density Functional Theory.....	80
3.2	<i>Multiconfigurational Methods</i>	88
3.3	QM program packages used in this work.....	90
4.	QM/MM Hybrid methods.....	91
4.1	QM/MM Energy: subtractive scheme	91
4.2	QM/MM Energy: additive scheme	92
4.3	Mechanical inclusion	92
4.4	Electrostatic inclusion.....	92
4.5	Polarisable inclusion.....	93
4.6	Saturation of broken bonds at the QM / MM boundary.....	93
4.7	QM/MM program packages used in this work.....	93
	REFERENCES	95

Chapter 3. Structures, intermolecular interactions and chemical hardness of binary water- organic solvents 101

	Abstract	102
1.	Introduction.....	102
2.	Methods and computational details	104
2.1	Molecular dynamics simulations	104
2.2	Dynamic total hardness	106
3.	Results and discussion.....	107
3.1	Thermodynamic properties: validation of the MD simulations.....	107
3.2	Structures of the aqueous binary solvents	109
3.3	Analysis of the intermolecular interaction types	113
3.4	Averaged hardness from MD simulations	117
4.	Conclusions.....	119
5.	References.....	120

Chapter 4. Solvent effect on the structure and dynamics of lignocellulose model compounds	127
Abstract	128
1. Introduction.....	128
2. Computational details	130
2.1 Models	130
2.2 Classical Molecular Dynamics simulations.....	131
2.3 Density Functional Theory calculations	136
2.4 Molecular dynamics simulations at the QM/MM level (BOMD/MD).....	136
3. Results and discussion.....	137
3.1 Modelling of the interaction of water-organic solvents with cellulose nanocrystal	137
3.2 Dynamics of xylan tetramer model in water-organic solvents	147
3.3 Dynamics of lignin in water-organic solvents	153
3.4 Polarization effects on lignocellulose polymers, evaluated at the QM (DFT) level.....	160
3.5 MD and BOMD/MD evaluation of the stability of lignin dimer structures	164
4. Conclusions.....	167
References.....	169
Chapter 5. Solvent effect on the adhesion between lignocellulose compounds	174
Abstract	175
1. Introduction.....	175
2. Computational details	177
2.1 Models	177
2.2 Molecular Dynamics simulations	178
3. Results and discussion.....	181
3.1 Solvent effect on the adhesion between lignin and cellulose	181
3.2 Solvent effect on the adhesion between xylan and lignin.....	189
4. Conclusions.....	195
References.....	196
Chapter 6. Discussion: the solvent effect on lignocellulose interactions	201
Abstract	202
1. Cellulose-lignin adhesion: correlation between interaction energies and the solvent's preferential interaction	202

2. Cellulose-solvent interactions: the water exclusion and the entropic effect	204
3. Solvent effects on molecular wood structures: possible implications in lignocellulose adhesion and wood swelling.....	205
4. Assessment of the achieved scope of this work, the reliability of model size and simulation time scales, and future perspectives.....	207
References.....	210
Chapter 7. First-Principles Simulations of Two-Dimensional Electronic Spectroscopy of lignin model monomers and their β -O-4 linked dimer	212
Abstract	213
1. Introduction.....	213
2. Computational methods.....	216
2.1 Selected structures	216
2.2 Excited-State calculations of monomers and dimers.....	217
2.3 Simulations of 2D Electronic Spectra	217
3. Results and discussion.....	218
3.1 Electronic spectra of anisole and lignin monomers.....	218
3.2 CAS vs RAS 2D electronic spectra of lignin monomers.....	223
3.3 2D Electronic spectra of unstacked and stacked lignin dimer conformers.....	224
4. Conclusions.....	229
Chapter 8. Summary of the conclusions and perspectives	234
Supporting Information	240
2. SI Chapter 3. Structures, intermolecular interactions and chemical hardness of binary water- organic solvents.....	242
3.1 Thermodynamic properties of pure solvents.....	242
3.2 Structures of the aqueous binary solvents.....	244
3.3 Hydrogen bond autocorrelation functions.....	245
3.4 Dynamic hardness as a function of solvent composition.....	245
3. SI Chapter 4. Solvent effect on the structure and dynamics of lignocellulose model compounds.....	247
4.1 Cellulose-solvent radial distribution functions in mixed solvents.....	247
4.2 Xylose-solvent radial distribution functions in mixed solvents.....	249

4.3	Lignin-solvent radial distribution functions in mixed solvents.....	251
4.4	Polymer-solvent LJ and Coulomb interactions energies	252
4.	SI Chapter 5. Solvent effect on the adhesion between lignocellulose compounds.....	255
5.1	Cellulose-lignin radial distribution functions.....	255
5.2	LJ, Coulomb and Total xylan-lignin interaction energies.....	256
5.	SI Chapter 6. General discussion: the solvent effect on lignocellulose compounds interactions, conformations, and dynamics.....	257
6.1	Radial distributions of cellulose-solvent and lignin-solvent in cellulose-lignin complex	257
6.2	Number of solvent molecules within 0.7 nm cellulose-solvent shell in cellulose-lignin complex ..	261
6.	SI Chapter 7. First-Principles Simulations of Two-Dimensional Electronic Spectroscopy of lignin model monomers and their β -O-4 linked dimer	262
7.1	Vertical excitation energies of lignin monomers.....	262
7.2	Vertical excitation energies of the unstacked lignin dimer	264
7.3	Vertical excitation energies of the stacked lignin dimer	266
7.4	Comparison of 2D spectra of lignin dimer in gas phase and in water	268
7.5	2D spectra of HVER monomer and monomer G' from the unstacked lignin dimer in gas phase ...	272
7.6	TDM Modified 1-color 2D spectra of unstacked and stacked lignin dimer in solution	273
	References.....	273
	Abstract	275
	Résumé	275

General Introduction

Use of petroleum as a source for fuel and chemicals during the last centuries has provided a cheap way to satisfy the needs of society by leaving, however, a footprint in the environment. The awareness of its negative impact along with the anticipated depletion of petroleum reserves in the next few decades has arisen an intense interest from industry and academy towards finding next-generation feedstocks. Advances in genetics, biotechnology, process chemistry, and engineering are leading to a new manufacturing concept for converting renewable biomass to valuable fuels and chemical products [1]. Wood is highly abundant and can be used as an inexpensive feedstock [2, 3] without compromising the food supply [4] while doing sustainable and responsible use of the resources [5].

Wood –i.e., lignocellulosic biomass- is a natural, composite material constructed from a variety of organic polymers, which possesses unique structural and chemical characteristics rendering it desirable for a broad range of applications [6, 7]. The skeletal material of wood cell walls is cellulose, a long chain, linear sugar polymer composed of glucose monomers and it accounts for about 40-45% of the dry weight of normal wood tissue. According to how it is treated, cellulose can be used in a plethora of applications ranging from paper, film, explosives, plastics, textiles, in addition to having many other industrial uses, such as biofuel production [8-15]. Cellulose is organised into fibrils [16] and interacts with a surrounding matrix formed by hemicellulose and lignin [17, 18]. Hemicellulose, a lower molecular weight polysaccharide that contains short chains of carbohydrates, serves as a matrix substance for the cellulose superstructure. The carbohydrates in hemicellulose represent mostly combinations of various five-carbon sugars (xylose and arabinose) and six-carbon sugars (glucose, mannose, and galactose) [7]. Once isolated, the potential applications for hemicellulose include barrier materials for food packaging or biopolymers with new properties [19], among others [19-25].

Lignin, the third major constituent of wood after hemicelluloses, composes around 25-35% of the total dry weight of the polysaccharide wall substances [7]. It permeates both cell walls and intercellular regions (middle lamella) and plays a significant role in imparting rigidity to the polysaccharide wall substances [7]. Lignin interactions with cellulose determine wood mechanical properties, such as strength, creep, durability and ageing [26]. Lignin is a three dimensional, highly branched, and polyphenolic molecule complex structure with high molecular weight and a variety of lignin units and linkages. The phenylpropanoid unit is the basic structure of lignin, consisting of an aromatic ring and a 3-C side chain [27]. Two types of inter-unit linkages (C-O and C-C linkages) bind several aromatic units, with the β -O-4 ether bond being the predominant inter-unit linkage [28]. Lignin can be used in the production of a variety of chemicals, such as antioxidants, binders,

and dispersants, among others [29-35]. Moreover, minor low-molar-mass extraneous materials, mostly organic extractives and inorganic minerals, are also present in wood (usually 4%-10%)[36]

Disassembling of lignocellulose matter in paper pulping is a significant economic activity, enabling the use of renewable resources to satisfy the world needs of more than 400 Mt of paper and cardboard per year [38]. In most current chemical pulping, the only product targeted is oligocellulose, and the solubilised lignin fraction is slightly more than a waste, which is only thermally valorised in the wet combustion process needed to recycle the salts and the aggressive inorganic chemicals used in the pulping process [39]. The introduction of greener processes for disassembling plant biomass remains a major environmental issue, which could allow a better valorisation of the rich chemistry of lignin components. The higher purity requirement for cellulose used in second-generation ethanol biorefineries is fostering development of organosolv pulping processes, favouring the recovery of lignin coproducts [40-42].

The use of multicomponent solvents has provided remarkable results in fractionation processes of several levels of severity, ranging from swelling, dilute acid, hydrothermal, steam explosion, alkaline treatment, to organosolv pulping [4, 43-53]. In organosolv fractionation [43], for example, wood is in contact with different solvents to produce treated fractions of cellulose pulp, soluble lignin, and hemicelluloses-derived products. Despite all invested efforts, no fractionation method is currently capable of valorising all functionalities of lignocellulose components in an economically viable commercial-scale deconstruction process.

Therefore, the complex and versatile nature of lignocellulose puts fractionation at the core of most successful valorisation processes. Nevertheless, these biopolymers form a complex matrix, intricately cross-linked with a network of interactions. Hemicellulose and lignin are covalently and non-covalently linked, while cellulose and hemicelluloses are connected exclusively by non-covalent interactions. Moreover, the polysaccharides of the cell wall, cellulose and hemicellulose, have a strong affinity for water molecules in either their liquid or vapour state. Lignin, on the other hand, is considered mostly hydrophobic. Therefore, investment for techniques and technologies for lignocellulose deconstruction into its structural constituents represents a challenge since the fractionation process is crucial to the subsequent valorisation of the fractions [43]. Furthermore, to make end processes feasible, the fractionation step should produce easily treated fractions, such as cellulose pulp, soluble lignin, and hemicellulose-derived products (oligomers, sugars, furans, organic acids, etc.) [43].

Understanding of the underlying physicochemical mechanisms is expected to lead to essential innovations in the biomass fractionation technology, rendering bio-sourced synthons and platform molecules available in economically sustainable ways. Understanding the mechanisms of biomass breakdown and its interactions with the solvent medium during fractionation will lead to more efficient use of biomass. Furthermore, the

reduction of the energetic costs of the fractionation processes passes through the optimisation of the interactions between the different components of the biomass and a solvent medium [4]. This has stimulated a significant amount of recent experimental and theoretical investigations of lignocellulosic polymers in mono-component solvents [54-89], aiming to shed light on the microscopic interactions between crystalline or amorphous biopolymers. Moreover, a better understanding of the interactions between biomass components and solvents is also an essential requirement for the design of more effective catalytic processes dealing with renewable resources and could provide useful information for the study of the reactivity of any kind of composites.

Project Description

This work is motivated by the experimental research carried out in our team (ICGM-MACS) on the fractionation of wood with mixed water-organic solvents. Swelling, the first step in wood dissolution, has been studied to understand the solvent effect on the volumetric change of wood in water and organic solvents [90-95]. The observed unusual swelling of wood in a series of mixed aqueous solutions (discussed in more details in Chapter 1- *State of the Art*) has shown not only higher swelling values for selected mixed aqueous solvents, but also a nonlinear relation between the volume swollen and the concentration of the organic component in the solution [90, 94, 96-99]. This evolution of wood volume with the concentration of the organic solvents shows a behaviour similar to gels, with a significant swelling for solutions of intermediate polarity [90]. Thus, these observations have led to the hypothesis that synergic interactions of solvent components with biomass play an important role in the cohesion of composite biomaterials, a subject of high fundamental and technological relevance. There is a consensus in the literature [85, 89] that changes in the morphology of lignocellulosic biomass as exposed to mixed solvent play a vital role in the increase of its solubility, i.e. swelling, and therefore, promotes the fractionation of the material. However, the factors responsible for such behaviour at a molecular scale are still scarcely available. Moreover, understanding the effect of various compositions of multi-component solvents on the molecular structures, reactivity and solute-solvent specific interactions can be most appropriately achieved employing theoretical methods. Molecular-scale studies performed by computational methods are very promising, but the results are scarce. Furthermore, the study of binary mixtures effects on lignocellulosic biomass remains unexplored.

In view of the aforesaid, **the objectives of this thesis focus on atomistic understanding of the interactions between lignocellulosic biomass and binary water-organic solvent mixtures.** To achieve this scope, we carried out predominantly molecular dynamic (MD) simulations on model lignocelluloses compounds in various solvents. We aim to gain knowledge on wood polymer-solvent specific interactions, and lignocellulose properties from an atomistic scale. To this end, we elaborated a computational protocol based on combination of molecular dynamics (MD), combined quantum mechanics/molecular mechanics (QM/MM)

scheme and quantum mechanics (QM) approaches in conjunction with *ab-initio* Born-Oppenheimer molecular dynamics (BOMD) to evaluate the solvent effect on wood molecular structures, their solvent-specific interactions and properties explicitly. The QM density functional theory (DFT) was used to compute the cellulose and lignin atomic polarisation, whereas BOMD/MD was employed to study the lignin stabilisation energy in water, ethanol, acetonitrile, and in their mixtures at 75%. QM/MM scheme was also undertaken for calculations of optical (UV-VIS) characteristic fingerprints of lignin-conformers and to estimate the effect of solvents, using higher-level wave-function QM methods.

This research project has been carried out in the Sustainable Industrial Chemistry (SINCHEM) Erasmus Mundus Joint Doctorate. The research activity has been conducted both at the *Ecole Nationale Supérieure de Chimie de Montpellier (ENSCM)*, as home university, under the supervision of Dr. Tzonka Mineva and Dr. Francesco Di Renzo, and at the *University of Bologna*, as host university, under the guidance of Prof. Stefania Albonetti in collaboration with Prof. Marco Garavelli and Prof. Ivan Rivalta.

The present manuscript is organised in the following way. The state of the art is presented in **Chapter 1**. **Chapter 2** contains a summary of the variety of computational chemistry methods used throughout the work. **Chapter 3** presents the study of the structures and the intermolecular interactions of binary water-organic solvents (ethanol, isopropanol, and acetonitrile) using MD simulations, along with a chemical hardness analysis. In **Chapter 4**, we include the MD study of the solvent effect on wood molecules (cellulose, lignin, and hemicellulose), followed by presentation of the survey of cellulose-lignin and xylan-lignin adhesion in **Chapter 5**. In **Chapter 6**, a general discussion on the effects of different solvent components on the specific intramolecular and intermolecular H-binding, lignocellulose stabilization energies and solvent binding affinities (presented in *Chapter 4*) are discussed together with the results in *Chapter 5* on the lignin-cellulose and lignin-xylan interaction profiles, which permits an emphasis on the major conclusion and hypothesis drawn from the results in *Chapters 4, 5*. In this *Chapter 6*, an attempt of assessment on the validity of the model sizes, used in the simulations, the choice of the force field parameters and the conducted simulation time lengths is presented as well. **Chapter 7** presents the first-principles calculations of two-dimensional electronic spectroscopy of lignin model monomers and the β -O-4 linked dimer. Finally, the conclusions and perspectives are summarised in the **General Conclusions and perspectives** section.

References

1. Ragauskas, A.J., et al., *The Path Forward for Biofuels and Biomaterials*. Science, 2006. **311**(5760): p. 484-489.
2. Sannigrahi, P., A.J. Ragauskas, and G.A. Tuskan, *Poplar as a feedstock for biofuels: A review of compositional characteristics*. Biofuels, Bioproducts and Biorefining, 2010. **4**(2): p. 209-226.

3. van Osch, D.J.G.P., et al., *Ionic liquids and deep eutectic solvents for lignocellulosic biomass fractionation*. Physical Chemistry Chemical Physics, 2017.
4. Grande, P.M., et al., *Fractionation of lignocellulosic biomass using the OrganoCat process*. Green Chemistry, 2015. **17**(6): p. 3533-3539.
5. Mohan, D., C.U. Pittman, and P.H. Steele, *Pyrolysis of Wood/Biomass for Bio-oil: A Critical Review*. Energy & Fuels, 2006. **20**(3): p. 848-889.
6. McKendry, P., *Energy production from biomass (part 1): overview of biomass*. Bioresource Technology, 2002. **83**(1): p. 37-46.
7. Parham, R.A. and R.L. Gray, *Formation and Structure of Wood*, in *The Chemistry of Solid Wood*. 1984, American Chemical Society. p. 3-56.
8. Coseri, S., *Cellulose: To depolymerize... or not to?* Biotechnology Advances.
9. Cao, Y., et al., *Imidazolium-based ionic liquids for cellulose pretreatment: recent progresses and future perspectives*. Applied Microbiology and Biotechnology, 2017. **101**(2): p. 521-532.
10. Tan, H.-T., K.R. Corbin, and G.B. Fincher, *Emerging Technologies for the Production of Renewable Liquid Transport Fuels from Biomass Sources Enriched in Plant Cell Walls*. Frontiers in Plant Science, 2016. **7**(1854).
11. Vanneste, J., et al., *Unconventional Pretreatment of Lignocellulose with Low-Temperature Plasma*. ChemSusChem, 2017. **10**(1): p. 14-31.
12. Patinvoh, R.J., et al., *Innovative pretreatment strategies for biogas production*. Bioresource Technology, 2017. **224**: p. 13-24.
13. Endes, C., et al., *A critical review of the current knowledge regarding the biological impact of nanocellulose*. Journal of Nanobiotechnology, 2016. **14**(1): p. 78.
14. Ummartyotin, S. and H. Manuspiya, *A critical review on cellulose: From fundamental to an approach on sensor technology*. Renewable and Sustainable Energy Reviews, 2015. **41**: p. 402-412.
15. Suhas, et al., *Cellulose: A review as natural, modified and activated carbon adsorbent*. Bioresource Technology, 2016. **216**: p. 1066-1076.
16. Nishiyama, Y., P. Langan, and H. Chanzy, *Crystal Structure and Hydrogen-Bonding System in Cellulose I β from Synchrotron X-ray and Neutron Fiber Diffraction*. Journal of the American Chemical Society, 2002. **124**(31): p. 9074-9082.
17. Sun, J.X., et al., *Isolation and characterization of cellulose from sugarcane bagasse*. Polymer Degradation and Stability, 2004. **84**(2): p. 331-339.
18. Fernandes, A.N., et al., *Nanostructure of cellulose microfibrils in spruce wood*. Proceedings of the National Academy of Sciences, 2011. **108**(47): p. E1195-E1203.
19. Peng, F., et al., *Chemicals from Hemicelluloses: A Review*, in *Sustainable Production of Fuels, Chemicals, and Fibers from Forest Biomass*. 2011, American Chemical Society. p. 219-259.

20. Hansen, N.M.L. and D. Plackett, *Sustainable Films and Coatings from Hemicelluloses: A Review*. Biomacromolecules, 2008. **9**(6): p. 1493-1505.
21. Silva, A.E.d., et al., *Xylan, a Promising Hemicellulose for Pharmaceutical Use*, in *Products and Applications of Biopolymers*, J. Verbeek, Editor. 2012, InTEch.
22. Gírio, F.M., et al., *Hemicelluloses for fuel ethanol: A review*. Bioresource Technology, 2010. **101**(13): p. 4775-4800.
23. Sella Kapu, N. and H.L. Trajano, *Review of hemicellulose hydrolysis in softwoods and bamboo*. Biofuels, Bioproducts and Biorefining, 2014. **8**(6): p. 857-870.
24. Delidovich, I., K. Leonhard, and R. Palkovits, *Cellulose and hemicellulose valorisation: an integrated challenge of catalysis and reaction engineering*. Energy & Environmental Science, 2014. **7**(9): p. 2803-2830.
25. Scheller, H.V. and P. Ulvskov, *Hemicellulose: Annual Review of Plant Biology*. Annual Reviews, 2010. **61**: p. 1-720.
26. Glasser, W.G., et al. *Studies of the molecular interaction between cellulose and lignin as a model for the hierarchical structure of wood*. in *American Chemical Society Symposium series 688; Orlando, FL; 1996 August 25-29; Washington, DC, American Chemical Society: 265-282*. 1998.
27. Horwath, W., *Chapter 12 - Carbon Cycling: The Dynamics and Formation of Organic Matter*, in *Soil Microbiology, Ecology and Biochemistry (Fourth Edition)*, E.A. Paul, Editor. 2015, Academic Press: Boston. p. 339-382.
28. Guadix-Montero, S. and M. Sankar, *Review on Catalytic Cleavage of C–C Inter-unit Linkages in Lignin Model Compounds: Towards Lignin Depolymerisation*. Topics in Catalysis, 2018. **61**(3): p. 183-198.
29. Upton, B.M. and A.M. Kasko, *Strategies for the Conversion of Lignin to High-Value Polymeric Materials: Review and Perspective*. Chemical Reviews, 2016. **116**(4): p. 2275-2306.
30. Phillips, M., *The Chemistry of Lignin*. Chemical Reviews, 1934. **14**(1): p. 103-170.
31. Bruijninx, P.C.A., R. Rinaldi, and B.M. Weckhuysen, *Unlocking the potential of a sleeping giant: lignins as sustainable raw materials for renewable fuels, chemicals and materials*. Green Chemistry, 2015. **17**(11): p. 4860-4861.
32. Sen, S., S. Patil, and D.S. Argyropoulos, *Thermal properties of lignin in copolymers, blends, and composites: a review*. Green Chemistry, 2015. **17**(11): p. 4862-4887.
33. Liu, W.-J., H. Jiang, and H.-Q. Yu, *Thermochemical conversion of lignin to functional materials: a review and future directions*. Green Chemistry, 2015. **17**(11): p. 4888-4907.
34. Wang, Y.-Y., Y.-r. Chen, and S. Sarkanen, *Path to plastics composed of ligninsulphonates (lignosulfonates)*. Green Chemistry, 2015. **17**(11): p. 5069-5078.
35. Yunpu, W., et al., *Review of microwave-assisted lignin conversion for renewable fuels and chemicals*. Journal of Analytical and Applied Pyrolysis, 2016. **119**: p. 104-113.
36. Rowell, R.M., *The chemistry of solid wood*. 1984.

37. Ebringerová, A. and T. Heinze, *Xylan and xylan derivatives – biopolymers with valuable properties, 1. Naturally occurring xylans structures, isolation procedures and properties*. *Macromolecular Rapid Communications*, 2000. **21**(9): p. 542-556.
38. Nations, F.a.A.O.o.t.U., *Forest Products Yearbook 2016*. 2018: Roma.
39. Zakzeski, J., et al., *The Catalytic Valorization of Lignin for the Production of Renewable Chemicals*. *Chemical Reviews*, 2010. **110**(6): p. 3552-3599.
40. Doherty, W.O.S., P. Mousavioun, and C.M. Fellows, *Value-adding to cellulosic ethanol: Lignin polymers*. *Industrial Crops and Products*, 2011. **33**(2): p. 259-276.
41. Rinaldi, R., et al., *Paving the Way for Lignin Valorisation: Recent Advances in Bioengineering, Biorefining and Catalysis*. *Angewandte Chemie International Edition*, 2016. **55**(29): p. 8164-8215.
42. Rinaldi, R., *A Tandem for Lignin-First Biorefinery*. *Joule*, 2017. **1**(3): p. 427-428.
43. Li, M.-F., S. Yang, and R.-C. Sun, *Recent advances in alcohol and organic acid fractionation of lignocellulosic biomass*. *Bioresource Technology*, 2016. **200**: p. 971-980.
44. Lali, A.M., J.S. Varavadekar, and P.C. Wadekar, *Process for fractionation of biomass 2011*.
45. Carneiro, A.P., O. Rodríguez, and E.A. Macedo, *Dissolution and fractionation of nut shells in ionic liquids*. *Bioresource Technology*, 2017. **227**: p. 188-196.
46. Yoo, C.G., *Pretreatment and fractionation of lignocellulosic biomass for production of biofuel and value-added products*. 2012, Iowa State University.
47. Mussatto, S.I. and G.M. Dragone, *Chapter 1 - Biomass Pretreatment, Biorefineries, and Potential Products for a Bioeconomy Development*, in *Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery*. 2016, Elsevier: Amsterdam. p. 1-22.
48. Romero, I., E. Ruiz, and E. Castro, *Chapter 10 - Pretreatment With Metal Salts A2 - Mussatto, Solange I*, in *Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery*. 2016, Elsevier: Amsterdam. p. 209-227.
49. Santos, J.C., et al., *Chapter 9 - Biomass Pretreatment With Oxalic Acid for Value-Added Products A2 - Mussatto, Solange I*, in *Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery*. 2016, Elsevier: Amsterdam. p. 187-208.
50. Shuddhodana, et al., *Chapter 23 - Enzymatic Hydrolysis of Lignocellulosic Residues A2 - Mussatto, Solange I*, in *Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery*. 2016, Elsevier: Amsterdam. p. 543-560.
51. Wikandari, R., R. Millati, and M.J. Taherzadeh, *Chapter 12 - Pretreatment of Lignocelluloses With Solvent N-Methylmorpholine N-oxide A2 - Mussatto, Solange I*, in *Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery*. 2016, Elsevier: Amsterdam. p. 255-280.
52. Xu, J.K. and R.C. Sun, *Chapter 19 - Recent Advances in Alkaline Pretreatment of Lignocellulosic Biomass A2 - Mussatto, Solange I*, in *Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery*. 2016, Elsevier: Amsterdam. p. 431-459.

53. Zhang, C., et al., *Chapter 22 - Sulfite Pretreatment to Overcome the Recalcitrance of Lignocelluloses for Bioconversion of Woody Biomass A2 - Mussatto, Solange I*, in *Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery*. 2016, Elsevier: Amsterdam. p. 499-541.
54. Polavarapu, P.L. and C.S. Ewig, *Ab Initio computed molecular structures and energies of the conformers of glucose*. *Journal of Computational Chemistry*, 1992. **13**(10): p. 1255-1261.
55. Kroon-Batenburg, L.M.J. and J. Kroon, *Solvent effect on the conformation of the hydroxymethyl group established by molecular dynamics simulations of methyl- β -D-glucoside in water*. *Biopolymers*, 1990. **29**(8-9): p. 1243-1248.
56. van Eijck, B.P., R.W.W. Hooft, and J. Kroon, *Molecular dynamics study of conformational and anomeric equilibria in aqueous D-glucose*. *The Journal of Physical Chemistry*, 1993. **97**(46): p. 12093-12099.
57. Heiner, A.P. and O. Teleman, *Interface between Monoclinic Crystalline Cellulose and Water: Breakdown of the Odd/Even Duplicity*. *Langmuir*, 1997. **13**(3): p. 511-518.
58. Heiner, A.P., L. Kuutti, and O. Teleman, *Comparison of the interface between water and four surfaces of native crystalline cellulose by molecular dynamics simulations*. *Carbohydrate Research*, 1998. **306**(1-2): p. 205-220.
59. Matthews, J.F., M.E. Himmel, and J.W. Brady, *Simulations of the Structure of Cellulose*, in *Computational Modeling in Lignocellulosic Biofuel Production*. 2010, American Chemical Society. p. 17-53.
60. Matthews, J.F., et al., *Comparison of Cellulose I β Simulations with Three Carbohydrate Force Fields*. *Journal of Chemical Theory and Computation*, 2012. **8**(2): p. 735-748.
61. Matthews, J.F., et al., *High-Temperature Behavior of Cellulose I*. *The Journal of Physical Chemistry B*, 2011. **115**(10): p. 2155-2166.
62. Hadden, J.A., A.D. French, and R.J. Woods, *Unraveling Cellulose Microfibrils: A Twisted Tale*. *Biopolymers*, 2013. **99**(10): p. 746-756.
63. Petridis, L., et al., *Hydration Control of the Mechanical and Dynamical Properties of Cellulose*. *Biomacromolecules*, 2014. **15**(11): p. 4152-4159.
64. Matthews, J.F., et al., *Computer simulation studies of microcrystalline cellulose I β* . *Carbohydrate Research*, 2006. **341**(1): p. 138-152.
65. Yui, T., et al., *Swelling behavior of the cellulose I β crystal models by molecular dynamics*. *Carbohydrate Research*, 2006. **341**(15): p. 2521-2530.
66. Youssefian, S. and N. Rahbar, *Molecular Origin of Strength and Stiffness in Bamboo Fibrils*. *Scientific Reports*, 2015. **5**: p. 11116.
67. Silveira, R.L., et al., *Plant Biomass Recalcitrance: Effect of Hemicellulose Composition on Nanoscale Forces that Control Cell Wall Strength*. *Journal of the American Chemical Society*, 2013. **135**(51): p. 19048-19051.

68. Berglund, J., et al., *A molecular dynamics study of the effect of glycosidic linkage type in the hemicellulose backbone on the molecular chain flexibility*. The Plant Journal, 2016. **88**(1): p. 56-70.
69. Sauter, J. and A. Grafmüller, *Solution Properties of Hemicellulose Polysaccharides with Four Common Carbohydrate Force Fields*. Journal of Chemical Theory and Computation, 2015. **11**(4): p. 1765-1774.
70. Durbeej, B. and L.A. Eriksson, *A Density Functional Theory Study of Coniferyl Alcohol Intermonomeric Cross Linkages in Lignin - Three-Dimensional Structures, Stabilities and the Thermodynamic Control Hypothesis*, in *Holzforschung*, 2003. p. 150.
71. Sawada, D., et al., *Structure and dynamics of a complex of cellulose with EDA: insights into the action of amines on cellulose*. Cellulose, 2013. **20**(4): p. 1563-1571.
72. Durbeej, B., Y.-N. Wang, and L.A. Eriksson, *Lignin Biosynthesis and Degradation — a Major Challenge for Computational Chemistry*, in *High Performance Computing for Computational Science — VECPAR 2002: 5th International Conference Porto, Portugal, June 26–28, 2002 Selected Papers and Invited Talks*, J.M.L.M. Palma, et al., Editors. 2003, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 137-165.
73. Besombes, S. and K. Mazeau, *Molecular dynamics simulations of a guaiacyl β -O-4 lignin model compound: Examination of intramolecular hydrogen bonding and conformational flexibility*. Biopolymers, 2004. **73**(3): p. 301-315.
74. Besombes, S. and K. Mazeau, *The cellulose/lignin assembly assessed by molecular modeling. Part 1: adsorption of a threo guaiacyl beta-O-4 dimer onto a Ibeta cellulose whisker*. Plant physiology and biochemistry : PPB, 2005. **43**(3): p. 299-308.
75. Besombes, S. and K. Mazeau, *The cellulose/lignin assembly assessed by molecular modeling. Part 2: seeking for evidence of organization of lignin molecules at the interface with cellulose*. Plant Physiology and Biochemistry, 2005. **43**(3): p. 277-286.
76. Kumar, R., et al., *Cellulose–hemicellulose interactions at elevated temperatures increase cellulose recalcitrance to biological conversion*. Green Chemistry, 2018. **20**(4): p. 921-934.
77. Yang, H., et al., *Quantum Calculations on Plant Cell Wall Component Interactions*. Interdisciplinary Sciences: Computational Life Sciences, 2018.
78. Petridis, L. and J.C. Smith, *A molecular mechanics force field for lignin*. Journal of Computational Chemistry, 2009. **30**(3): p. 457-467.
79. Petridis, L. and J.C. Smith, *Cellulosic ethanol: progress towards a simulation model of lignocellulosic biomass*. Journal of Physics: Conference Series, 2008. **125**(1): p. 012055.
80. Petridis, L., R. Schulz, and J.C. Smith, *Simulation Analysis of the Temperature Dependence of Lignin Structure and Dynamics*. Journal of the American Chemical Society, 2011. **133**(50): p. 20277-20287.
81. Hong, L., L. Petridis, and J.C. Smith, *Biomolecular Structure and Dynamics with Neutrons: The View from Simulation*. Israel Journal of Chemistry, 2014. **54**(8-9): p. 1264-1273.
82. Petridis, L. and J.C. Smith, *Conformations of Low-Molecular-Weight Lignin Polymers in Water*. ChemSusChem, 2016. **9**(3): p. 289-295.

83. Petridis, L., et al., *Self-similar multiscale structure of lignin revealed by neutron scattering and molecular dynamics simulation*. *Physical Review E*, 2011. **83**(6): p. 061911.
84. Vermaas, J.V., et al., *Mechanism of lignin inhibition of enzymatic biomass deconstruction*. *Biotechnology for Biofuels*, 2015. **8**(1): p. 217.
85. Langan, P., et al., *Common processes drive the thermochemical pretreatment of lignocellulosic biomass*. *Green Chemistry*, 2014. **16**(1): p. 63-68.
86. Sangha, A.K., et al., *Molecular simulation as a tool for studying lignin*. *Environmental Progress & Sustainable Energy*, 2012. **31**(1): p. 47-54.
87. Schulz, R., et al., *Scaling of Multimillion-Atom Biological Molecular Dynamics Simulation on a Petascale Supercomputer*. *Journal of Chemical Theory and Computation*, 2009. **5**(10): p. 2798-2808.
88. Lindner, B., et al., *Solvent-Driven Preferential Association of Lignin with Regions of Crystalline Cellulose in Molecular Dynamics Simulation*. *Biomacromolecules*, 2013. **14**(10): p. 3390-3398.
89. Pingali, S.V., et al., *Morphological changes in the cellulose and lignin components of biomass occur at different stages during steam pretreatment*. *Cellulose*, 2014. **21**(2): p. 873-878.
90. Chang, S., et al., *Solvent polarity and internal stresses control the swelling behaviour of green wood during dehydration in organic solution*. *Bioresources*, 2012. **7**(2).
91. Stamm, A.J., *SHRINKING and SWELLING of WOOD*. *Industrial & Engineering Chemistry*, 1935. **27**(4): p. 401-406.
92. Nayer, A.N. and R.L. Hossfeld, *Hydrogen Bonding and the Swelling of Wood in Various Organic Liquids*. *Journal of the American Chemical Society*, 1949. **71**(8): p. 2852-2855.
93. Finch, C.A., *The chemistry of solid wood*. Edited by R. M. Rowell, *ACS Advances in Chemistry Series No. 207, American Chemical Society, Washington D.C., 1984. pp. x + 614, price \$79.95 (USA & Canada), \$95.95 (rest of world). ISBN 0-8412-0796-8*. *British Polymer Journal*, 1986. **18**(5): p. 352-352.
94. Mantanis George, I., A. Young Raymond, and M. Rowell Roger, *Swelling of Wood. Part II. Swelling in Organic Liquids*, in *Holzforschung - International Journal of the Biology, Chemistry, Physics and Technology of Wood*. 1994. p. 480.
95. Gordy, W. and S.C. Stanford, *Spectroscopic Evidence for Hydrogen Bonds: Comparison of Proton-Attracting Properties of Liquids. II*. *The Journal of Chemical Physics*, 1940. **8**(2): p. 170-177.
96. Hasselblatt, M., *Der Wasserdampfdruck und die elektrische Leitfähigkeit des Holzes in Abhängigkeit von seinem Wassergehalt*. *Zeitschrift für anorganische und allgemeine Chemie*, 1926. **154**(1): p. 375-385.
97. Mantanis, G.I., R.A. Young, and R.M. Rowell, *Swelling of wood*. *Wood Science and Technology*, 1994. **28**(2): p. 119-134.
98. MEIER, P., et al., *Multiple Swelling of Pinewood (Pinus Sylvestris) in Binary and Ternary Mixtures of Ethanol, Acetone and Water* *MATERIALS SCIENCE (MEDŽIAGOTYRA)*, 2006. **12**(1).

99. Meiera, P., et al., *Mechanical properties of pinewood (Pinus Sylvestris) swollen in organic liquids*
Proc. Estonian Acad. Sci. Eng., 2006. **12**(2): p. 125–133

Chapter 1. State of the art

Contents

Abstract	38
1.1 Wood structure and constituents	38
1.1.1 Cellulose	40
1.1.2 Hemicellulose	41
1.1.3 Lignin	42
1.2 Interactions of solvents with lignocellulose biomass	43
1.2.1 Pretreatment and fractionation processes	43
1.2.2 Swelling: first step towards polymer dissolution.....	44
1.2 Computational studies of lignocellulose.....	45
1.3.1 Cellulose	46
1.3.2 Hemicellulose	52
1.3.3 Lignin	53
1.3.4 Lignocellulose interactions.....	54
References	55

Abstract

In this chapter, we reviewed the lignocellulose structure and the chemical composition of its main constituents, the most common pre-treatment and fractionation methods including mixed solvents, as well as the study of swelling to understand the factors involved in lignocellulose solubilization. An emphasis is given on the computational studies dealing with lignocellulosic biomass. Lignocellulose is composed mainly of cellulose, hemicellulose, and lignin. These polymers form a complex matrix; thus, the access to hemicelluloses and cellulose is hindered by the complex chemical and physical structure of biomass. The conversion of cellulose into fermentable sugars in lignocellulose pretreatment is enhanced in the presence of organic solvents. Computational methods along with experimental studies have provided the first insights recently into the mechanisms of lignocellulose deconstructions, from dehydration, solubilization and reduction of the interactions between lignocellulose components, to selective interactions of the solvents with different lignocellulose domains. These studies also evidenced many questions about the mechanisms for biomass deconstruction that remain to be addressed.

1.1 Wood structure and constituents

Wood-i.e., lignocellulosic biomass- is a natural, composite material constructed from a variety of organic polymers. The skeletal material of wood cell walls is cellulose, a long chain, linear sugar polymer composed of glucose monomers, and it accounts for about 40-45% of the dry weight of normal wood tissue. Cellulose is organised into fibrils[1] (Fig. 1.1) and interacts with a surrounding matrix formed by hemicellulose and lignin[2, 3]. These biopolymers form a complex matrix, intricately cross-linked with a network of interactions. Hemicellulose and lignin are covalently and non-covalently linked, while cellulose and hemicelluloses are connected exclusively by non-covalent interactions. Moreover, the polysaccharides of the cell wall, cellulose and hemicellulose, have a strong affinity for water molecules in either their liquid or vapour state. Lignin, on the other hand, is considered mostly hydrophobic because of its aromatic nature.

The skeletal substance of the wood cell wall –cellulose- is aggregated to form microfibrils and macrofibrils, which further combine to form sheets of wall layers. This architectural scheme, illustrated in Figure 1.1, is widely considered to be formed by the middle lamella (ML), the primary cell wall (P), and the S1, S2, and S3 layers of the secondary cell wall (S) [4]. The relative composition of cellulose, hemicellulose, and lignin among the different cell wall layer is shown in Figure 1.2. The primary cell wall is composed of cellulose microfibrils that form a random, irregular, and interlinked network to facilitate cell expansion. The primary wall also contains a large portion of carbohydrates, particularly pectic materials and hemicelluloses. The region between two adjacent primary walls is generally known as the middle lamella. This region is 70-80% lignin by weight, and it is considered as responsible for binding all wood cells together. Nonetheless, about 70% or more of the total wood lignin is located in the cell wall itself. Finally, in the innermost region of the

primary wall, the secondary wall is formed by many lamellae, each with a specific orientation. Each of these S layers are easily distinguishable and have different thickness and orientations. Moreover, this region is rich in cellulose.

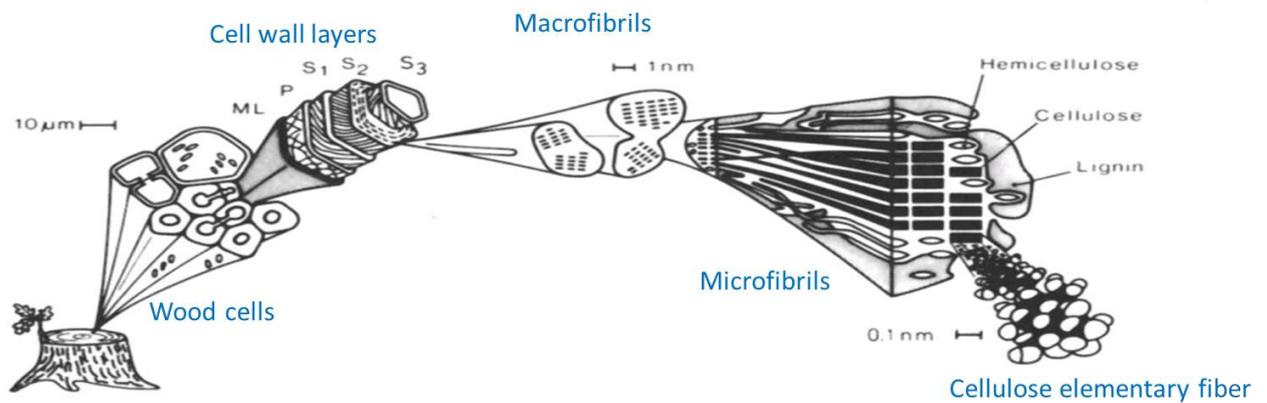


Figure 1. 1. Structure of wood cell walls [5].

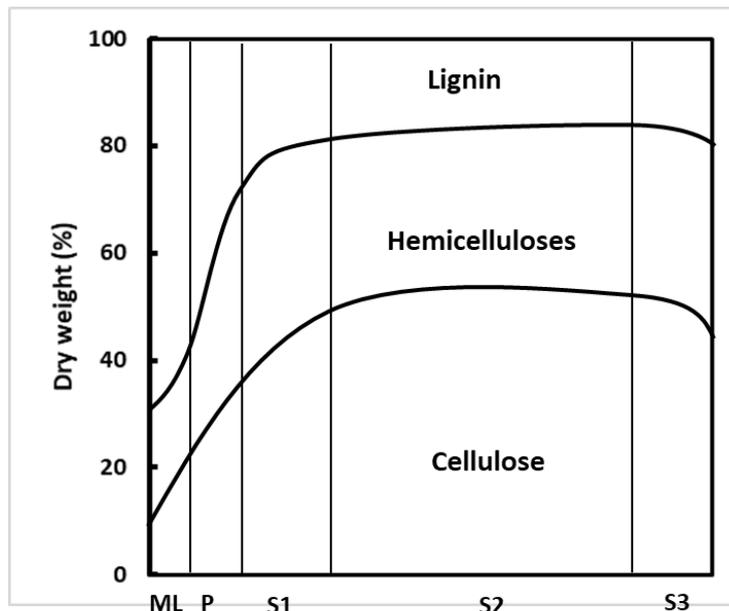


Figure 1. 2. Composition of wood constituents through cell wall[6]

The gain or loss of water or other liquids into or out of the wood cell wall can be significantly influenced by the nature, amount, and distribution of wood polysaccharides. Thus, swelling and shrinkage of wood subject to different environments are affected by sorption phenomena and architectural arrangements of wood cells. Furthermore, the cell wall is anisotropic, resulting in differences in swelling and shrinkage in different directions and different part in the cell wall [4]. Moreover, the cell structure varies among different kinds of tissue, that can be influenced by the part of the wood, time, weather, wood type, among others. In the following section, we describe the main features of the physical structure of the main constituents of wood – i.e., cellulose, hemicellulose, and lignin-.

1.1.1 Cellulose

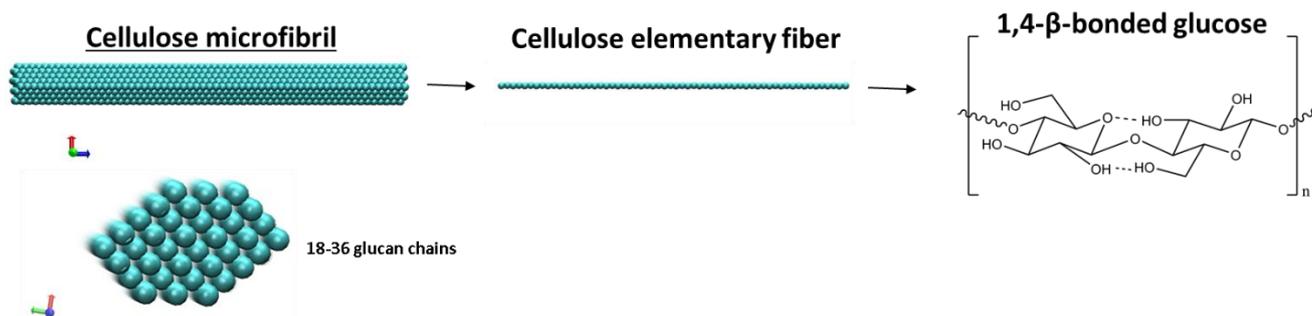


Figure 1. 3. The partial molecular structure of cellulose in the 1,4- β -D-glucopyranose form. Cellulose assembles into elementary fibres and cellulose microfibrils that vary from 18 to 36 glucan chains[1, 7].

Cellulose is a glucan polymer consisting of linear chains of 1,4- β -bonded anhydroglucose units. The number of sugar units in one molecular chain is referred to as the degree of polymerisation (DP). The wood cellulose has an average DP of at least 9000-10000 and possibly as high as 15000. A DP of 10000 means a linear chain length of approximately 5 μm in wood. Cellulose in the primary wall has a lower average DP and is thought to be polydisperse. Native cellulose is partially crystalline[8].

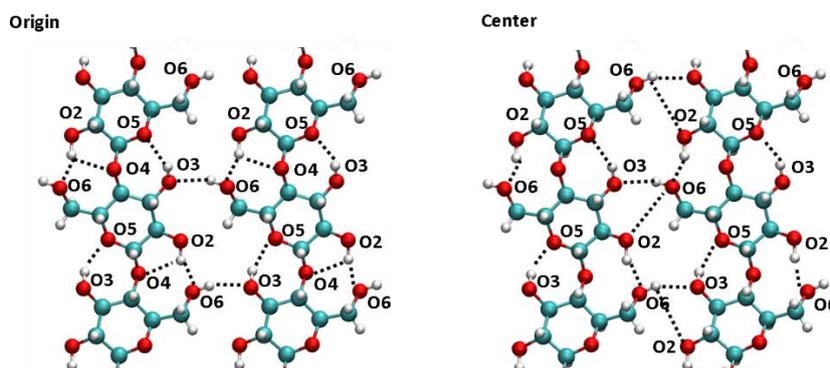


Figure 1. 4. Schematic representation of the hydrogen bonds in origin (left) and centre (right) sheets of cellulose I β . Carbon, oxygen, and hydrogen atoms are coloured cyan, red, white, respectively. Hydrogen bonds are represented by dotted lines [1].

Natural crystalline cellulose consists of two allomorphs, I α and I β , but for this review, we focus on allomorph I β , which dominates in higher organisms such as plants and therefore it is the industrially most relevant form [9]. The synchrotron X-ray and neutron experiments of Nishiyama et al. [1] have determined that the monoclinic unit cell of cellulose I β (space group symmetry P21) consists of two parallel chains having slightly different conformations (centre and origin) and they organised in sheets (Fig. 1.4). The cell dimensions are $a= 7.784$, $b= 8.201$ nm, $c=10.380$ nm and $\gamma=96.5^\circ$.

The Intramolecular hydrogen bonds O3-H...O5' and O2...O6' (more dynamic), parallel to the glycoside bond cause each cellulose to have extended rigid chains (Fig. 1.4). The hydroxymethyl groups of cellulose I are in the tg conformation, with O5-O6 trans [10]. Each hydroxymethyl (O6) in a cellulose chain is partly engaged

in an intramolecular hydrogen bond with the O2 atom of the adjacent glucosyl moiety of the same chain and partly in an intra-chain hydrogen bond with another chain of the same sheet. Moreover, there is no inter-sheet O-H...O hydrogen bonds in cellulose I β . Therefore, the cellulose sheets are held together only by hydrophobic interactions and weak C-H...O bonds. [1]

1.1.2 Hemicellulose

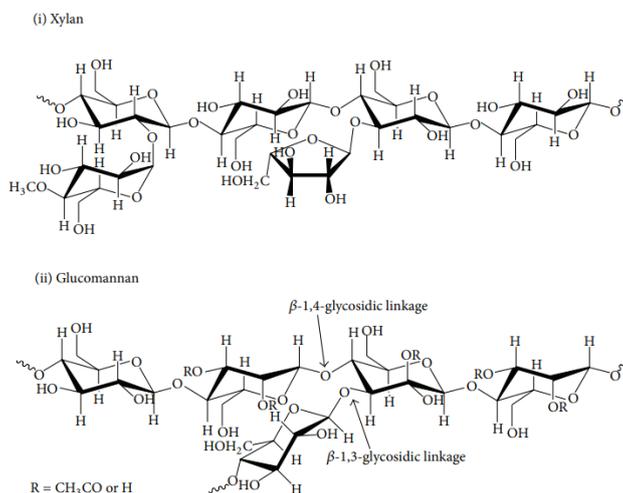


Figure 1. 5. Chemical structure of xylan and glucomannan hemicellulose compounds[11].

Hemicellulose (Fig. 1.5) is a low molecular weight polysaccharide that contains short chains of carbohydrates and serves as a matrix substance for the cellulose superstructure. The carbohydrates in hemicellulose represent mostly combinations of various five-carbon sugars (xylose and arabinose) and six-carbon sugars (glucose, mannose, and galactose) [4]. Common hemicelluloses include xylan, glucuronoxylan, arabinoxylan, glucomannan, and xyloglucan. Moreover, hemicellulose composition differs in different wood species. In this work, hemicellulose is regarded as xylan, the most abundant hemicellulose in plants [12]. The backbone of xylan consists of poly β -(1 \rightarrow 4) linked D-xylopyranosyl units, to which can be attached a variety of side group.

1.1.3 Lignin

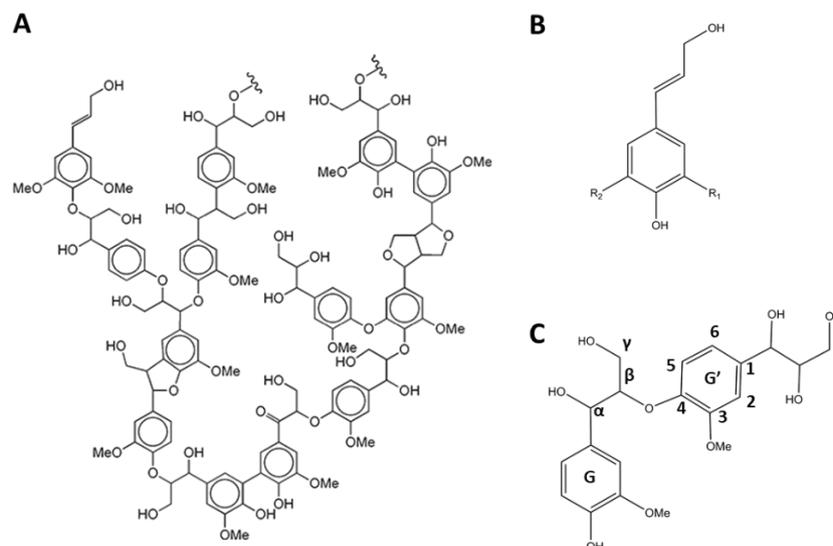


Figure 1. 6. A) Chemical structure of a lignin polymer with different linkages and monomer units [13]. B) The three monolignols before polymerisation: *p*-coumaryl ($R_1=R_2=H$), coniferyl ($R_1=H$, $R_2=OMe$), and sinapyl ($R_1=R_2=OMe$) alcohols. C) A lignin dimer formed by two guaiacyl (G, G') units linked with a β -O-4 ether bond.

Lignin, the third principal constituent of wood after hemicelluloses, composes around 25-35% of the total dry weight of the polysaccharide wall substances [4]. It permeates both cell walls and intercellular regions and plays a significant role in imparting rigidity to the polysaccharide wall substances [4]. Lignin interactions with cellulose determine wood mechanical properties, such as strength, creep, durability and ageing [14]. Lignin is a three dimensional, highly branched, and polyphenolic complex structure with high molecular weight and a variety of lignin units and linkages (Fig. 1.6A). The phenylpropanoid unit (Fig. 1.6B) is the basic structure of lignin, consisting of an aromatic ring and a 3-C side chain [15]. The monolignol subunits are derived from coniferyl, *p*-coumaryl, and sinapyl alcohols (Fig. 1.6B), and are referred to as, after polymerisation, *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) residues, respectively [13]. Two types of inter-unit linkages (C-O and C-C linkages) link several aromatic units, being the β -O-4 ether bond the predominant inter-unit linkage in the structure of lignin (Fig. 1.6C) [16].

The chemical structure of lignin and composition of lignin varies with source, type of lignin, and isolation method. Thus, the study of lignin is a considerable challenge due to the lack of a regular and ordered structure. The β -O-4 linkage represents the predominant inter-unit linkage in lignin, and it is an excellent model for studying major conformation features such as H-bonding and flexibility [17]. Thus, concerning the complexity of the lignin network, we have chosen here to explore the guaiacyl β -O-4 dimer model compound. Furthermore, the study of such model compounds can lead to the understanding of the behaviour of larger and more complex molecular systems.

1.2 Interactions of solvents with lignocellulose biomass

Lignocellulose biomass can be used in a plethora of applications [4, 18-40]. However, access to hemicelluloses and cellulose is hindered by the complex chemical and physical structure of biomass. Thus, biomass pretreatment is fundamental for modifying its structural and chemical characteristics. Nonetheless, this step is one of the main economic costs in the process. To make end processes feasible, the fractionation step should produce easily treated fractions, such as cellulose pulp, soluble lignin, and hemicellulose-derived products (oligomers, sugars, furans, organic acids, among others) [41]. Thus, the reduction of the energetic costs of the fractionation processes passes through the optimisation of the interactions between the different components of the biomass and a solvent medium[42].

1.2.1 Pretreatment and fractionation processes

The use of solvents has provided remarkable results in fractionation processes of several levels of severity, ranging from swelling, dilute acid, hydrothermal, steam explosion, alkaline treatment, to organosolv pulping[41-52]. Alkaline reagents such as ammonium hydroxide, sodium hydroxide, and calcium hydroxide has been used as pre-treatment reagents[53]. This treatment increases cellulose digestibility by affecting lignin in biomass. However, the degradation of hemicellulose and cellulose is minor compared to acid and hydrothermal treatments and has been more effective at treating hardwood than softwood. Concentrated or diluted acid has been used to solubilise hemicelluloses. However, the use of concentrated acid is not suitable for ethanol production due to the degradation of both hemicellulose and cellulose, and the formation of inhibiting compounds. [54, 55]. Biomass can also be pre-treated with liquid hot water at high pressure to solubilise hemicellulose and degrade lignin. However, the high demand for energy and water makes this process expensive[56]. The steam explosion pretreatment combines both chemical and mechanical effects, and the energy requirements are lower than for the liquid hot water process.

Moreover, organic solvents are also used to enhance the conversion of cellulose into fermentable sugars in lignocellulose pretreatment [57-59], and novel water-organosolv pretreatments have demonstrated enhanced sugar yields[60-62]. It is presumed that water-organosolv pretreatment disrupt the lignocellulose interactions, resulting in structural changes. Tetrahydrofuran-water (THF-water) and γ -valerolactone-water (GVL-water) systems have demonstrated to increase to solubilise lignin and hemicellulose as well as partially de-crystallise cellulose[61, 63, 64]. Ammonia and ethylenediamine (EDA) can efficiently penetrate cellulose fibres, and they are being explored for use in pulping and lignocellulose pre-treatment to improve its conversion [65-67]. Swatloski et al.[68]reported that ionic liquids could be used as solvents cellulose under relatively mild conditions. Good solubilities of cellulose in ILs as high as 25 wt% have been observed; however, the major drawback is the high viscosity of the ionic liquids, which hinders cellulose dissolution and post-processes. The inclusion of aprotic solvents as cosolvents to decrease the viscosity without the precipitation of cellulose has

been proposed to overcome this problem. [69-72] Polar aprotic solvents can accept a hydrogen bond through lone pair electron but are unable to donate any acidic protons and form hydrogen bonding. Biomass conversion in various polar aprotic solvents is promising for the solubilization and depolymerisation of carbohydrates[73].

Thus, there is considerable interest in understanding how cosolvents can lead to an increase in pre-treatment efficiency. Moreover, non-linear effects of aqueous solutions of organics have been observed in solvent fractionation techniques, where aqueous-organic solvent solutions led to higher lignin solubilisation and extraction of polyphenols than pure solvents [74, 75]. However, the factors responsible for such behaviour are still not well understood.

1.2.2 Swelling: first step towards polymer dissolution

Swelling of dry and green wood in liquid water and organic solvents has been studied by several authors[76-84]. Early, it has been established an inverse relation between the weight and size of the molecule with the extent of wood swelling due to diffusion barriers [76]. Nayer and Hossfeld [77] tried to correlate the volumetric swelling of thin sections against the dielectric constant, dipole moment and surface tension of liquids. However, weak correlations were observed. Nevertheless, they were able to find a relationship between the degree of swelling and the extent of hydrogen bonding between constituents of wood and the swelling agent. Similarly, Mantanis [85] evaluated the swelling rates of forty organic liquids, and he concluded that the swelling of wood in organic solvents is mainly influenced by three solvent properties: the solvent basicity, the molar volume, and the hydrogen bonding capability [86]. Some other authors have performed experiments to correlate the effect of temperature [80].

Higher swelling values for wood when using non-pure organic solvents were suggested earlier by Hasseblatt [79]. Ishimaru and coworkers [81, 85] studied water-ethanol and water-acetone mixtures and concluded that the increase of swelling of wood was explained thanks to a positive deviation of the constituent's activity from the ideal solution meaning some stresses related with the cohesion force of the mixture were released and, thus, absorption sites in the polymer seem to increase. More recent work on wood swelling of mixed solvents in dry and green wood [87-91] (Fig. 1.7) has shown not only higher swelling values for selected mixed aqueous solutions, but also a nonlinear relation between the volume swollen and the concentration of the organic solvent. This evolution of wood volume with the concentration of the organic solvents shows a behaviour similar to gels, with a significant swelling for solutions of intermediate polarity [91]. Other authors have explained this behaviour due to changes in the wood structure thanks to a partial dissolution of lignin from the matrix [89, 90].

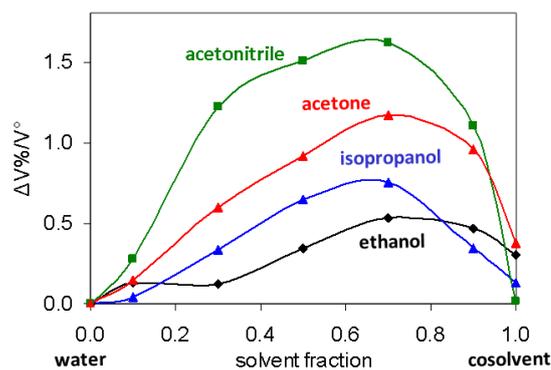


Figure 1. 7 Volumetric swelling of green wood at different concentrations of mixed solvents, taken from ref. [91]

Recent experimental work based on optical microscopy experiments has evidenced a detachment between the primary wall and middle lamella of wood cells in pure and ethanol aqueous solutions [92]. There is a consensus in the literature that changes in the morphology of lignocellulosic biomass as exposed to mixed solvent plays a vital role in the increase of its solubility, promoting the fractionation of the material. However, the current understanding of the solvent effect on the interactions within the cell wall is limited. Hence, molecular-scale studies achieved by experimental and theoretical approaches are needed for the understanding and optimisation of biorefinery processes [93-97].

1.2 Computational studies of lignocellulose

Understanding how pre-treatment will affect biomass chemical and physical features can offer meaningful improvements in biomass conversion [98]. Furthermore, a molecular-level description is needed to understand the interplay between the solvents and the lignocellulose structures, which can be useful for the design of solvent systems for use in biomass pretreatment processes. In this regard, computational methods such as molecular dynamics (MD) and quantum chemistry (QM), among others, are an indispensable investigation tool to provide insight into the structural and dynamic properties of lignocellulose systems at the atomic level.

Molecular-scale studies achieved by computational methods often in combination with experimental investigations have been promising to understand the conformations of the principal wood components –i.e. cellulose[9, 99-110], hemicelluloses [111-115], and lignin [17, 116-124]-, but the solvent effect on lignocellulosic biomass remains widely unexplored [10, 64, 92, 125-130]. Although the study of lignin and hemicellulose has proven to be a considerable challenge due to the lack of a regular and ordered structure, theoretical studies have provided valuable information about the most common units and linkages present in these polymer structures [17, 111-124, 131, 132]. In this section, we summarise the findings of the theoretical studies performed with MD and QM methods. We also include some of the models used in these studies to highlight the evolution of the field over the past years.

1.3.1 Cellulose

Molecular dynamics (MD) simulations along with experimental studies have provided valuable information about the cellulose bulk structure [99, 100], the orientation of the hydroxymethyl group, and the hydrophilic/hydrophobic cellulose surface site interactions with water [9, 101-110], and the degree of water structuring due to hydrogen bonding with the cellulose surfaces [108, 109].

1.3.1.1 First studies on cellulose monomers

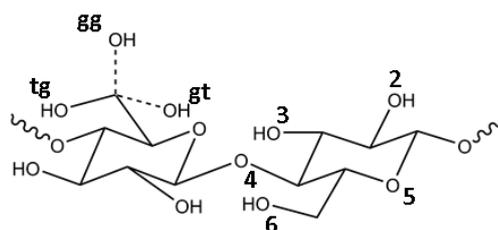


Figure 1.8. (A) Cellulose topology and oxygen notation. Three hydroxymethyl group conformers tg (bulk, full line), gt, and gg (dashed lines), and their gauche or trans positions of O6 relative to both O5 and C4 are present in cellulose. Ideal tg, gt, and gg conformations are characterised by the hydroxymethyl torsion angle $\omega = 180^\circ$, 60° , and -60° to O5, respectively. In the bulk phase, the hydroxymethyl is found in the tg conformer, and participates in hydrogen bonds with the O2 of the adjacent residue (See Fig. 1.4) [1].

Early work concerned mostly the evaluation of the stability of glucose anomers, the hydroxymethyl group conformation, and the solvent effect on glucose oligomers. van Eijck et al.[101] calculated free energy differences of six conformations of D-glucose α/β and the transitions between the three staggered conformations of the hydroxymethyl group (gg/gt/tg) by MD using the GROMOS force field. By NMR studies, it is known that in both α and β forms, the gg and gt forms dominate. Experimentally, $G\beta-G\alpha$ is -0.34 kcal mol $^{-1}$, and the study provided a calculated value of -0.86 kcal mol $^{-1}$ in good agreement with the experiment. The free energy differences between gt and gg forms are almost zero, in excellent agreement with NMR measurements. However, the tg form was overestimated. This study highlights the limitations of the model in terms of the force field and cutoff radius and the methods used.

Ha et al [133] and Schmidt et al.[134] performed free energy calculations of D-glucose and D-xylose in aqueous solution to analyse the anomeric equilibrium when solvated. In both papers, the simulations predicted the wrong anomers to be preferred in solution, and this was attributed to the used force field [135]. Thermodynamic integration analysis of the component values showed that the free energy difference results from a balance between an internal term favouring the α anomer and a solvation term favoring the β anomer. Nevertheless, hydrogen bonding analysis of the MD provided a mechanistic explanation of the solvation preference for the β anomer. Improved hydrogen bonding of the anomeric hydroxyl group results as a consequence of an increased accessible surface area and determined by geometric features of the solute and

solvent structuring. Thus, this work highlighted the importance of including the solvent medium in the simulation, to get the right conformational features of the system.

Both QM [99] and MD [100] have shown that solvation effects determine the orientation of the hydroxymethyl group. Similarly, Lee et al. [10] observed by H-NMR and MD that the intramolecular hydrogen bond O3-H3...O5 over the glycosidic linkage in methyl β -cellobioside is present in apolar solvents but absent in polar solvents. Due to solvation/desolvation of the hydroxyl groups, solute-solvent hydrogen bonds are often energetically favoured over solute-solute hydrogen bonds, resulting in different hydroxyl conformations for well-hydrated carbohydrates compared to carbohydrates in vacuum or an apolar environment. Moreover, Their results indicate that intramolecular hydrogen bonding does not contribute significantly to the rigidity of cellulose chains during the spinning of fibres in aqueous solutions.

1.3.1.2 Cellulose in aqueous medium

The results highlighted in the previous section provided a base for the understanding of the solvation effect on the hydroxymethyl conformation and intramolecular hydrogen bonding of cellulose. However, the results are presented for minimal models that do not capture the complexity of the hydrophobic/hydrophilic nature of cellulose.

Two decades ago (1997), the first paper on the interface between a crystal of cellulose I β and water was published by Heiner and Teleman [102]. Favoured At the surface, half of the cellulose intermolecular hydrogen bonding is lost, but this it is compensated by hydrogen bonds with water molecules. According to this study, the energy of water molecules in the first hydration layer is lower by 2 kJ/mol, and the cellulose surface contains in average five exposed hydroxyl groups per square nanometer, which explains the excellent hydration of this surface. In this study, six cellulose layers, each layer consisting of six chains of three cellobiose units were placed in a monoclinic periodic box. The simulation was performed during 500 ps under the NVT ensemble, using GROMOS87 force field and the GROMOS87 program. The interplanar distance was monitored, and it was found to decrease by 3%. Besides, for “odd” and “even” chains, the camber angle between the glucose ring plane and the 200 plane was monitored, and between layer II and III, it was found the camber differs by 8.3°. For layer I odd molecules, the camber increases to 7.5° for C6-outward glucose rings, which is believed to be a consequence of the asymmetry of the hydrogen bond network of cellulose chains at the surface.

In further work, Heiner et al.[9] assessed the hydrophilicity of different surfaces of native crystalline cellulose. Upon association, hydroxyl groups of either substrate or ligand may reorient, and the reorientations are thought to be due to solvation/desolvation processes, and therefore entropy-driven. Based on the previous studies, it is expected that the hydrophilicity of the surface, by means of the extent of hydrogen bonding, depends on the precise distribution of the hydroxyl groups and their ability to adjust to the solvent to

maximise hydrogen bonding. This work focused on the qualitative description of the surface structure and solvent-induced structural changes on the four surfaces of native crystalline cellulose (monoclinic 110 and 1-10 and triclinic 100 and 010) exposed to water. It was found that for all cases, only the topmost surface layer of cellulose is structurally affected by the water of the exterior. The monoclinic 1-10 and triclinic 100 surfaces were found to be denser and more hydrophilic than the monoclinic 110 and 010 surfaces since the former made more hydrogen bonds with the solvent.

Mazeau and Heux [110] found that the conformational behaviour of the amorphous phase of cellulose differs significantly from that of the crystal phases. For instance, torsion angles explore an extensive angular range, and hydroxymethyl groups explore mostly the *gg* and *gt* orientations, in contrast with the crystal structure. Thus, the total number of hydrogen bonds per residue is more significant for the crystalline microstructure than for the amorphous one.

In the work of Matthews et al. [108], different extended surfaces (1 1 0, 1 $\bar{1}$ 0, 1 0 0, and 010) of two small crystals of 36 and 32 chains of 14 monomer units of cellulose I β surrounded by water were analysed. The simulations were carried out under the NVE ensemble during 1 ns using the CHARMM program along with the CHARMM Carbohydrate Solution Force Field (CSFF) [136] and the modified TIP3P water model. As water molecules interact with individual functional groups in the cellulose surface, they will become localised at particular positions relative to the surface. Contour maps to show the regions with high and low water densities. These simulations demonstrated that cellulose surfaces highly structure the water solution in contact with them primarily through direct hydrogen bonding.

Moreover, during the simulation of Matthews et al. [108], the crystal structures showed deviations in *rmsd* ranging from 1.46 to 1.72 Å in comparison with the initial structure, differing from the unit cell dimensions in previous diffraction studies. The unit cell varied such that *a* increased from 7.784 to 8.470 Å, *c* increased from 10.380 to 10.512, and the γ -angle decreased from 96.5° to almost 90°. Also, many of the C-6 primary alcohols underwent rotational transitions from *tg* to *gg* conformations. In contrast with the alcohol groups in the surface, whose structure is influenced by the solvent medium, conformations in the internal layers of the crystal should not be affected. Besides, it was observed an increased disorder and swelling of the glucose units at each end of the chains. The transitions to the *gg* conformation in the simulations seem to be expected since the CSFF force field was parameterized to reproduce the lower energy conformation of isolated glucose residues, instead of cellobiose units with glycoside bonds. Nevertheless, in this simulation both microcrystals developed a right-hand twist of about 1.5° per cellobiose unit, similar to the twist of β -sheets in proteins, and the authors indicate that such twist is not related with the transition to *gg* conformation of the primary alcohol of the centre chains, since constrained simulations of these groups in the *tg* conformation were also performed, showing the same twist.

Yui et al. [109] performed a similar study also using atomic coordinates, hydroxyl-group orientations, and lattice parameters from the fibre-diffraction results of Nishiyama et al.[1] using the GLYCAM04 force field and AMBER 8 package. The size of the crystals varied from 24 to 48 chains with DP of 10 or 20 monomer units. Simulations using the GLYCAM [109, 137] and GROMOS [9] force field revealed the similar right-handed twist of the chains, arising the question of whether the differences with the experiment are due to force field parameterization or real transitions in the material. The authors suggested that a possible cause for such a transformation would be an inherent tension probably involved in both the origin- and centre-chain sheets.

In the following year after this discussion regarding the twist was aroused, Guvench et al. [138] presented the all-atom additive force field for hexopyranose monosaccharides. Parametrisation target included vibrational frequencies, crystal geometries, solute-water interaction energies, molecular volumes, heats of vaporisation, and conformational energies for over 1800 monosaccharide conformations at the MP2/cc-pVTZ/MP2/6-31G(d) level of theory. Free energies of aqueous solvation for model compounds compared favourably with experimental values. Furthermore, shortly after Guvench et al. [139] presented an extension of the CHARMM hexopyranose monosaccharide force field to enable modelling of glycosidic-linked hexopyranose polysaccharides. This extension allowed for MD simulations of linear, branched, and cyclic hexopyranose glycosides both alone and in heterogeneous systems including proteins, nucleic acids, and lipids. Similarly, the GLYCAM06 [140] and GROMOS56A(CARBO)[141] force fields were released after optimisation to balance the interaction energies due to inconsistency between classes of molecules and to reproduce the correct properties of these systems. Also, parameters for carbohydrate derivatives such as xylose, fucose, glucuronic acid [142], among others, and solvent-free coarse-grain models for crystalline and amorphous cellulose fibrils was developed [143].

Matthews et al. [103] performed a conformational analysis of cellobiose and simulation of hydrated cellulose I β with CSFF and GLYCAM06 force fields using the CHARMM program. 3-ns simulations of hydrated DP 40 diagonal crystals with either 36 or 16 chains were run with both force fields to explore the effect of chain length and diameter of fiber structure. Also, additional 10-ns simulations of the diagonal crystal were run. In contrast with previous simulations of the same group with CSFF, for this simulation, the treatment of electrostatics was handled by particle mesh Ewald whereas the earlier simulations ignored interactions more than 13 Å away. The GLYCAM06 simulations of this study reproduced the hydroxymethyl conformation and hydrogen bond pattern from diffraction studies, but unit cell dimensions do not match exactly. In 36-chain fibrils of CSFF and GLYCAM06 simulations, the interior of the GLYCAM06 crystal remains stable, with surface chains more free to move whereas CSFF shows more dynamic behaviour in the interior, with occasional transitions from gt back to tg in the origin layers. Moreover, for the 10-ns MD of the I β diagonal with CSFF, the twist mentioned in previous work went away as the simulation progressed, as direct result of the development of the regular three-dimensional hydrogen bond pattern. However, this twist is observed in a higher magnitude for the GLYCAM06 force field due the fact that in this model the hydrogen bond patter is

mostly two dimensional. Many essential features found in these simulations are sensitive to differences in force field parameters and treatment of long-range interactions, giving rise to dramatically different structures.

Further work by the same research group [104] compared fully hydrated 36 chain microfibril with DP 40 for the CHARMM35, GLYCAM06, and Gromos45a4 up to the near-microsecond scale. For all simulations, the box sizes are 6 x 6 x 23.3 nm and contains 1.5 nm of water solvation in each dimension. It was found the behaviour of cellulose with the GROMOS 45a4 force field to significantly diverge from the initial cellulose I β crystal structure and with the other two force fields. For GLYCAM and C35 force fields, it was found that the microfibrils initially twisted while retaining the tg conformation in the interior. After several tens to hundreds of ns, the microfibrils untwist. The loss of twist is due to the conformation of a three-dimensional hydrogen-bond pattern, which is not present in the cellulose I β crystal structure. These events seem to be promoted by surface-mediated conformational changes, which produce both gt and gg conformers and an expanded unit cell. It was demonstrated that for GLYCAM, both the initial hydrogen bond pattern and 1,4 electrostatic scaling factors increase the rate of structural divergence from the I β structure, while for C35 this rate is not significantly affected. Also, surface chain conformations for both force fields equilibrate within the first few tens of nanoseconds, and the twist disappears after hundreds of nanoseconds. From these comparison simulations, the time scale achieved is likely sufficient to converge interactions of interest which occur at surfaces, such as solvent-cellulose or catalyst-cellulose interactions. Also, it was noted that simulations with all three force fields diverge from the initial I β structure. It is unknown from the experiment whether particles of this width and length may have the same structure as in the bulk phase. Experimental work addressing these questions is needed.

It was discussed in simulations of finite length crystals with explicit chain ends microfibrils can adopt twist along a chain axis, as a result of chirality amplification [105]. This can be solved by using an alternative simulation approach in which periodic boundary conditions are applied at the chain-ends, by simulating a periodic infinite crystal where chains are covalently linked across the periodic boundary such that there are no chain ends [104]. This behaviour has also been observed under some experimental conditions, for instance, with individual microfibrils with a cross-sectional thickness of 20-50 nm. However, due to insufficient conformational sampling in MD simulations and experimental results, it remains unclear the certainty of this behaviour.

Hadden et al. [106] studied a cellulose crystal to understand the driving forces behind this twisting, as well as to determine the role of computational methodology, microfibril dimensions, solvent effect and solvent model, charge set and bond constrains. Their findings suggest that vdW interactions favour microfibril twisting, and are counteracted by both intrachain hydrogen bonds and solvents effects at the surface. A microfibril of 81 chains, 9 per face, and DP 20 was taken as model of this study, using the GLYCAM06h force field and the TIP3P model. Model microfibrils were solvated with a 1.2 nm water buffer and were simulated during 10 ns, with equilibration times of 1 ns. Simulations in vacuo showed great deformation indicating that the presence

of water plays a critical role in mitigating the extent of twisting. Also, simulations of implicit solvents produced microfibrils that fragmented into constituent cellulose chains. TIP4PEW and TIP3P produced comparable results, however, TIP5P, which includes lone pairs on oxygen atoms to better reproduce both bulk water properties and the tetrahedral geometry of hydrogen bonds, resulted in a noticeable reduction of the twist. This means that solvent H-bond plays an essential role in stabilising the structure. TIP5P binds more tightly and displays higher occupancy in the first solvation shell as compared to TIP3P. Also, the interaction energy between microfibril and solvent from the force field showed that TIP5P is preferred over TIP3P due to more favourable electrostatic interactions. It could be that increased order structuring of solvent serves to restrict twisting motion and, therefore, results in better stabilisation of the structure. Using GLYCAM06EP (include lone pairs in carbohydrate oxygen atoms) and TIP5P yields considerable reduction of the twisting, in comparison to GLYCAM06 and TIP3P models.

Empirical force field methods keep being an affordable option to study the dynamics of large systems such as lignocellulose. MD simulations are a promising tool to quantitatively compare with experimental results as newer force field versions improve the parametrisation. It is essential to mention that previous versions of the GROMOS force field reproduce better the experimental data than more recent releases such as the GROMOS56Acarbo, that has been specially optimised for carbohydrates. Chen et al. [144] studied the influence of the nonbonded parameters, i.e., Lennard-Jones and the partial atomic charges using the GROMOS force field. It was found that the atomic partial charges and the LJ parameter of C1 are crucial to model cellulose allomorphs accurately. The unit cell parameters gave a prediction within 0.2% from the experimental value when the united-atom the repulsion term of CH1 was optimised, and the CHARMM charge set was imported.

Petridis et al. [107] studied the hydration of the hydrophilic surfaces of cellulose using the CHARMM36 force field and the TIP4P-EW model of water. The simulations were used to determine how hydration changes the rigidity of cellulose I α fibres. Their results show that hydration leads to an increased disorder in hydroxymethyl conformation on the cellulose surface and that the structure of in the bulk phase is well preserved. Moreover, an overall increase of local packing and water-cellulose bonds increase the stiffness when hydrated than dry.

The theoretical and experimental results of Langan et al. [145] sensitive at different scales have elucidated processes responsible for morphological changes in biomass during steam explosion pretreatment. Cellulose dehydration is partially responsible for such an outcome. The authors have suggested a mechanism in which the core water molecules leave the microfibrillar bundles and, as a consequence, the surrounding matrix goes through a phase separation, with more hydrophobic lignin component aggregation into crumpled globules. This phase separation and the pressure of lignin globules to separate seem to cause swelling and will expand the size of small voids by a few nanometres.

1.3.1.3 Cellulose in non-aqueous solvents

Whereas an extensive work on cellulose in water has been carried out, only a hand of computational studies approaches the interactions of the wood components as mentioned earlier with non-water or mixed solvents [10, 64, 92, 125-130].

Simulations of a methyl β -cellobioside [10] have shown that the intramolecular O3-H...O5' H-bond over the glycosidic linkage is present in apolar solvents but absent in polar solvents. More recently, quantum chemical calculations and MD simulations of cellulose with EDA [125] shows that the solvent strongly disrupts the naturally occurring hydrogen bond pattern, including the well preserved O3-H...O5' H-bond, as O3 prefers to donate to EDA instead of O5. Simulation of glucose solvation in water, THF, DMSO, and DMF[128] showed that the organic solvents compete with water to be in the first solvation shell of glucose and a significant amount of water is pushed to the second solvation shell.

Simulations of cellulose with ionic liquids, DMSO, and water [71] show that anions interact much more strongly with the cellulose surface than the cations, which is responsible for the dissolution of cellulose. Furthermore, the addition of DMSO or water strongly affects cellulose-chloride interactions. Zhao et al.[146] observed preferential solvation of anions by protic solvents with a reduction of the interactions between the anions and cellulose. Aprotic solvents form weak hydrogen bonds with the cations, thus enhancing the anion-cellulose interactions.

Moreover, a phase separation of binary water-organic solvents (THF, acetone, ethanol, and γ -valerolactone) on cellulose surfaces was reported [64, 129]. Chen et al. [147] demonstrated that the primary molecular mechanisms for urea absorption on cellulose are the Lennard-Jones dispersion energies as well as the gain in the translational entropy of water expelled from the cellulose surface, in water-urea mixtures. Moreover, the authors show that the decrease in the entropy of urea adsorbed on the cellulose surface is always lower than the increase in the entropy of water.

1.3.2 Hemicellulose

The most abundant hemicellulose backbone components have served as a model to study the conformation and flexibility of hemicelluloses. Simulations of low molecular weight of polysaccharides [111] have shown that the structure of the glycosidic linkage is highly flexible in xylose and can explore conformations that depend on the surrounding environment. Mazeau et al. [148] have reported that three low energy conformers stabilised by interresidue hydrogen bonds can be found. The two lowest minima are the extended (2_1) and threefold helical (3_2) conformers which are stabilised by hydrogen bonds between O3 and O5, whereas the third conformer forms hydrogen bonds between O3 and O2. It has been reported that free xylan adopts a left-hand threefold helical (3_2) conformation in water, whereas a twofold helical (2_1) structure is stabilised in the

presence of cellulose [112-114]. This conformation emerges from the distortion of the OH3 rotamer allowing water molecules to move freely around the glycosidic region and exchange with bulk water [113].

Moreover, simulations of polysaccharides [115] have shown lower solute-water and solute-solute hydrogen bonds and less hydrated surfaces in xylose when compared to glucose polymers. Hemicelluloses have been reported experimentally to turn into a gel with the addition of ethanol [149, 150], and previous MD have also reported a decrease of xylan's diffusion coefficient in ethanol-water mixture [130]. A xylan chain with 30 monomers in water-THF mixtures was studied by Smith et al. [151]. MD simulations, along with experimental results, indicate that the solvation of xylan depends on the phase miscibility of the water-THF mixture. Xylan is preferentially solvated by water in the immiscible mixture of THF and water. Furthermore, xylan's solubilisation is slower in the water-THF than in water.

1.3.3 Lignin

Density functional theory (DFT) methods have been used to examine structural features and stabilities of most common dimerisation products of lignin. Durbeej and Eriksson [124] have found that, in the absence of external hydrogen bond acceptors, intramolecular hydrogen bonding acts as a stabilising force in the lowest-energy conformers of several different dimeric coniferil alcohol structures.

The β -O-4 linkage represents the predominant inter-unit linkage in lignin, and it is an excellent model for studying major conformation features such as H-bonding and flexibility [17]. Both experimental and theoretical studies suggest that β -O-4 structures are flexible molecules that can adopt a large number of conformations [17, 121, 122]. The results of Charlier and Mazeau [123] and Besombes et al. [122] on the different torsion angles of the β -O-4 linkages in linkages have shown that the torsional angles simultaneously explore multiple orientations, thus indicating that lignin very flexible coil conformation. Charlier and Mazeau [123] have reported that the mobility of water molecules in the proximity of lignin is 1.5 times larger than that of carbohydrates. This is in line with the results of NMR data. MD studies have provided useful insights into the structures of lignin [17, 131, 152, 153]; however, by using force fields that have been developed to model proteins or other biomolecules, instead of lignin.

Petridis and Smith [154, 155] developed a molecular mechanics force field for lignin parametrised based on reproducing quantum mechanical data of model compounds. Thus, along with the already existing force fields for polysaccharides, the force field has opened the door towards full simulations of lignocellulose. Moreover, the force field has been recently reviewed, and a new version of the force field [156, 157] has been released, improving some features of the previous version and including the parametrisation of more units and linkages present in lignocellulose biomass.

All-atom MD simulations to examine the structure of lignin at four different temperatures and tetrahydrofuran (THF) concentrations [126, 127] address the interactions of mixed solvents with lignocellulosic biomass at a

molecular level. The simulation showed that THF, a polar aprotic ether, preferentially solvates lignin, and in doing so, shifts the configurational equilibrium distribution of the biopolymer from a crumpled globule to coil, independent of temperature. Whereas pure water is a bad solvent for lignin, the THF-water cosolvent acts as a θ -solvent, in which solvent-lignin and lignin-lignin interactions are approximately equivalent in strength. Under these conditions, polymers do not aggregate, thus providing a mechanism for the observed lignin solubilization that facilitates unfettered access of cellulolytic enzymes to cellulose [126].

Work has been done by the same research group in study the temperature dependence of lignin physics [118, 119], dependence of the shape of lignin polymers on chain length [117], its aggregation [120], lignin inhibition mechanism and enzymatic biomass deconstruction [158], and processes that drive the biomass pretreatment using models of hemicellulose and lignin [145], among others [132, 159, 160]. Most relevant contributions of their work show the tendency of softwood lignin polymers (degree of polymerisation from 6 to 41) to self-aggregate, even at high temperature. The shape of low polymerisation degree lignin is aspherical, and it has been explained due to favourable interactions of polar surface hydroxyl groups with water, which decrease the surface tension. On the other hand, spherical conformations for longer chain lengths seem to be favoured by monomer-monomer interactions and dominance of the entropic confinement of surface water [117].

1.3.4 Lignocellulose interactions

Molecular dynamics (MD) simulations, often in combination with experimental investigations, have provided valuable information about the adhesion between lignocellulose components. Besombes et al.[152, 153] observed that the aromatic rings of lignin adopt a preferential parallel orientation relative to the cellulose surface in water. Moreover, electrostatic interactions between the alcohol groups of lignin and the hydroxyls of cellulose are favoured, where lignin acts as an H-bond donor[152]. Thus, cellulose-lignin adhesion is controlled by both van der Waals and H-bond interactions [152]. Linder et al. [160] have observed that lignin binding to cellulose is favoured in the presence of water. Furthermore, it has been shown that lignin molecules preferentially aggregate into the hydrophobic faces of crystalline cellulose fibres [158, 160].

Kumar et al.[161] demonstrated that the cellulose-hemicellulose interactions at elevated temperatures increase cellulose recalcitrance. Falcoz-Vigne et al. [114]and Mazeau et al. [112]have observed that the interaction between xylan and cellulose is stronger when xylan is in the twofold conformation than when it is in the threefold one. Charlier and Mazeau [123]performed MD on a model of secondary cell plant composed by cellulose, lignin, hemicellulose, and water. The models showed a well-defined interface between cellulose and xylan, as well as interpenetration of xylan into the lignin domain.

The theoretical and experimental results of Langan et al. [145, 162] sensitive at different scales have elucidated processes responsible for morphological changes in biomass during steam explosion pretreatment.

Cellulose dehydration and lignin-hemicellulose phase separation are responsible for such an outcome. The authors have suggested a mechanism in which the core water molecules leave the microfibrillar bundles and, as a consequence, the surrounding matrix goes through a phase separation, with more hydrophobic lignin component aggregation into crumpled globules. This phase separation and the pressure of lignin globules to separate seem to cause swelling and will expand the size of small voids by a few nanometres.

Very recent NMR spectroscopy and dynamic nuclear polarisation experiments have revealed that lignin has abundant electrostatic interactions with the polar motifs of xylan [95]. Recently, simulations of the lignin-carbohydrate complex to study the elastic moduli and adhesion energies between the lignocellulose compounds [163] have shown that hemicellulose model has stronger mechanical properties than lignin whereas lignin exhibits greater tendency to adhere to cellulose microfibrils. This work also suggests that the various presence of hydrogen bonds in hemicellulose chains are responsible for improving the mechanical behaviour of the lignocellulosic complex. Moreover, it is highlighted that strong van de Waals forces between lignin molecules and cellulose microfibrils are responsible for higher adhesion energy in the complex. The effect of hemicellulose composition and the nano forces that control the cell wall strength have been addressed by using 3D-RISM-KH molecular theory of solvation [164].

However, the non-aqueous solvent effect on the adhesion between lignocellulose components is still unexplored. Recently, Yang et al. [165] studied at the quantum level the solvent effect on the pairwise adhesion between several tetramer models of cell wall components using polarised continuum models to represent water, methanol, and chloroform. Their results show that implicit solvation methods does not replicate the interaction energies calculated with explicit solvation of water molecules, as expected, because polarised continuum models do not consider the strong short-range interactions such as hydrogen bonding. Moreover, their results show that the variations in model interaction with the solvent dielectric constant in the polarised continuum model are significant and non-systematic [165]. Thus, the explicit inclusion of the solvent is fundamental to consider the strong short-range interactions. The quantum inclusion of the solvent is computationally expensive and can result in local potential energy minima [165]; however, MD simulation of explicitly solvated interactions can provide a compromise between cost and accuracy.

References

1. Nishiyama, Y., P. Langan, and H. Chanzy, *Crystal Structure and Hydrogen-Bonding System in Cellulose I β from Synchrotron X-ray and Neutron Fiber Diffraction*. Journal of the American Chemical Society, 2002. **124**(31): p. 9074-9082.
2. Sun, J.X., et al., *Isolation and characterization of cellulose from sugarcane bagasse*. Polymer Degradation and Stability, 2004. **84**(2): p. 331-339.
3. Fernandes, A.N., et al., *Nanostructure of cellulose microfibrils in spruce wood*. Proceedings of the National Academy of Sciences, 2011. **108**(47): p. E1195-E1203.

4. Parham, R.A. and R.L. Gray, *Formation and Structure of Wood*, in *The Chemistry of Solid Wood*. 1984, American Chemical Society. p. 3-56.
5. Hoffmann, P. and M.A. Jones, *Structure and Degradation Process for Waterlogged Archaeological Wood*, in *Archaeological Wood*. 1989, American Chemical Society. p. 35-65.
6. Muzamal, M., *Structural Modifications in Spruce Wood during Steam Explosion Pretreatment*, in *Department of Chemistry and Chemical Engineering*. 2016, Chalmers University of Technology: SE-412 96 Gothenburg.
7. Gomes, T.C.F. and M.S. Skaf, *Cellulose-Builder: A toolkit for building crystalline structures of cellulose*. *Journal of Computational Chemistry*, 2012. **33**(14): p. 1338-1346.
8. Pettersen, R.C., *The Chemical Composition of Wood*, in *The Chemistry of Solid Wood*. 1984, American Chemical Society. p. 57-126.
9. Heiner, A.P., L. Kuutti, and O. Teleman, *Comparison of the interface between water and four surfaces of native crystalline cellulose by molecular dynamics simulations*. *Carbohydrate Research*, 1998. **306**(1-2): p. 205-220.
10. Leeflang, B.R., et al., *A 1H-NMR and MD study of intramolecular hydrogen bonds in methyl β -cellobioside*. *Carbohydrate Research*, 1992. **230**(1): p. 41-61.
11. Lee, H.V., S.B.A. Hamid, and S.K. Zain, *Conversion of Lignocellulosic Biomass to Nanocellulose: Structure and Chemical Process*. *The Scientific World Journal*, 2014. **2014**: p. 20.
12. Ebringerová, A. and T. Heinze, *Xylan and xylan derivatives – biopolymers with valuable properties, 1. Naturally occurring xylnans structures, isolation procedures and properties*. *Macromolecular Rapid Communications*, 2000. **21**(9): p. 542-556.
13. ACS. *Molecule of the week*. 2012; Available from: <https://www.acs.org/content/acs/en/molecule-of-the-week/archive/l/lignin.html>.
14. Glasser, W.G., et al. *Studies of the molecular interaction between cellulose and lignin as a model for the hierarchical structure of wood*. in *American Chemical Society Symposium series 688; Orlando, FL; 1996 August 25-29; Washington, DC, American Chemical Society: 265-282*. 1998.
15. Horwath, W., *Chapter 12 - Carbon Cycling: The Dynamics and Formation of Organic Matter*, in *Soil Microbiology, Ecology and Biochemistry (Fourth Edition)*, E.A. Paul, Editor. 2015, Academic Press: Boston. p. 339-382.
16. Guadix-Montero, S. and M. Sankar, *Review on Catalytic Cleavage of C–C Inter-unit Linkages in Lignin Model Compounds: Towards Lignin Depolymerisation*. *Topics in Catalysis*, 2018. **61**(3): p. 183-198.
17. Besombes, S. and K. Mazeau, *Molecular dynamics simulations of a guaiacyl β -O-4 lignin model compound: Examination of intramolecular hydrogen bonding and conformational flexibility*. *Biopolymers*, 2004. **73**(3): p. 301-315.
18. McKendry, P., *Energy production from biomass (part 1): overview of biomass*. *Bioresource Technology*, 2002. **83**(1): p. 37-46.
19. Coseri, S., *Cellulose: To depolymerize... or not to?* *Biotechnology Advances*.

20. Cao, Y., et al., *Imidazolium-based ionic liquids for cellulose pretreatment: recent progresses and future perspectives*. Applied Microbiology and Biotechnology, 2017. **101**(2): p. 521-532.
21. Tan, H.-T., K.R. Corbin, and G.B. Fincher, *Emerging Technologies for the Production of Renewable Liquid Transport Fuels from Biomass Sources Enriched in Plant Cell Walls*. Frontiers in Plant Science, 2016. **7**(1854).
22. Vanneste, J., et al., *Unconventional Pretreatment of Lignocellulose with Low-Temperature Plasma*. ChemSusChem, 2017. **10**(1): p. 14-31.
23. Patinvoh, R.J., et al., *Innovative pretreatment strategies for biogas production*. Bioresource Technology, 2017. **224**: p. 13-24.
24. Endes, C., et al., *A critical review of the current knowledge regarding the biological impact of nanocellulose*. Journal of Nanobiotechnology, 2016. **14**(1): p. 78.
25. Ummartyotin, S. and H. Manuspiya, *A critical review on cellulose: From fundamental to an approach on sensor technology*. Renewable and Sustainable Energy Reviews, 2015. **41**: p. 402-412.
26. Suhas, et al., *Cellulose: A review as natural, modified and activated carbon adsorbent*. Bioresource Technology, 2016. **216**: p. 1066-1076.
27. Peng, F., et al., *Chemicals from Hemicelluloses: A Review*, in *Sustainable Production of Fuels, Chemicals, and Fibers from Forest Biomass*. 2011, American Chemical Society. p. 219-259.
28. Hansen, N.M.L. and D. Plackett, *Sustainable Films and Coatings from Hemicelluloses: A Review*. Biomacromolecules, 2008. **9**(6): p. 1493-1505.
29. Silva, A.E.d., et al., *Xylan, a Promising Hemicellulose for Pharmaceutical Use*, in *Products and Applications of Biopolymers*, J. Verbeek, Editor. 2012, InTEch.
30. Gírio, F.M., et al., *Hemicelluloses for fuel ethanol: A review*. Bioresource Technology, 2010. **101**(13): p. 4775-4800.
31. Sella Kapu, N. and H.L. Trajano, *Review of hemicellulose hydrolysis in softwoods and bamboo*. Biofuels, Bioproducts and Biorefining, 2014. **8**(6): p. 857-870.
32. Delidovich, I., K. Leonhard, and R. Palkovits, *Cellulose and hemicellulose valorisation: an integrated challenge of catalysis and reaction engineering*. Energy & Environmental Science, 2014. **7**(9): p. 2803-2830.
33. Scheller, H.V. and P. Ulvskov, *Hemicellulose: Annual Review of Plant Biology*. Annual Reviews, 2010. **61**: p. 1-720.
34. Upton, B.M. and A.M. Kasko, *Strategies for the Conversion of Lignin to High-Value Polymeric Materials: Review and Perspective*. Chemical Reviews, 2016. **116**(4): p. 2275-2306.
35. Phillips, M., *The Chemistry of Lignin*. Chemical Reviews, 1934. **14**(1): p. 103-170.
36. Bruijninx, P.C.A., R. Rinaldi, and B.M. Weckhuysen, *Unlocking the potential of a sleeping giant: lignins as sustainable raw materials for renewable fuels, chemicals and materials*. Green Chemistry, 2015. **17**(11): p. 4860-4861.
37. Sen, S., S. Patil, and D.S. Argyropoulos, *Thermal properties of lignin in copolymers, blends, and composites: a review*. Green Chemistry, 2015. **17**(11): p. 4862-4887.

38. Liu, W.-J., H. Jiang, and H.-Q. Yu, *Thermochemical conversion of lignin to functional materials: a review and future directions*. Green Chemistry, 2015. **17**(11): p. 4888-4907.
39. Wang, Y.-Y., Y.-r. Chen, and S. Sarkanen, *Path to plastics composed of ligninsulphonates (lignosulfonates)*. Green Chemistry, 2015. **17**(11): p. 5069-5078.
40. Yunpu, W., et al., *Review of microwave-assisted lignin conversion for renewable fuels and chemicals*. Journal of Analytical and Applied Pyrolysis, 2016. **119**: p. 104-113.
41. Li, M.-F., S. Yang, and R.-C. Sun, *Recent advances in alcohol and organic acid fractionation of lignocellulosic biomass*. Bioresource Technology, 2016. **200**: p. 971-980.
42. Grande, P.M., et al., *Fractionation of lignocellulosic biomass using the OrganoCat process*. Green Chemistry, 2015. **17**(6): p. 3533-3539.
43. Lali, A.M., J.S. Varavadekar, and P.C. Wadekar, *Process for fractionation of biomass* 2011.
44. Carneiro, A.P., O. Rodríguez, and E.A. Macedo, *Dissolution and fractionation of nut shells in ionic liquids*. Bioresource Technology, 2017. **227**: p. 188-196.
45. Yoo, C.G., *Pretreatment and fractionation of lignocellulosic biomass for production of biofuel and value-added products*. 2012, Iowa State University.
46. Mussatto, S.I. and G.M. Dragone, *Chapter 1 - Biomass Pretreatment, Biorefineries, and Potential Products for a Bioeconomy Development*, in *Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery*. 2016, Elsevier: Amsterdam. p. 1-22.
47. Romero, I., E. Ruiz, and E. Castro, *Chapter 10 - Pretreatment With Metal Salts A2 - Mussatto, Solange I*, in *Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery*. 2016, Elsevier: Amsterdam. p. 209-227.
48. Santos, J.C., et al., *Chapter 9 - Biomass Pretreatment With Oxalic Acid for Value-Added Products A2 - Mussatto, Solange I*, in *Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery*. 2016, Elsevier: Amsterdam. p. 187-208.
49. Shuddhodana, et al., *Chapter 23 - Enzymatic Hydrolysis of Lignocellulosic Residues A2 - Mussatto, Solange I*, in *Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery*. 2016, Elsevier: Amsterdam. p. 543-560.
50. Wikandari, R., R. Millati, and M.J. Taherzadeh, *Chapter 12 - Pretreatment of Lignocelluloses With Solvent N-Methylmorpholine N-oxide A2 - Mussatto, Solange I*, in *Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery*. 2016, Elsevier: Amsterdam. p. 255-280.
51. Xu, J.K. and R.C. Sun, *Chapter 19 - Recent Advances in Alkaline Pretreatment of Lignocellulosic Biomass A2 - Mussatto, Solange I*, in *Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery*. 2016, Elsevier: Amsterdam. p. 431-459.
52. Zhang, C., et al., *Chapter 22 - Sulfite Pretreatment to Overcome the Recalcitrance of Lignocelluloses for Bioconversion of Woody Biomass A2 - Mussatto, Solange I*, in *Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery*. 2016, Elsevier: Amsterdam. p. 499-541.
53. Kim, J.S., Y.Y. Lee, and T.H. Kim, *A review on alkaline pretreatment technology for bioconversion of lignocellulosic biomass*. Bioresource Technology, 2016. **199**: p. 42-48.

54. Zhang, T., R. Kumar, and C.E. Wyman, *Sugar yields from dilute oxalic acid pretreatment of maple wood compared to those with other dilute acids and hot water*. Carbohydrate Polymers, 2013. **92**(1): p. 334-344.
55. Lee, C., Y. Zheng, and J.S. VanderGheynst, *Effects of pretreatment conditions and post-pretreatment washing on ethanol production from dilute acid pretreated rice straw*. Biosystems Engineering, 2015. **137**: p. 36-42.
56. Zhuang, X., et al., *Liquid hot water pretreatment of lignocellulosic biomass for bioethanol production accompanying with high valuable products*. Bioresource Technology, 2016. **199**: p. 68-75.
57. Zhao, X., K. Cheng, and D. Liu, *Organosolv pretreatment of lignocellulosic biomass for enzymatic hydrolysis*. Applied Microbiology and Biotechnology, 2009. **82**(5): p. 815.
58. Silveira, M.H.L., et al., *Current Pretreatment Technologies for the Development of Cellulosic Ethanol and Biorefineries*. ChemSusChem, 2015. **8**(20): p. 3366-3390.
59. Sathitsuksanoh, N., A. George, and Y.-H.P. Zhang, *New lignocellulose pretreatments using cellulose solvents: a review*. Journal of Chemical Technology & Biotechnology, 2013. **88**(2): p. 169-180.
60. Renders, T., et al., *Synergetic Effects of Alcohol/Water Mixing on the Catalytic Reductive Fractionation of Poplar Wood*. ACS Sustainable Chemistry & Engineering, 2016. **4**(12): p. 6894-6904.
61. Motagamwala, A.H., et al., *An engineered solvent system for sugar production from lignocellulosic biomass using biomass derived γ -valerolactone*. Green Chemistry, 2016. **18**(21): p. 5756-5763.
62. Sannigrahi, P., S.J. Miller, and A.J. Ragauskas, *Effects of organosolv pretreatment and enzymatic hydrolysis on cellulose structure and crystallinity in Loblolly pine*. Carbohydrate Research, 2010. **345**(7): p. 965-970.
63. Cai, C.M., et al., *Coupling metal halides with a co-solvent to produce furfural and 5-HMF at high yields directly from lignocellulosic biomass as an integrated biofuels strategy*. Green Chemistry, 2014. **16**(8): p. 3819-3829.
64. Mostofian, B., et al., *Local Phase Separation of Co-solvents Enhances Pretreatment of Biomass for Bioenergy Applications*. Journal of the American Chemical Society, 2016. **138**(34): p. 10869-10878.
65. Chundawat, S.P.S., et al., *Restructuring the Crystalline Cellulose Hydrogen Bond Network Enhances Its Depolymerization Rate*. Journal of the American Chemical Society, 2011. **133**(29): p. 11163-11174.
66. Xiao, M. and M.W. Frey, *The role of salt on cellulose dissolution in ethylene diamine/salt solvent systems*. Cellulose, 2007. **14**(3): p. 225-234.
67. Teymouri, F., et al., *Optimization of the ammonia fiber explosion (AFEX) treatment parameters for enzymatic hydrolysis of corn stover*. Bioresource Technology, 2005. **96**(18): p. 2014-2018.
68. Swatloski, R.P., et al., *Dissolution of Cellose with Ionic Liquids*. Journal of the American Chemical Society, 2002. **124**(18): p. 4974-4975.
69. Gericke, M., T. Liebert, and T. Heinze, *Interaction of Ionic Liquids with Polysaccharides, 8 – Synthesis of Cellulose Sulfates Suitable for Polyelectrolyte Complex Formation*. Macromolecular Bioscience, 2009. **9**(4): p. 343-353.

70. Ferreira, D.C., et al., *Binary mixtures of ionic liquids-DMSO as solvents for the dissolution and derivatization of cellulose: Effects of alkyl and alkoxy side chains*. Carbohydrate Polymers, 2019. **212**: p. 206-214.
71. Huo, F., Z. Liu, and W. Wang, *Cosolvent or Antisolvent? A Molecular View of the Interface between Ionic Liquids and Cellulose upon Addition of Another Molecular Solvent*. The Journal of Physical Chemistry B, 2013. **117**(39): p. 11780-11792.
72. Gericke, M., et al., *Tailored Media for Homogeneous Cellulose Chemistry: Ionic Liquid/Co-Solvent Mixtures*. Macromolecular Materials and Engineering, 2011. **296**(6): p. 483-493.
73. Ghosh, A., "Production of solubilized carbohydrates from lignocellulosic biomass using solvent liquefaction" (2018). , Iowa State University.
74. Domínguez-Robles, J., et al., *Aqueous acetone fractionation of kraft, organosolv and soda lignins*. International Journal of Biological Macromolecules, 2018. **106**: p. 979-987.
75. Yang, Y. and F. Zhang, *Ultrasound-assisted extraction of rutin and quercetin from Euonymus alatus (Thunb.) Sieb.* Ultrasonics Sonochemistry, 2008. **15**(4): p. 308-313.
76. Stamm, A.J., *SHRINKING and SWELLING of WOOD*. Industrial & Engineering Chemistry, 1935. **27**(4): p. 401-406.
77. Nayer, A.N. and R.L. Hossfeld, *Hydrogen Bonding and the Swelling of Wood in Various Organic Liquids*. Journal of the American Chemical Society, 1949. **71**(8): p. 2852-2855.
78. Adebawo, F.G., et al., *Morphological and interfacial properties of chemically-modified tropical hardwood*. RSC Advances, 2016. **6**(8): p. 6571-6576.
79. Hasselblatt, M., *Der Wasserdampfdruck und die elektrische Leitfähigkeit des Holzes in Abhängigkeit von seinem Wassergehalt*. Zeitschrift für anorganische und allgemeine Chemie, 1926. **154**(1): p. 375-385.
80. Finch, C.A., *The chemistry of solid wood*. Edited by R. M. Rowell, ACS Advances in Chemistry Series No. 207, American Chemical Society, Washington D.C., 1984. pp. x + 614, price \$79.95 (USA & Canada), \$95.95 (rest of world). ISBN 0-8412-0796-8. British Polymer Journal, 1986. **18**(5): p. 352-352.
81. Mantanis, G.I., R.A. Young, and R.M. Rowell, *Swelling of wood*. Wood Science and Technology, 1994. **28**(2): p. 119-134.
82. Ishimaru, Y., S. Narimoto, and I. Iida, *Mechanical properties of wood swollen in organic liquids with two or more functional groups for hydrogen bonding in a molecule*. Journal of Wood Science, 2001. **47**(3): p. 171.
83. Rafsanjani, A., et al., *Hygroscopic swelling and shrinkage of latewood cell wall micropillars reveal ultrastructural anisotropy*. Journal of The Royal Society Interface, 2014. **11**(95).
84. Schuerch, C., *The Solvent Properties of Liquids and Their Relation to the Solubility, Swelling, Isolation and Fractionation of Lignin*. Journal of the American Chemical Society, 1952. **74**(20): p. 5061-5067.
85. Mantanis George, I., A. Young Raymond, and M. Rowell Roger, *Swelling of Wood. Part II. Swelling in Organic Liquids*, in *Holzforschung - International Journal of the Biology, Chemistry, Physics and Technology of Wood*. 1994. p. 480.

86. Gordy, W. and S.C. Stanford, *Spectroscopic Evidence for Hydrogen Bonds: Comparison of Proton-Attracting Properties of Liquids. II*. The Journal of Chemical Physics, 1940. **8**(2): p. 170-177.
87. Prusov, A.N., S.M. Prusova, and A.G. Zakharov, *Interaction of cellulose and lignocellulosic polymers with water and aqueous systems*. Russian Chemical Bulletin, 2014. **63**(9): p. 1926-1945.
88. MEIER, P., T. KAPS, and U. KALLAVUS, *Swelling of Pinewood (Pinus Sylvestris) in Binary Aqueous Solutions of Organic Substances*. MATERIALS SCIENCE (MEDŽIAGOTYRA), 2005. **11**(2).
89. MEIER, P., et al., *Multiple Swelling of Pinewood (Pinus Sylvestris) in Binary and Ternary Mixtures of Ethanol, Acetone and Water* MATERIALS SCIENCE (MEDŽIAGOTYRA), 2006. **12**(1).
90. Meiera, P., et al., *Mechanical properties of pinewood (Pinus Sylvestris) swollen in organic liquids* Proc. Estonian Acad. Sci. Eng., 2006. **12**(2): p. 125–133
91. Chang, S., et al., *Solvent polarity and internal stresses control the swelling behaviour of green wood during dehydration in organic solution*. Bioresources, 2012. **7**(2).
92. Aguilera-Segura, S.M., et al., *Synergistic sorption of mixed solvents in wood cell walls: experimental and theoretical approach*. Macromolecular symposia, 2019. **386**(1): p. 1900022.
93. Wierzbicki, M.P., et al., *Xylan in the Middle: Understanding Xylan Biosynthesis and Its Metabolic Dependencies Toward Improving Wood Fiber for Industrial Processing*. Frontiers in Plant Science, 2019. **10**(176).
94. Petridis, L. and J.C. Smith, *Molecular-level driving forces in lignocellulosic biomass deconstruction for bioenergy*. Nature Reviews Chemistry, 2018. **2**(11): p. 382-389.
95. Kang, X., et al., *Lignin-polysaccharide interactions in plant secondary cell walls revealed by solid-state NMR*. Nature Communications, 2019. **10**(1): p. 347.
96. Zhang, Y., et al., *Recent progress in theoretical and computational studies on the utilization of lignocellulosic materials*. Green Chemistry, 2019. **21**(1): p. 9-35.
97. Donev, E., et al., *Engineering Non-cellulosic Polysaccharides of Wood for the Biorefinery*. Frontiers in Plant Science, 2018. **9**(1537).
98. Lambert, E., et al., *Real Time and Quantitative Imaging of Lignocellulosic Films Hydrolysis by Atomic Force Microscopy Reveals Lignin Recalcitrance at Nanoscale*. Biomacromolecules, 2019. **20**(1): p. 515-527.
99. Polavarapu, P.L. and C.S. Ewig, *Ab Initio computed molecular structures and energies of the conformers of glucose*. Journal of Computational Chemistry, 1992. **13**(10): p. 1255-1261.
100. Kroon-Batenburg, L.M.J. and J. Kroon, *Solvent effect on the conformation of the hydroxymethyl group established by molecular dynamics simulations of methyl-β-D-glucoside in water*. Biopolymers, 1990. **29**(8-9): p. 1243-1248.
101. van Eijck, B.P., R.W.W. Hooft, and J. Kroon, *Molecular dynamics study of conformational and anomeric equilibria in aqueous D-glucose*. The Journal of Physical Chemistry, 1993. **97**(46): p. 12093-12099.
102. Heiner, A.P. and O. Teleman, *Interface between Monoclinic Crystalline Cellulose and Water: Breakdown of the Odd/Even Duplicity*. Langmuir, 1997. **13**(3): p. 511-518.

103. Matthews, J.F., M.E. Himmel, and J.W. Brady, *Simulations of the Structure of Cellulose*, in *Computational Modeling in Lignocellulosic Biofuel Production*. 2010, American Chemical Society. p. 17-53.
104. Matthews, J.F., et al., *Comparison of Cellulose I β Simulations with Three Carbohydrate Force Fields*. *Journal of Chemical Theory and Computation*, 2012. **8**(2): p. 735-748.
105. Matthews, J.F., et al., *High-Temperature Behavior of Cellulose I*. *The Journal of Physical Chemistry B*, 2011. **115**(10): p. 2155-2166.
106. Hadden, J.A., A.D. French, and R.J. Woods, *Unraveling Cellulose Microfibrils: A Twisted Tale*. *Biopolymers*, 2013. **99**(10): p. 746-756.
107. Petridis, L., et al., *Hydration Control of the Mechanical and Dynamical Properties of Cellulose*. *Biomacromolecules*, 2014. **15**(11): p. 4152-4159.
108. Matthews, J.F., et al., *Computer simulation studies of microcrystalline cellulose I β* . *Carbohydrate Research*, 2006. **341**(1): p. 138-152.
109. Yui, T., et al., *Swelling behavior of the cellulose I β crystal models by molecular dynamics*. *Carbohydrate Research*, 2006. **341**(15): p. 2521-2530.
110. Mazeau, K. and L. Heux, *Molecular Dynamics Simulations of Bulk Native Crystalline and Amorphous Structures of Cellulose*. *The Journal of Physical Chemistry B*, 2003. **107**(10): p. 2394-2403.
111. Berglund, J., et al., *A molecular dynamics study of the effect of glycosidic linkage type in the hemicellulose backbone on the molecular chain flexibility*. *The Plant Journal*, 2016. **88**(1): p. 56-70.
112. Mazeau, K. and L. Charlier, *The molecular basis of the adsorption of xylans on cellulose surface*. *Cellulose*, 2012. **19**(2): p. 337-349.
113. Almond, A. and J.K. Sheehan, *Predicting the molecular shape of polysaccharides from dynamic interactions with water*. *Glycobiology*, 2003. **13**(4): p. 255-264.
114. Falcoz-Vigne, L., et al., *Quantification of a tightly adsorbed monolayer of xylan on cellulose surface*. *Cellulose*, 2017. **24**(9): p. 3725-3739.
115. Sauter, J. and A. Grafmüller, *Solution Properties of Hemicellulose Polysaccharides with Four Common Carbohydrate Force Fields*. *Journal of Chemical Theory and Computation*, 2015. **11**(4): p. 1765-1774.
116. Stark, N.M., D.J. Yelle, and U.P. Agarwal, *4 - Techniques for Characterizing Lignin*, in *Lignin in Polymer Composites*, O. Faruk and M. Sain, Editors. 2016, William Andrew Publishing. p. 49-66.
117. Petridis, L. and J.C. Smith, *Conformations of Low-Molecular-Weight Lignin Polymers in Water*. *ChemSusChem*, 2016. **9**(3): p. 289-295.
118. Petridis, L., R. Schulz, and J.C. Smith, *Simulation Analysis of the Temperature Dependence of Lignin Structure and Dynamics*. *Journal of the American Chemical Society*, 2011. **133**(50): p. 20277-20287.
119. Hong, L., L. Petridis, and J.C. Smith, *Biomolecular Structure and Dynamics with Neutrons: The View from Simulation*. *Israel Journal of Chemistry*, 2014. **54**(8-9): p. 1264-1273.
120. Petridis, L., et al., *Self-similar multiscale structure of lignin revealed by neutron scattering and molecular dynamics simulation*. *Physical Review E*, 2011. **83**(6): p. 061911.

121. Stomberg, R. and K. Lundquist, *Stereochemistry of lignin structures of the β -O-4 type*, in *Nordic Pulp & Paper Research Journal*. 1994. p. 37.
122. Besombes, S., et al., *Molecular Modeling of Lignin β -O-4 Model Compounds. Comparative Study of the Computed and Experimental Conformational Properties for a Guaiacyl β -O-4 Dimer*, in *Holzforschung*. 2003. p. 266.
123. Charlier, L. and K. Mazeau, *Molecular Modeling of the Structural and Dynamical Properties of Secondary Plant Cell Walls: Influence of Lignin Chemistry*. *The Journal of Physical Chemistry B*, 2012. **116**(14): p. 4163-4174.
124. Durbeej, B. and L.A. Eriksson, *A Density Functional Theory Study of Coniferyl Alcohol Intermonomeric Cross Linkages in Lignin - Three-Dimensional Structures, Stabilities and the Thermodynamic Control Hypothesis*, in *Holzforschung*. 2003. p. 150.
125. Sawada, D., et al., *Structure and dynamics of a complex of cellulose with EDA: insights into the action of amines on cellulose*. *Cellulose*, 2013. **20**(4): p. 1563-1571.
126. Smith, M.D., et al., *Cosolvent pretreatment in cellulosic biofuel production: effect of tetrahydrofuran-water on lignin structure and dynamics*. *Green Chemistry*, 2016. **18**(5): p. 1268-1277.
127. Smith, M.D., et al., *Enhanced sampling simulation analysis of the structure of lignin in the THF-water miscibility gap*. *Physical Chemistry Chemical Physics*, 2016. **18**(9): p. 6394-6398.
128. Vasudevan, V. and S.H. Mushrif, *Insights into the solvation of glucose in water, dimethyl sulfoxide (DMSO), tetrahydrofuran (THF) and N,N-dimethylformamide (DMF) and its possible implications on the conversion of glucose to platform chemicals*. *RSC Advances*, 2015. **5**(27): p. 20756-20763.
129. Smith, M.D., et al., *Organosolv-Water Cosolvent Phase Separation on Cellulose and its Influence on the Physical Deconstruction of Cellulose: A Molecular Dynamics Analysis*. *Scientific Reports*, 2017. **7**(1): p. 14494.
130. Umemura, M. and Y. Yuguchi, *Solvation of xyloglucan in water/alcohol systems by molecular dynamics simulation*. *Cellulose*, 2009. **16**(3): p. 361-371.
131. Durbeej, B., Y.-N. Wang, and L.A. Eriksson, *Lignin Biosynthesis and Degradation — a Major Challenge for Computational Chemistry*, in *High Performance Computing for Computational Science — VECPAR 2002: 5th International Conference Porto, Portugal, June 26–28, 2002 Selected Papers and Invited Talks*, J.M.L.M. Palma, et al., Editors. 2003, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 137-165.
132. Sangha, A.K., et al., *Molecular simulation as a tool for studying lignin*. *Environmental Progress & Sustainable Energy*, 2012. **31**(1): p. 47-54.
133. Ha, S., et al., *Solvent effect on the anomeric equilibrium in D-glucose: a free energy simulation analysis*. *Journal of the American Chemical Society*, 1991. **113**(5): p. 1553-1557.
134. Schmidt, R.K., M. Karplus, and J.W. Brady, *The Anomeric Equilibrium in d-Xylose: Free Energy and the Role of Solvent Structuring*. *Journal of the American Chemical Society*, 1996. **118**(3): p. 541-546.
135. Ha, S.N., et al., *A revised potential-energy surface for molecular mechanics studies of carbohydrates*. *Carbohydrate Research*, 1988. **180**(2): p. 207-221.

136. Kuttel, M., J.W. Brady, and K.J. Naidoo, *Carbohydrate solution simulations: Producing a force field with experimentally consistent primary alcohol rotational frequencies and populations*. Journal of Computational Chemistry, 2002. **23**(13): p. 1236-1243.
137. Bonanno, G., R. Noto, and S.L. Fornili, *Water interaction with [small alpha],[small alpha]-trehalose: molecular dynamics simulation*. Journal of the Chemical Society, Faraday Transactions, 1998. **94**(18): p. 2755-2762.
138. Guvench, O., et al., *Additive empirical force field for hexopyranose monosaccharides*. J Comput Chem, 2008. **29**.
139. Guvench, O., et al., *CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses*. J Chem Theory Comput, 2009. **5**.
140. Kirschner, K.N., et al., *GLYCAM06: A generalizable biomolecular force field. Carbohydrates*. Journal of Computational Chemistry, 2008. **29**(4): p. 622-655.
141. Hansen, H.S. 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*. Journal of Computational Chemistry, 2011. **32**(6): p. 998-1032.
142. Guvench, O., et al., *CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling*. J Chem Theory Comput, 2011. **7**.
143. Srinivas, G., X. Cheng, and J.C. Smith, *A Solvent-Free Coarse Grain Model for Crystalline and Amorphous Cellulose Fibrils*. Journal of Chemical Theory and Computation, 2011. **7**(8): p. 2539-2548.
144. Chen, P., Y. Nishiyama, and K. Mazeau, *Atomic partial charges and one Lennard-Jones parameter crucial to model cellulose allomorphs*. Cellulose, 2014. **21**(4): p. 2207-2217.
145. Langan, P., et al., *Common processes drive the thermochemical pretreatment of lignocellulosic biomass*. Green Chemistry, 2014. **16**(1): p. 63-68.
146. Zhao, H., et al., *Designing enzyme-compatible ionic liquids that can dissolve carbohydrates*. Green Chemistry, 2008. **10**(6): p. 696-705.
147. Chen, P., et al., *Translational Entropy and Dispersion Energy Jointly Drive the Adsorption of Urea to Cellulose*. The Journal of Physical Chemistry B, 2017. **121**(10): p. 2244-2251.
148. Mazeau, K., et al., *Conformational analysis of xylan chains*. Carbohydrate Research, 2005. **340**(18): p. 2752-2760.
149. Yuguchi, Y., et al., *Gelation of xyloglucan ID water/alcohol systems*. Cellulose, 2004. **11**(2): p. 203-208.
150. Yamanaka, S., et al., *Gelation of tamarind seed polysaccharide xyloglucan in the presence of ethanol*. Food Hydrocolloids, 2000. **14**(2): p. 125-128.
151. Smith, M.D., et al., *Temperature-dependent phase behaviour of tetrahydrofuran–water alters solubilization of xylan to improve co-production of furfurals from lignocellulosic biomass*. Green Chemistry, 2018. **20**(7): p. 1612-1620.

152. Besombes, S. and K. Mazeau, *The cellulose/lignin assembly assessed by molecular modeling. Part 1: adsorption of a threo guaiacyl beta-O-4 dimer onto a Ibeta cellulose whisker*. Plant physiology and biochemistry : PPB, 2005. **43**(3): p. 299-308.
153. Besombes, S. and K. Mazeau, *The cellulose/lignin assembly assessed by molecular modeling. Part 2: seeking for evidence of organization of lignin molecules at the interface with cellulose*. Plant Physiology and Biochemistry, 2005. **43**(3): p. 277-286.
154. Petridis, L. and J.C. Smith, *A molecular mechanics force field for lignin*. Journal of Computational Chemistry, 2009. **30**(3): p. 457-467.
155. Petridis, L. and J.C. Smith, *Cellulosic ethanol: progress towards a simulation model of lignocellulosic biomass*. Journal of Physics: Conference Series, 2008. **125**(1): p. 012055.
156. Vermaas, J.V., et al., *Automated Transformation of Lignin Topologies into Atomic Structures with LigninBuilder*. ACS Sustainable Chemistry & Engineering, 2019. **7**(3): p. 3443-3453.
157. Vermaas, J.V., et al., *Systematic parameterization of lignin for the CHARMM force field*. Green Chemistry, 2019. **21**(1): p. 109-122.
158. Vermaas, J.V., et al., *Mechanism of lignin inhibition of enzymatic biomass deconstruction*. Biotechnology for Biofuels, 2015. **8**(1): p. 217.
159. Schulz, R., et al., *Scaling of Multimillion-Atom Biological Molecular Dynamics Simulation on a Petascale Supercomputer*. Journal of Chemical Theory and Computation, 2009. **5**(10): p. 2798-2808.
160. Lindner, B., et al., *Solvent-Driven Preferential Association of Lignin with Regions of Crystalline Cellulose in Molecular Dynamics Simulation*. Biomacromolecules, 2013. **14**(10): p. 3390-3398.
161. Kumar, R., et al., *Cellulose-hemicellulose interactions at elevated temperatures increase cellulose recalcitrance to biological conversion*. Green Chemistry, 2018. **20**(4): p. 921-934.
162. Pingali, S.V., et al., *Morphological changes in the cellulose and lignin components of biomass occur at different stages during steam pretreatment*. Cellulose, 2014. **21**(2): p. 873-878.
163. Youssefian, S. and N. Rahbar, *Molecular Origin of Strength and Stiffness in Bamboo Fibrils*. Scientific Reports, 2015. **5**: p. 11116.
164. Silveira, R.L., et al., *Plant Biomass Recalcitrance: Effect of Hemicellulose Composition on Nanoscale Forces that Control Cell Wall Strength*. Journal of the American Chemical Society, 2013. **135**(51): p. 19048-19051.
165. Yang, H., et al., *Quantum Calculations on Plant Cell Wall Component Interactions*. Interdisciplinary Sciences: Computational Life Sciences, 2018.

Chapter 2. Theoretical background

Contents

Abstract	67
1. Fundamentals of Statistical Mechanics	67
2. Molecular Dynamics Simulations	69
2.2 Initialization	70
2.2 Force Calculation	70
2.3 Integrating the Equations of motion	71
2.4 Commonly used software for MD simulations.....	72
2.5 Force Field	72
3. Quantum Chemistry (QC) background	77
3.1 Density Functional Theory.....	80
3.2 <i>Multiconfigurational Methods</i>	88
3.3 QM program packages used in this work	90
4. QM/MM Hybrid methods.....	91
4.1 QM/MM Energy: subtractive scheme	91
4.2 QM/MM Energy: additive scheme	92
4.3 Mechanical inclusion	92
4.4 Electrostatic inclusion.....	92
4.5 Polarisable inclusion	93
4.6 Saturation of broken bonds at the QM / MM boundary.....	93
4.7 QM/MM program packages used in this work.....	93
REFERENCES	95

Abstract

In this chapter, we outline the theoretical framework behind the computational methodology presented in this thesis. The generalities of classical mechanics and molecular dynamics simulations are summarized. The basis of the quantum chemical approximations to solve the Schrodinger equation for many-electron systems is introduced. The theoretical formalism of Density Functional Theory is described. The fundamentals of the multiconfigurational methods with complete and restrictive actives spaces are reviewed. And finally, the QM / MM method was introduced.

1. Fundamentals of Statistical Mechanics

Most computer simulations are based on the assumption that classical mechanics can be used to describe the motions of atoms and molecules. A quantum mechanical system can be found in different states. Those quantum states are limited for general purposes to quantum states that are eigenvectors of the Hamiltonian H of the system (energy eigenstates). For any such state i , $H_i=E_i$, where E_i is the energy of state i . For the systems of interest to statistical mechanics (systems with up to 10^{23} particles), the degeneracy of energy levels is very large. It is denoted by $\Omega(E,V,N)$ the number of eigenstates with energy E of a system with N particles in a volume V . The basic assumption of statistical mechanics states that a system with fixed N , V , and E is equally likely to be found in any of its $\Omega(E)$ eigenstates [1].

The $\ln\Omega$ of the total system is at a maximum when the thermal equilibrium is reached. This suggests that $\ln\Omega$ is somehow related to the thermodynamic entropy S . The second law of thermodynamics states that the entropy of a system N , V , and E is at its maximum when the system is in thermal equilibrium. Therefore, it is established a direct relationship between the entropy and the degeneracy of the system. The entropy is related to the density of states of a system with energy E , volume V , and number of particles N . The entropy S will be defined as follows[1]

$$S(N, V, E) = k_B \ln\Omega(N, V, E) \quad (1)$$

Where k_B is Boltzmann's constant. With this, it is seen that if all degenerate eigenstates of a quantum system are equally likely immediately implies that, in thermal equilibrium, the entropy of a composite system is at a maximum [1]. The thermodynamic definition of temperature is

$$\frac{1}{T} = \left(\frac{\partial S}{\partial E} \right)_{V,N} \quad (2)$$

Now it is supposed that there is a system (denoted by A) that is in thermal equilibrium with a large heat bath (system B). The total system is closed, the total energy E is defined by E_A+E_B , and system A is prepared in one

specific quantum state i with energy E_i . The probability P_i to find system A in state i determined by degeneracy can be defined as [1]:

$$P_i = \frac{\Omega_B(E - E_i)}{\sum_j \Omega_B(E - E_j)} \quad (3)$$

Expanding $\ln\Omega(E-E_i)$ around $E_i=0$,

$$P_i = \frac{\exp\left(-\frac{E_i}{k_B T}\right)}{\sum_j \exp\left(-\frac{E_j}{k_B T}\right)} \quad (4)$$

Equation 4 is known as the Boltzmann distribution for a system at temperature T . Knowledge of the energy distribution allows us to compute the average energy $\langle E \rangle$ of the system at the given temperature T [1]:

$$\langle E \rangle = \sum_i E_i P_i \quad (5)$$

$$= \frac{\sum_i E_i \exp\left(-\frac{E_i}{k_B T}\right)}{\sum_j \exp\left(-\frac{E_j}{k_B T}\right)}$$

$$= \frac{\partial \ln \sum_i \exp\left(-\frac{E_i}{k_B T}\right)}{\partial 1/k_B T}$$

$$= \frac{\partial \ln Q}{\partial 1/k_B T} \quad (6)$$

Where the partition function Q is defined in the last line of *Equation 6*. If this equation is compared with the thermodynamic relation

$$E = \frac{\partial F/T}{\partial 1/T} \quad (7)$$

Where A is the Helmholtz free energy, it is seen that A is related to the partition function Q [1]:

$$A = -k_B T \ln Q = -k_B T \ln \left(\sum_i \exp\left(-\frac{E_i}{k_B T}\right) \right) \quad (8)$$

The relation between the Helmholtz free energy and the partition function is often more convenient to use than the relation between $\ln\Omega$ and the entropy. As a consequence, *Equation 8* is the workhorse of equilibrium statistical mechanics [1]. Statistical mechanics deals with ensemble averages. For the canonical ensemble, in

which the temperature T and number of particles N are fixed, the equilibrium average of some quantity G is expressed in terms of phase-space integrals involving the potential energy $U(r_1, \dots, r_{Nm})$ [2],

$$\langle G \rangle = \frac{\int G(r_1, \dots, r_{Nm}) e^{-\beta U(r_1, \dots, r_{Nm})} dr_1 \dots r_{Nm}}{\int e^{-\beta U(r_1, \dots, r_{Nm})} dr_1 \dots r_{Nm}} \quad (9)$$

Where $(r_i, i= 1, \dots, N_m)$ are the coordinates and $\beta=1/k_B T$. This average corresponds to a series of measurements over an ensemble of independent systems. The ergodicity hypothesis relates the ensemble average to measurements carried out for a single equilibrium system during the course of its natural evolution. It is expected that both kinds of measurement should produce the same result. Molecular dynamics simulation follows the dynamics of a single system and produces averages of the form [2]:

$$\langle G \rangle = \frac{1}{M} \sum_{k=1}^M G_k(r_1, \dots, r_{Nm}) \quad (10)$$

over a series of M measurements made as the system evolves. Assuming that the sampling is sufficiently complete to capture the typical behaviour, the two kinds of averaging will be identical [2].

2. Molecular Dynamics Simulations

Molecular Dynamics (MD) is a technique for computing the equilibrium and transport properties of a many-body system. In this system, the nuclear motion of the constituent particles obeys the laws of classical mechanics which is an excellent approximation for a wide range of materials. In an MD simulation a sample is prepared, a model consisting of N particles is selected, and Newton's equations of motion for this system ($F=ma$) is solved until the properties of the system no longer change with time (known as an equilibration). After equilibration, the actual measure of the properties of interest is performed. MD simulations are in many aspects similar to real experiments since they follow a similar protocol and some of the most common mistakes are similar. Some of these errors are sample preparation, short time of measurement, the system undergoes an irreversible change during the experiment, or the measure does not represent what it is supposed to be [1].

To measure an observable quantity in a MD simulation, it is needed to first express this quantity as a function of the position and momenta of the particles in the system. The definition of temperature in MD makes use of the equilibration of energy over all the degrees of freedom that enter quadratically in the Hamiltonian of the system. The total kinetic energy of the system should be measured and then divided by the number of degrees of freedom N_f ($3N - 3$, for a system of N particles with fixed total momentum). As the total kinetic energy of a system fluctuates, so does the instantaneous temperature [1]:

$$T(t) = \sum_{i=1}^N \frac{m_i v_i^2(t)}{k_B N_f} \quad (11)$$

A MD program usually reads in the parameters that specify the conditions of the run (e.g., initial temperature, number of particles, density, time steep) and initialize the system (i.e., select initial positions and velocities). Then the program computes the forces on all particles and integrates Newton's equations of motion. This steep make up the core of the simulation. They are repeated until the time evolution of the system has been completed for the desired length of time. After completion of the central loop, the averages of measured quantities are computed and printed and stop [1]. Those steps will be described as follows.

2.2 Initialization

It is needed to assign initial positions and velocities to all particles in the system to start the simulation. The particle positions should be chosen compatible with the structure that it is aimed to simulate. The particles should not be positioned at positions that result in an appreciable overlap of the atomic or molecular cores for any reason. This is achieved by initially placing the particles on a cubic lattice. First, each particle is located on its lattice site, and then a value that is drawn from a uniform distribution is attributed to each velocity of each particle on its lattice. Subsequently, all velocities are shifted such that the total momentum is zero and then the resulting velocities are scaled to adjust the mean kinetic energy to the desired value. In thermal equilibrium, the following relation should hold [1].

$$\langle v_a^2 \rangle = \frac{k_B T}{m} \quad (12)$$

Where v_a is the a component of the velocity of a given particle. The instantaneous temperature $T(t)$ can be adjusted to match the desired temperature T by scaling all velocities with a factor $(T/T(t))^{1/2}$ [1]. The velocities themselves are not really used to solve Newton's equations of motion. Rather, the positions of all particles at the present (x) and previous (x_m) time steps, combined with the knowledge of the force (F) acting on the particles, to predict the positions at the next time are used. With the knowledge of conservation of linear momentum, a position x for a particle in a direction is approximated by $x_m(i) = x(i) - v(i) * dt$ [1].

2.2 Force Calculation

The calculation of the force acting on every particle is the most time-consuming part of almost all MD simulations. In classical MD, it is usually considered a model system with pairwise additive interactions, so it is needed to consider the contribution to the force on each particle due to all its neighbours. If only interactions between a particle and the nearest image of another particle are considered, this implies that, for a system of N particle, it is needed to evaluate $N * (N - 1)/2$ pair distances. This means that the time needed for the evaluation of forces scales as N^2 . However, there are methods to reduce the time needed to N [1]. As it

is presented further, a force field is a set of equations, known as potential functions, which are used to generate the potential energies and forces of a simulation system. They are not a part of the simulation method itself, and their selection depends on the system of study [3].

2.3 Integrating the Equations of motion

Newton's equations are numerically integrated usually by one of two methods: Verlet and LeapFrog algorithms. Both start with a Taylor expansion, but the first one makes it around the position and the second one make it around both the position and the velocity. The former algorithm is not only one of the simplest, but also usually the best [1].

2.3.1 Verlet algorithm

The Verlet algorithm only uses a numerical integration equation. The acceleration, actual and past positions are needed for the integration. Verlet algorithm does not use the velocity to compute the new position. However, it can be known using finite differences. Verlet algorithm starts with a Taylor expansion of the coordinate of a particle, around time t [1],

$$r(t + \Delta t) = r(t) + \frac{dr}{dt}(t)\Delta t + \frac{d^2r}{dt^2}(t) \frac{\Delta t^2}{2} \quad (13)$$

Similarly,

$$r(t - \Delta t) = r(t) - \frac{dr}{dt}(t)\Delta t + \frac{d^2r}{dt^2}(t) \frac{\Delta t^2}{2} \quad (14)$$

Adding these two equations,

$$r(t + \Delta t) + r(t - \Delta t) = 2r(t) + \Delta t^2 \frac{d^2r}{dt^2}(t) \quad (15)$$

Or

$$r(t + \Delta t) + r(t - \Delta t) = 2r(t) + \Delta t^2 \frac{f(t)}{m} \quad (16)$$

2.3.2 LeapFrog algorithm

Leapfrog algorithm is analogous to Verlet's, but it uses the velocity instead of the position. It uses two numerical integration equations, and the acceleration, past velocity and actual position must be known [1].

$$v(t + \Delta t) = v(t) + \frac{dv}{dt}(t) \frac{\Delta t}{2} + \frac{1}{2} \frac{d^2v}{dt^2}(t) \frac{\Delta t^2}{2} \quad (17)$$

Similarly,

$$v(t - \Delta t) = v(t) - \frac{dv}{dt}(t) \frac{\Delta t}{2} + \frac{1}{2} \frac{d^2v}{dt^2}(t) \frac{\Delta t^2}{2} \quad (18)$$

Subtracting these two equations,

$$v(t + \Delta t) + v(t - \Delta t) = \Delta t \frac{dv}{dt}(t) \quad (19)$$

or

$$r(t + \Delta t) + r(t - \Delta t) = \Delta t \frac{f(t)}{m} \quad (20)$$

2.4 Commonly used software for MD simulations

There are different MD software differing on the possibility of modelling different molecular systems, solvent environment, flexibility, computation efficiency, implementation, tools or modules for trajectory analyses, among many others. Some common molecular dynamics engines that can satisfy the above criteria are LAMMPS, CHARMM, GROMACS, NAMD, AMBER, DESMOND, and TINKER, among others. In this work, MD simulations were carried out using the GROMACS [4-8] package versions 5.1.2 and 2016.3.

2.4.1 GROMACS

The Groningen Machine for Chemical Simulations (GROMACS) is primarily designed for biochemical molecules like proteins, lipids and nucleic acids that have a lot of complicated bonded interactions. GROMACS is extremely fast at calculating the non-bonded interactions, that usually dominate simulations. Therefore, it has been used for research on non-biological systems as polymers. The entire GROMACS package is available under the GNU General Public License and the source code and selected set of binary packages are available on the home page [4-8]. A Molecular Dynamics simulation in GROMACS usually comprises a set of steps that involve the preparation, minimization and equilibration of the system and finally a molecular dynamics production. Those steps are described in detail in the computational details of each chapter throughout this manuscript.

2.5 Force Field

As mentioned before, a force field is a set of equations known as potential functions, which are used to generate the potential energies and forces of a simulation system. They are not a part of the simulation method

itself, and their selection depends on the system of study. The potential functions can be subdivided into three parts: non-bonded, bonded and restraints interactions [3].

2.5.1 Non-bonded interactions

The non-bonded interactions contain a repulsion term, a dispersion term, and a Coulomb term. The repulsion and dispersion term are usually combined in the Lennard-Jones (6-12) interaction. In addition, partially charged atoms act through the Coulomb term. These interactions are computed on the basis of a neighbour list (a list of non-bonded atoms within a certain radius), in which exclusions are already removed. Non-bonded interactions are pair-additive and centrosymmetric [3]:

$$U_{NB}(r_1, \dots, r_N) = \sum_{i < j} V_{ij}(r_{ij}) \quad (21)$$

$$F_i = - \sum_j \frac{dU_{ij}(r_{ij})}{dr_{ij}} \frac{\mathbf{r}_{ij}}{r_{ij}} = -F_j \quad (22)$$

Lennard-Jones potential

The Lennard-Jones (LJ) potential U_{LJ} between two atoms is:

$$U_{LJ}(r_{ij}) = \frac{C_{ij}^{12}}{r_{ij}^{12}} - \frac{C_{ij}^6}{r_{ij}^6} \quad (23)$$

The parameters C_{ij}^{12} and C_{ij}^6 depend on a pair of atom types; consequently, they are taken from a matrix of LJ parameters. The force derived from this potential is [3]:

$$F_i(r_{ij}) = \left(12 \frac{C_{ij}^{12}}{r_{ij}^{13}} - 6 \frac{C_{ij}^6}{r_{ij}^7} \right) \frac{\mathbf{r}_{ij}}{r_{ij}} \quad (23)$$

Coulomb interaction

The Coulomb interaction between [3] two charged particles is given by [3]:

$$U_c(r_{ij}) = f \frac{q_i q_j}{\epsilon_r r_{ij}} \quad (24)$$

The force derived from this potential is [3]:

$$F_i(r_{ij}) = f \frac{q_i q_j}{\epsilon_r r_{ij}^2} \frac{\mathbf{r}_{ij}}{r_{ij}} \quad (25)$$

Where $f = \frac{1}{4\pi\epsilon_0} = 138.935$

2.5.2 Bonded interactions

Bonded interactions are based on a fixed list of atoms. They include not only pair interactions but also include 3- and 4-body interactions. There are bond stretching, bond angle and dihedral angle for 2-, 3- and 4-body interactions, respectively. The improper dihedral is a special type of dihedral interaction in which the atoms are forced to remain in a plane. It is also used to prevent transition to a configuration of opposite chirality (a mirror image) [3].

Bond stretching

The harmonic potential for the bond stretching between two covalently bonded atoms i and j is represented by [3]:

$$U_b(r_{ij}) = \frac{1}{2}k_{ijbond}(r_{ij} - b_{ij})^2 \quad (26)$$

The force derived from this potential is [3]:

$$F_i(r_{ij}) = k_{ijbond}(r_{ij} - b_{ij}) \frac{\mathbf{r}_{ij}}{r_{ij}} \quad (27)$$

Harmonic angle potential

The bond-angle vibration between a triplet of atoms i - j - k is also represented by a harmonic potential on the angle θ_{ijk} [3]:

$$U_a(\theta_{ijk}) = \frac{1}{2}k_{ijkangle}(\theta_{ijk} - \theta_{0ijk})^2 \quad (28)$$

The force equations are given by:

$$F_i = -\frac{dV_a(\theta_{ijk})}{dr_i} \quad (29)$$

$$F_k = -\frac{dV_a(\theta_{ijk})}{dr_k} \text{ where } \theta_{ijk} = \arccos \frac{(\mathbf{r}_{ij} \cdot \mathbf{r}_{kj})}{r_{ij}r_{kj}} \quad (30)$$

$$F_j = -F_i - F_k \quad (31)$$

Dihedral angle potential

The form of the dihedral angle potential follows a fourth-order polynomial [3]:

$$U_d(\theta_{ijk}) = \sum_{n=0}^4 C_n (\theta_{ijk} - \theta_{0ijk})^n \quad (32)$$

Improper dihedrals

Improper dihedrals are meant to keep planar groups (like aromatic rings) planar or to prevent molecules from flipping over their mirror images. The harmonic potential is the simplest of all, and it is discontinuous [3].

$$U_{id}(\varepsilon_{ijkl}) = \frac{1}{2} k_\varepsilon (\varepsilon_{ijkl} - \varepsilon_0)^2 \quad (33)$$

Proper dihedrals

A proper dihedral consists of four atoms, joined linearly through bonds. These dihedrals can be represented by a periodic potential. A periodic dihedral is represented by a cosine function of the form [3]

$$U_d(\varphi_{ijkl}) = k_\varphi (1 + \cos(n\varphi - \varphi_s)) \quad (34)$$

2.5.3 Restraints

Restraints must be imposed on the motion of the system, either to avoid extreme deviations or to include knowledge from experimental data. There are position, angle, dihedral, distance, and orientation restraints. They are not really part of the force field. Some of these restraints are used to restrain particles to fixed reference positions, to restrain the angles and dihedrals between particles and to add penalties to the potential when the distance/angle/dihedral between specified atoms exceeds a threshold value [3].

2.5.4 Commonly used Force Fields

A force field is built up from the combination of the set of equations included on the potential functions and of the parameters used in each set of equations. These parameters are usually derived from experiments and will depend on the type of atoms involved in the system [3]. There are different fields of forces differing in the expression of the potential energy and the parameters used. The most used set of force fields are AMBER[9], CHARMM [10-12], OPLS [13, 14], GROMOS [15, 16].

Throughout this work, we used the CHARMM36 and the OPLS force fields. The first was used to compute the potential energies of lignocellulose and the solvent molecules in MD using GROMACS, whereas the second was used during the QM/MM simulations (see below) of lignin and cellulose in the presence of ethanol and acetonitrile in deMon2q. Furthermore, we used the TIP3P and TIP4 models of water. The former was used with both Gromacs and deMon2k, whereas the second was only used with Gromacs.

All-atom MD simulations were carried out using the GROMACS [4-8] package versions 5.1.2 and 2016.3, along with the CHARMM36 additive force field [17, 18] and the CHARMM-compatible force field for lignin. These force fields enable the modelling of hexopyranose monosaccharides and linear, branched, or cyclic glycosidic-linked hexopyranose polysaccharides, both alone and in heterogeneous systems, along with other organic molecules and biomolecules including proteins, nucleic acids, and/or lipids. The lignin force field has been parametrized to model all the monomers and linkages available in the lignin polymers.

In this study, we consider the 3 and 4-sites Transferable Intermolecular Potential (TIP) for liquid water [19-21], in order to assess the potential that reproduces better the target properties. TIP3P is a 3-sites model, with charges located in the centre of mass of the atoms (Figure 2.1), whereas TIP4P is a 4-sites model with a delocalized charge (dummy atom M) for oxygen. TIP3P is widely used in simulation and, in addition, it has been selected for parameterization of the CHARMM force field. Thus, parameters for the solvent-solute interactions should have been optimized for this potential. However, this model has been previously studied by several authors and has been found that it poorly reproduces properties such as the density and structure of water, and for this reason, a more sophisticated potential for water may be useful. TIP4P has been parameterized to better reproduce the electrostatic distribution around the water molecule. Thus, liquid density and structure at 298 K and 1 bar reproduces better than some 3-sites models, including the TIP3P. However, this is achieved by adding a 4th site to the model which makes it more time-consuming. Nevertheless, during the GROMACS implementation has been found that it is only 7% slower than the other 3-sites potential [22]. The parameters describing each one of the models are presented in Table 1.

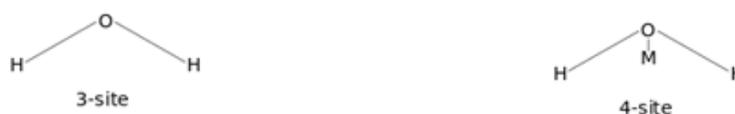


Figure 2.1. Representation of water models

Table 1. Parameters for TIP3P and TIP4P models of water. $r(\text{OH})$ and HOH are the bond distance and bond angle, respectively, in a water molecule, A and B are the constants in the Lennard-Jones potential, and q denotes the atomic charge parameters.

TIP3P		TIP4P	
$r(\text{OH}), \text{\AA}$	0.9572	$r(\text{OH}), \text{\AA}$	0.9572
$\text{HOH}, \text{degrees}$	104.52	$\text{HOH}, \text{degrees}$	104.52
		$r(\text{OM}), \text{\AA}$	0.15
$A \times 10^{-3}, \text{kcal } \text{\AA}^{12}/\text{mol}$	582.0	$A \times 10^{-3}, \text{kcal } \text{\AA}^{12}/\text{mol}$	600.0
$B, \text{kcal } \text{\AA}^6/\text{mol}$	595.0	$B, \text{kcal } \text{\AA}^6/\text{mol}$	610.0
$q(\text{O})$	-0.834	$q(\text{M})$	-1.04
$q(\text{H})$	+0.417	$q(\text{H})$	+0.52

On the other hand, the OPLS force field of Jorgensen [13] has a simple form and have been parameterized directly to reproduce experimental thermodynamic and structural data on fluids. The parameters were obtained and tested primarily in conjunction with Monte Carlo statistical mechanics simulations of organic liquids and aqueous solutions of organic ions representative of subunits in the side chains and backbones of proteins. Bond stretch, angle bend, and torsional terms have been adopted from the AMBER united atom force field [14].

3. Quantum Chemistry (QC) background

The main objective of quantum chemistry is to solve the time-independent non-relativistic Schrödinger equation for many-electron systems [23-28]:

$$\mathbf{H}\Psi = E\Psi \quad (35)$$

$$\hat{H}\Psi_i(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N, \vec{R}_1, \vec{R}_2, \dots, \vec{R}_M) = E_i \Psi_i(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N, \vec{R}_1, \vec{R}_2, \dots, \vec{R}_M)$$

where \hat{H} is the Hamilton operator for a molecular system of M nuclei with a radius-vector \vec{R}_M and N electrons with radius-vectors \vec{x}_N in the absence of magnetic or electric field. The Hamiltonian, expressed in atomic units, is of the type:

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 - \frac{1}{2} \sum_{A=1}^M \frac{1}{M_A} \nabla_A^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}} \quad (36)$$

In the above formula, A and B label the atomic nuclei, and i and j label the electrons. First, two terms are the kinetic energy operators of the electrons and the nuclei, where the Laplace operator is of the following type

$\nabla_{i(A)}^2 = \frac{\partial^2}{\partial x_{i(A)}^2} + \frac{\partial^2}{\partial y_{i(A)}^2} + \frac{\partial^2}{\partial z_{i(A)}^2}$, and M_A Z_A are the mass and the charge at nucleus A, respectively. The other

three terms are the potential energy operators, which describe the Coulomb nucleus-electron attraction potential, the repulsion potential between each couple (i,j) electrons and the repulsion between each (A, B) nuclei, respectively.

The solution of the Schrödinger equation gives the energy and wave function, from which a large variety of properties can be computed.

The mass, M_A of a nucleus is nearly 2000 larger than that of an electron. This allows to consider the nuclei as static, with zero kinetic energy and constant potential energy of the nucleus-nucleus repulsion, and to separate the variables of the nuclei and electrons. This is the well-known Principe of Born-Oppenheimer [29, 30]. Applying the Principe of Born-Oppenheimer allows to reduce the above Hamiltonian (Eq. 36) to an only electronic part:

$$\hat{H}_{elec} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} = \hat{T} + \hat{V}_{Ne} + \hat{V}_{ee} \quad (33)$$

Therefore, solving the electronic Schrödinger equation consists of finding the eigenvalues (the electron energy E_{elec}) and the eigenfunctions (the electron wave function $\hat{\Psi}_{elec}$) of the Hamiltonian \hat{H}_{elec} .

According to the Variational principle, the true solution will give us the minimum energy E_0 (the lowest possible energy), which corresponds to the system wave function Ψ_0 of this minimum energy state. This is mathematically expressed as follows:

$$E_0 = \min_{\Psi \rightarrow N} E[\Psi] = \min_{\Psi \rightarrow N} \langle \Psi | \hat{T} + \hat{V}_{Ne} + \hat{V}_{ee} | \Psi \rangle \quad (34)$$

The anti-symmetric wave function of an N-electron system expressed by the one-electron wave functions (spin-orbits) $\chi_i(\vec{x}_i)$ is represented by the Slater determinant (Φ_{SD}).

$$\Psi_0 \approx \Phi_{SD} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\vec{x}_1) \chi_2(\vec{x}_1) \dots \chi_N(\vec{x}_1) \\ \chi_1(\vec{x}_2) \chi_2(\vec{x}_2) \dots \chi_N(\vec{x}_2) \\ \vdots \\ \chi_1(\vec{x}_N) \chi_2(\vec{x}_N) \dots \chi_N(\vec{x}_N) \end{vmatrix} \quad (35)$$

Each spin-orbit is the product of the space wave function $\phi_i(\vec{r})$ and either the spin-function with spin $s = 1/2$, $\alpha(s)$, or that with spin $s = -1/2$, $\beta(s)$.

$$\chi(\vec{x}) = \phi(\vec{r}) \sigma(s), \quad \sigma = \alpha, \beta \quad (36)$$

An important property of these spin-functions is that they are normalized and orthogonal functions $\langle \alpha | \alpha \rangle = \langle \beta | \beta \rangle = 1$ и $\langle \alpha | \beta \rangle = \langle \beta | \alpha \rangle = 0$. Spin orbitals have the following physical meaning: $|\chi(\vec{x})|^2 d\vec{x}$ presents the probability of finding an electron with spin-function σ in a volume $d\vec{r}$. The normalization factor $(N!)^{-1/2}$ in eq. 2.6 ensures the satisfaction of the Φ_{SD} normalization condition:

$$\int \dots \int |\Psi(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N)|^2 d\vec{x}_1 d\vec{x}_2 \dots d\vec{x}_N = 1 \quad (36)$$

In the following part, we give a brief introduction to the single-determinant Hartree-Fock (HF) theory as the basic method for solving approximately the Schrödinger equation for a many-electron system.

The main goal is by applying the variational principle to find the ground-state Slater determinant. In other words, to obtain a set of spin-orbitals that produces the minimum energy for the N-electron system in its ground-state. To this aim, we vary the HF-spin orbitals $\{\chi_i\}$ while maintaining the conditions for orthogonal and normalized spin-orbitals up to the finding of a set of spin-orbitals producing the lowest possible energy:

$$E_{HF} = \min_{\Phi_{SD} \rightarrow N} E[\Phi_{SD}] \quad (37)$$

The Hartree-Fock energy is represented as:

$$E_{HF} \langle \Phi_{SD} | \hat{H} | \Phi_{SD} \rangle = \sum_i^N (i | \hat{h} | i) + \frac{1}{2} \sum_i^N \sum_j^N (ii | jj) - (ij | ji), \quad (38)$$

$$\text{where } (i | \hat{h} | i) = \int \chi_i^*(\vec{x}_1) \left\{ -\frac{1}{2} \nabla^2 - \sum_A^M \frac{Z_A}{r_{1A}} \right\} \chi_i(\vec{x}_1) d\vec{x}_1$$

is the kinetic plus nuclear-electron attraction energies. The two-electron interactions are described by the Coulomb ($ii|jj$) and exchange ($ij|ji$) integrals as given below with eq. 40).

$$(ii | jj) = \iint |\chi_i(\vec{x}_1)|^2 \frac{1}{r_{12}} |\chi_j(\vec{x}_2)|^2 d\vec{x}_1 d\vec{x}_2 \quad (39)$$

$$(ij | ji) = \iint \chi_i(\vec{x}_1) \chi_j^*(\vec{x}_1) \frac{1}{r_{12}} \chi_j(\vec{x}_2) \chi_i^*(\vec{x}_2) d\vec{x}_1 d\vec{x}_2$$

To satisfy the normalization conditions, $\{\chi_i\}$ a Lagrange multiplier ε_i is introduced in the initial set of equations:

$$\hat{f} \chi_i = \varepsilon_i \chi_i, \quad i = 1, 2, \dots, N \quad (40)$$

Equations 39 are called Hartree-Fock equations. They represent a system of N equations with ε_i , being the eigenvalues of the Fock operator \hat{f} . The physical meaning of ε_i is the orbital energy, and the Fock operator is the effective one-electron operator defined as:

$$\hat{f}_i = -\frac{1}{2} \nabla_i^2 - \sum_A^M \frac{Z_A}{r_{iA}} + V_{HF}(i) \quad (41)$$

The first two members are the kinetic and electron-nuclear Coulomb operators. $V_{HF}(i)$ is the so-called Hartree-Fock potential. This is the mean repulsion potential of electron i with the $N-1$ electrons in the N -electronic system.

In this way, the exact, but explicitly unknown bi-electron operator $\frac{1}{r_{ij}}$ in the Hamiltonian is approximated by the one-electron operator $V_{HF}(i)$, accounting in an averaged way for the bi-electron repulsion. V_{HF} consists of two terms:

$$V_{HF}(\vec{x}_1) = \sum_j^N (\hat{J}_j(\vec{x}_1) - \hat{K}_j(\vec{x}_1)) \quad (42)$$

The Coulomb operator is given by $\hat{J}_j(\vec{x}_1) = \int |\chi_j(\vec{x}_2)|^2 \frac{1}{r_{12}} d\vec{x}_2$, and the exchange operator by

$$\hat{K}_j(\vec{x}_1) \chi_i(\vec{x}_1) = \int \chi_j^*(\vec{x}_2) \frac{1}{r_{12}} \chi_i(\vec{x}_2) d\vec{x}_2 \chi_j(\vec{x}_1) \quad .$$

Numerically, the HF equations are solved iteratively. The procedures proposed by Roothaan [31, 32] and known as restricted HF (RHF) method for the closed-shell systems, is applied, and for the open-shell systems – the Pople-Nesbet approach, known as unrestricted HF method (UHF) is applied [33]:

$$FC = SC\varepsilon \quad (43)$$

$$\begin{aligned} F^\alpha C^\alpha &= SC^\alpha \varepsilon^\alpha \\ F^\beta C^\beta &= SC^\beta \varepsilon^\beta \end{aligned} \quad (44)$$

In the above equations, F is the Fock matrix, and S is the overlap matrix build from the atomic wave functions, called basis functions.

3.1 Density Functional Theory

3.1.1 Thomas-Fermi Model

The Density Functional Theory [34, 35] has its conceptual root in the semi-classical model of Thomas-Fermi, which was proposed for the electronic structure computations of many-electron system in 1927. In this model quantum statistical theory is employed to describe the kinetic energy of the electrons, whereas the potential energy due to the electron-electron repulsion and electron-nuclei attraction is computed classically:

$$\hat{H}_{elec} = \hat{V}_{Ne} + \hat{V}_{ee} + \hat{T}_K \quad (45)$$

In this model, the density, (ρ) , is a statistically averaged quantity and the model is derived by assuming that :

- 1) the electrons are homogeneously distributed in a 6-dimensional phase-space defined by the electron coordinates q_x, q_y, q_z and electron momenta $\vec{p}(p_x, p_y, p_z)$ in a (volume) cell h^3 ($h^3 = \int q_x q_y q_z dx dy dz$);
- 2) each electron moves in the mean-field of all the other electrons and the nuclei.

For the N-electronic system in the phase-space with a volume , the mean density can be easily computed as

$\rho = \frac{N}{\mathcal{G}_{e^{-n}}}$. The maximum momentum of these electrons will be that of the fastest electron (Fermi electron),

noted as \vec{p}_f . The momenta of all other electrons are lower than \vec{p}_f , $\vec{p} < \vec{p}_f$. Thus, the volume of the occupied energy levels in the momentum space is:

$$\mathcal{G}_{occ} = \frac{4}{3} \pi p_f^3 \quad (46)$$

The total volume of those states in the phase space will be thus

$$\mathcal{G}_T = \mathcal{G}_{occ} V = \frac{4}{3} \pi p_f^3 V \quad (47)$$

According to the indetermination principle of Heisenberg, the phase space can be divided into h^3 cells, and each cell could contain a maximum of only two electrons with opposite spin (Pauli principle). Then:

$$N = \frac{4}{3} \pi p_f^3 \frac{V}{h^3 / 2} = \frac{8}{3} \pi p_f^3 \frac{V}{h^3} \quad (48)$$

By replacing in eq. 48 N with $N = \rho V$ one obtains:

$$\rho = \frac{8}{3} \frac{\pi p_f^3}{h^3} \frac{V}{V} = c_1 p_f^3 \quad (49)$$

By defining as a constant c_1 , $c_1 = \frac{8}{3} \frac{\pi}{h^3}$, it follows that:

$$p_f = c_1^{1/3} \rho^{1/3} \quad (50)$$

The total energy, $\mu = \frac{mV^2}{2} + V_p$, of all moving electrons in this space will be thus obtained after some algebra in the following way:

$$\mu = \frac{1}{2m} \left(\frac{3h^3}{8\pi} \rho \right)^{2/3} + V_p = \frac{1}{2m} \left(\frac{3h^3}{8\pi} \right)^{2/3} \rho^{2/3} + V_p \quad (51)$$

This shows that the kinetic energy is proportional only to $\rho^{2/3}$. From eq. 51 follows that the total energy μ can be analogously interpreted as a chemical potential.

The energy of any electron in position \mathbf{r} and with a momentum in an interval $(p, p + dp)$ is:

$$I_r dp \begin{cases} = \frac{4\pi p^2 dp}{4/3 \pi p_f^3(\vec{r})}, p < p_f(\vec{r}) \\ = 0, p \geq p_f(\vec{r}) \end{cases}, \text{ in this way the kinetic energy density } t \text{ (the kinetic energy in a unit}$$

volume) can be obtained as follows:

$$\begin{aligned} t &= \int_0^{p_f} \rho(\vec{r}) \frac{p^2}{2m} \frac{3p^2}{p_f^3} dp = \rho(\vec{r}) \frac{3}{2m} \frac{p_f^{5/2}}{5} \frac{1}{p_f^3} = \\ &= \rho(\vec{r}) \frac{3}{10m} \left(\frac{3h^3}{8\pi} \right)^{2/3} \rho^{2/3} = \frac{3h^2}{10m} \left(\frac{3}{8\pi} \right)^{2/3} \rho^{5/3} = c_2 [\rho]^{5/3} \end{aligned} \quad (52)$$

And the total kinetic energy thus becomes a function of the density defined by:

$$\hat{T}_K = c_2 \int [\rho(\vec{r})]^{5/3} d\vec{r} \quad (53)$$

3.1.2 Hohenberg-Kohn theorems

The idea behind DFT is to show that the ground state observables of a many-particle quantum system can be described exactly by the total density. It should, however, be noted already now, that although the theory in itself is exact, approximations will still enter in actual calculation due to our inability to find an exact functional which describes any system. The first actual formulation of a density functional theory was derived by Hohenberg and Kohn in 1964 [36]. It is based on two theorems – the Hohenberg-Kohn theorems.

First theorem: The first theorem states that the external potential $\mathcal{G}(\vec{r})$ of any many-electron system in its ground state Ψ is uniquely determined by the true electron density $\rho(\vec{r})$, up to an additive constant. The first

Hohenberg-Kohn theorem provides the fundamental theoretical bases for all modern variants of Density Functional Theory. It assures that the total energy depends explicitly on the electron density as it was suggested in the Thomas-Fermi model.

Second theorem: The total energy $E_g[\rho]$ of an N-electron system is the minimum energy of the system for the true electron density. This implies that there exists a unique map between the external potential \mathcal{G} and the ground state density. This implies that all the energies, including the total energy, is a *function* of the density, $E_g[\rho]$. The density determines the Hamiltonian, and thereby, the wavefunction.

3.1.3 Kohn-Sham Self Consistent Equations

The ground state energy can be obtained as the minimum of the energy functional:

$$E[\rho] = \int \rho(\mathbf{r})\mathcal{G}(\mathbf{r})d\mathbf{r} + F[\rho] \quad (54)$$

where

$$F[\rho] = T[\rho] + V_{ee}[\rho] \quad (55)$$

The ground-state electron density is the density for which $E[\rho]$ has a minimum and satisfies the Euler' equation:

$$\mu = \mathcal{G}(\mathbf{r}) + \frac{\delta F[\rho]}{\delta \rho(\mathbf{r})} \quad (56)$$

with μ being the Lagrange multiplier, introduced because the energy minimization is carried out under the condition that the electron density is constant in the system

$$\int \rho(\mathbf{r})d\mathbf{r} = N \quad (57)$$

Among all the possible solutions of eq. 56 there is a unique solution that minimizes $E[\rho]$.

In the Thomas-Fermi model, the functionals $T[\rho]$ и $V_{ee}[\rho]$ are introduced as density functionals, but with a classical treatment of $V_{ee}[\rho]$. In 1965 Kohn and Sham (KS) [37] suggested an original approach to obtain the kinetic energy functional by introducing the so-called “non-interacting” system. For such a system, the kinetic energy is known:

$$T_s[\rho] = \sum_i^N \langle \psi_i | -1/2\nabla^2 | \psi_i \rangle \quad (58)$$

and

$$\rho(\mathbf{r}) = \sum_i^N \sum_s |\psi_i(\mathbf{r}, s)|^2 \quad (59)$$

The kinetic energy of the interacting systems is added to the equation in the way:

$$F[\rho] = T_s[\rho] + J[\rho] + E_{xc}[\rho] \quad (60)$$

where:

$$E_{xc}[\rho] \equiv T[\rho] - T_s[\rho] + V_{ee}[\rho] - J[\rho]$$

$E_{xc}[\rho]$ is called exchange-correlation energy and contains the difference between the unknown exact kinetic energy and the known kinetic energy for the non-interacting system $T[\rho] - T_s[\rho]$. This difference is expected to be very small compared to the other energy terms. Analogously, the difference $V_{ee}[\rho] - J[\rho]$ accounts also for the interaction in the systems due to the coulomb and Pauli repulsion.

The Euler equation can be thus re-written as:

$$\mu = \mathcal{G}_{eff}(\mathbf{r}) + \frac{\delta T_s[\rho]}{\delta \rho(\mathbf{r})} \quad (61)$$

Where the effective Kohn-sham potential is defined as follows:

$$\mathcal{G}_{eff}(\mathbf{r}) = \mathcal{G}(\mathbf{r}) + \frac{\delta J[\rho]}{\delta \rho(\mathbf{r})} + \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})} = \mathcal{G}(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \mathcal{G}_{xc}(\mathbf{r}) \quad (62)$$

and

$$\mathcal{G}_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})} \quad (63)$$

In the canonical form, the Kohn-Sham equations can be written as:

$$\left[-1/2\nabla^2 + v_{eff}\right]\psi_i = \varepsilon_i\psi_i \quad (64)$$

$$v_{eff}(\mathbf{r}) = v(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' + v_{xc}(\mathbf{r}) \quad (65)$$

$$\rho(\mathbf{r}) = \sum_i^N \sum_s |\psi_i(\mathbf{r}, s)|^2 \quad (66)$$

The above equations are non-linear equations and are solved iteratively.

The total KS energy is obtained as:

$$E = \sum_i^N \varepsilon_i - \frac{1}{2} \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + E_{xc}[\rho] - \int v_{xc}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \quad (67)$$

where :

$$\sum_i^N \varepsilon_i = \sum_i^N \langle \psi_i | -1/2\nabla^2 + v_{eff}(\mathbf{r}) | \psi_i \rangle = T_s[\rho] + \int v_{eff}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \quad (68)$$

Analogously to the HF SCF equations, in the KS equations the spin-orbitals are built as a linear combination of the atomic wave functions (basis functions):

$$\phi_i = \sum_{\alpha}^M c_{\alpha i} \chi_{\alpha} \quad (69)$$

The iterative computations proceed by solving the equations sets below

$$\mathbf{h}_{KS} \mathbf{C} = \mathbf{S} \mathbf{C} \boldsymbol{\varepsilon}$$

$$h_{\alpha\beta} = \langle \chi_{\alpha} | \mathbf{h}_{KS} | \chi_{\beta} \rangle$$

$$S_{\alpha\beta} = \langle \chi_{\alpha} | \chi_{\beta} \rangle$$

$$\mathbf{h}_{KS} = -\frac{1}{2}\nabla^2 + V_{ne} + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' + V_{xc}$$

The \mathbf{h}_{KS} operator is analogous to the HF one and is called Kohn-Sham operator.

3.1.4 Approximations for the exchange-correlation formula

Although DFT is an exact theory, the explicit mathematical expression of the exchange-correlation functional for interacting electron systems is not known. There exists however approximations, allowing to build computational tools based on the DFT.

Usually, the exchange-correlation energy, E_{XC} , is written as a sum of the exchange energy, E_X , and the correlation energy, E_C , analogously as the exchange-correlation hole. The exchange-correlation energy is also represented by the exchange (ε_X) and correlation (ε_C) energy densities:

$$E_{XC}[\rho] = E_X[\rho] + E_C[\rho] = \int \rho(r) \varepsilon_X[\rho(r)] dr + \int \rho(r) \varepsilon_C[\rho(r)] dr \quad (70)$$

The exchange (correlation) energy is defined as the sum of the exchange energy of the electrons with spin α and that of the electrons with spin β :

$$\begin{aligned} E_X[\rho] &= E_X^\alpha[\rho_\alpha] + E_X^\beta[\rho_\beta] \\ E_C[\rho] &= E_C^{\alpha\alpha}[\rho_\alpha] + E_C^{\beta\beta}[\rho_\beta] + E_C^{\alpha\beta}[\rho_\alpha, \rho_\beta] \end{aligned} \quad (71)$$

In the **local density approximation** (LDA) [38, 39] the density is assumed to be a density of uniform homogeneous electron gas. The Dirak formula [40] is used to describe the exchange energy:

$$\begin{aligned} E_X^{LDA}[\rho] &= -C_X \int \rho^{4/3}(r) dr \\ \varepsilon_X^{LDA} &= -C_X \rho^{1/3} \end{aligned} \quad (72)$$

In the case of spin polarization, this approximation is:

$$\begin{aligned} E_X^{LSDA}[\rho] &= -2^{1/3} C_X \int (\rho_\alpha^{4/3} + \rho_\beta^{4/3}) dr \\ \varepsilon_X^{LSDA} &= -2^{1/2} C_X [\rho_\alpha^{1/3} + \rho_\beta^{1/3}] \end{aligned} \quad (73)$$

The X_α method of Slater [41] includes the correlation energy by introducing the appropriately fitted parameter:

$$\varepsilon_{X_\alpha} = -\frac{3}{2} \alpha C_X \rho^{1/3} \quad (74)$$

The LDA functional mostly implement in the DFT based computer programs are based on the general formula:

$$\varepsilon_{xc}(\rho) = \frac{A(\text{const.})}{r_s} - B(\text{const.})h_x(\vec{r}_1, \vec{r}_2) \frac{r_s}{C} \quad (75)$$

Where r_s is the Wigner-Seitz radius

$$r_s = \left[\frac{3}{4\pi} \frac{1}{\rho} \right]^{1/3} \text{ and } h_x \text{ is the exchange hole.}$$

For the spin-polarised case the density is given as:

$$\xi = \frac{\rho_\alpha(\vec{r}) - \rho_\beta(\vec{r})}{\rho(\vec{r})}$$

The **generalized gradient approximations** methods (GGA) [42, 43] includes a correction as a density gradient in order to represent better the inhomogeneous electron density. The general formula is:

$$E_{xc}(\rho) = \int \varepsilon_{xc}(\rho) \rho(\vec{r}) d\vec{r} + \int B_{xc}(\rho) |\nabla \rho|^2 d\vec{r} \quad (76)$$

B_{xc} is usually computed within the random phase approximations

$$B_{xc}(\rho) = \frac{C(r_s)}{\rho^{4/3}}$$

Many different GGA approximations have been so far proposed that have been parameterised for specific property computations.

For a better description of the exchange part, an appropriate combination between the exact Hartree-Fock exchange and GGA type of the correlation energy have been proposed. These functionals are called **hybrid exchange-correlation functionals**. It was firstly introduced by Becke [44, 45], who developed a formula including a portion of the exact HF exchange energy and several approximations of the exchange, and of the correlation energies as shown below:

$$E_{xc}^{B3LYP} = (1-a)E_x^{LSDA} + aE_x^{exact} + b\Delta E_x^{B88} + (1-c)E_c^{LSDA} + cE_c^{LYP} \quad (77)$$

Various improvements to the above formula 76 and 77 have been proposed. For example, in the meta-GGA functionals, the density Laplacian and the kinetic energy dependence (the orbital energy Laplacian) are included in addition to the energy gradients. Functionals derived in the frame of the optimized effective

potential formalism have also shown promising. For a recent review, we refer to the work of S. Kummel and L. Kronik [46].

3.1.5 Advantages and drawbacks of DFT

The DFT theory has become the method of choice in many solid-state and molecular electronic structure calculations. Its original formulation for the ground-state electron system in the absence of electric and magnetic field have been extended. Electron structure computations of higher electron energy states can be now performed with sufficient accuracy based on the Time-Dependent Density Functional Formalism (TDDFT) [47]. Methodologies to include magnetic property calculations have been also developed [48].

The main drawbacks lie in the description of the electron density in systems with inhomogeneous and non-local density. This includes a description of weakly bound non-covalently interacting systems. Various developments in this direction by separating the interaction range and applying a proper combination between the DFT based methodologies and post-Hartree Fock methods have been very recently proposed, and the computer realization of this approaches are in progress.

3.2 Multiconfigurational Methods

Multiconfigurational methods generate qualitatively correct reference state of molecules in cases where the Hartree-Fock and density functional theory are not adequate. The Configurational Interaction (CI) Theory is a post-HF method that provides the correct electronic structure of molecules, not only at the equilibrium and the dissociation limit, but for the full potential energy surface (PES). This theory has been applied successfully to many chemical applications, such as spectroscopy, photochemistry and excited states, among others [49]. The calculation of electronic excited states (in the scope of this work) is typically a multiconfigurational problem, and therefore it should preferably be treated with multiconfigurational methods such as CASSCF and CASPT2.

3.2.1 The Multiconfigurational (MC) Self-Consistent Field (SCF) method

One of the different variants of the CI theory is the multiconfigurational self-consistent field (MCSCF) methods. Here, the wave function comprises several electronic configurations in order to qualitatively describe situations where the electrons undergo substantial arrangements [50]. The MCSCF method [51] involves the simultaneous optimization of both the CI vector and the molecular orbitals (MO) coefficients. The coefficients from a CI calculation which comprises all single excitations of the reference MC function are used to construct a unitary rotation of the orbitals in each iteration [50]. Then, the MOs are optimized based on a linear combination of electronic configurations [49].

3.2.2 Complete active space self-consistent field (CASSCF)

In the CASSCF method [50] the number of determinants or CSFs used in the expansion of the CI vector are defined by dividing the orbitals into three subspaces (Fig. 2.2). In the first subspace, termed as inactive space, all orbitals are doubly occupied. The second subspace is known as the active space, and within this orbital space, a full-CI expansion is considered. The electrons and orbitals included in the active space are those that contribute most to the multireference character of the particular system of study. The occupation number of the orbitals in the active space will be a noninteger number between 0 and 2. Finally, the third subspace, known as the virtual space, consists of orbitals that are kept unoccupied. This method decreases the number of determinants or CSFs in the CI expansion since it limits the MC wave function within a specific subset of electrons and orbitals rather than on all electrons in all orbitals. The number of electrons and orbitals that can be included in the active space of a CASSCF calculation, however, is limited since the CI expansion increases exponentially with respect to the number of electrons and orbitals. The current limitations of the conventional CASSCF implementations are 18 active electrons in 18 orbitals. [49]

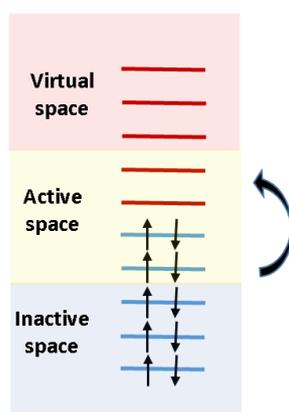


Figure 2.2. Schematic representation of the three orbital subspaces of a CASSCF set of orbitals. The arrow represents the electron excitations from the occupied orbitals to the empty orbitals in the active space.

3.2.3 Restricted Active Space (RASSCF)

The number of CSFs or Slater determinants and, thus, the computational cost quickly increases with the number of active orbitals. Therefore, it may be desirable to use a smaller set of CSFs. The RASSCF method [52] is an extension of the CASSCF formalism, where the inactive orbitals are doubly occupied in all configuration, and active orbitals are partitioned into three subspaces (Fig. 2.3). The subspace RAS1 contains orbitals that are doubly occupied up to a given maximum number of holes; in RAS2, all possible distributions of electrons are allowed, as in the CASSCF orbitals; and RAS3 contains orbitals that are unoccupied except up to a given maximum number of electrons.

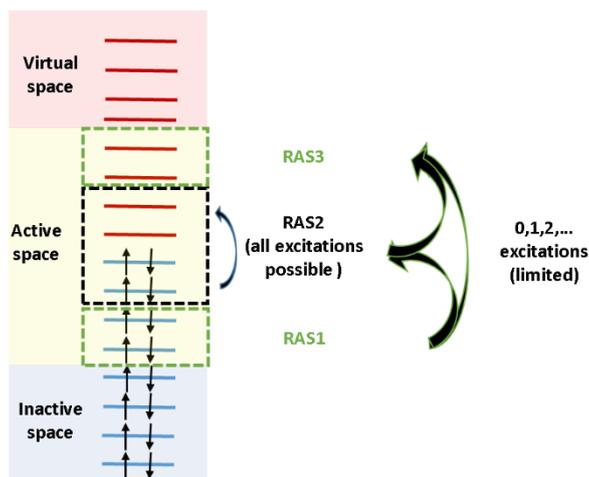


Figure 2.3 Schematic representation of the active spaces (RAS1, RAS2, and RAS3) used in RASSCF calculations. Curved arrows denote excitations from occupied to virtual orbitals.

3.3.4 Complete active space perturbation theory CASPT2

CASSCF wave functions are typically good enough, but this is not the case for electronic energies, and the dynamic correlation effects have to be included. The CASPT2 [53, 54] method uses as reference the states described with the CASSCF or RASSCF wave function and applies the second-order multiconfigurational perturbation theory to compute the correlation energy. The implementation and theory of the method can be further revised in ref. [55]

3.3 QM program packages used in this work

3.3.1 *deMon2k*

deMon2k is a program specialized in Kohn-Sham^[56] Density Functional Theory (DFT) within the Auxiliary DFT framework [57]. The latter permits remarkably fast evaluation of energies, potentials and properties. It is, therefore, a very promising basis for conducting hybrid QM/MM simulations with DFT as the electronic-structure method. Details of the program and implementation can be found in ref [58, 59].

3.3.2 *MOLCAS*

Molcas is an ab initio quantum chemistry software package able to treat general electronic structures for molecules consisting of atoms from most of the periodic table. Thus, the primary focus of the package is on multiconfigurational methods with applications typically connected to the treatment of highly degenerate states. Details of the program and implementation can be found in refs. [60-62].

4. *QM/MM Hybrid methods*

The size and complexity of a typical biomolecule, but also the time scale that must be reached make the ab initio calculations of biomolecules is difficult to access and make the use of molecular mechanics a necessity. However, molecular mechanics methods, heavily parameterized, do not allow a proper study of phenomena involving significant electronic rearrangement. The idea proposed by Warshel and Levitt [63] is to combine the QM method and MM method. This hybrid approach has the abbreviation QM/MM and has been promising in many fields spanning biomolecules, inorganic / organometallic systems, solids and the study of processes solvent medium 86-93 [64-67]

In the QM / MM methods, this involves dividing the modelled system into two parts: a precisely described region (QM method) and a region described with conventional methods (MM) that is less computationally expensive. The region described in a quantum way corresponds to the region where the chemical reactions take place (formation, breaking of bonds, excited state, etc.) while regions which undergo little chemical change (typically solvents) are described in a conventional manner thus allowing a saving of computing time very important. A third region can be used to describe the border area. The spatial division of the system into sub-regions is justified by the fact that most chemical reactions have a very local character.

The potential energy of the quantum region and the classical region are calculated in the usual way, but the total energy of the system can not be limited to the sum of the energies of these two regions since the latter interact together. We must, therefore, deal with the interaction between these two regions and this challenge is even more difficult if the border between these two regions intercepts a chemical bond. The border region may then contain additional atoms that do not exist in the base model. The QM / MM energy of a system can be evaluated according to two major schemes: additive schemes and subtractive schemes.

4.1 QM/MM Energy: subtractive scheme

In subtractive schemes, QM / MM energy is obtained in three steps. First, the energy of the total system is evaluated at the MM level. Then the QM energy of the QM region is added. Finally, the MM energy of the QM region is subtracted from the total energy obtained in order to correct the fact that the energy of the QM region is counted twice:

$$E_{QM/MM} = E_{MM}(QM + MM) + E_{QM}(QM) - E_{MM}(QM) \quad (77)$$

The main advantage of this scheme is the lack of communication between the QM and MM routines making the implementation relatively simple. But in return, a field of force is required for the part QM, but this force field is not always existing. Moreover, this method does not simulate the polarization of the electronic cloud in the QM part by the environment MM.

4.2 QM/MM Energy: additive scheme

In additive schemes, the total energy of the system is the sum of the QM energy of the QM region, the MM energy of the MM region, and a QM / MM coupling term. The interaction between region QM and region MM, at the origin of this term of coupling, can be described with different levels of sophistication:

$$E_{QM/MM} = E_{MM}(MM) + E_{QM}(QM) + E_{QM-MM}(QM + MM) \quad (78)$$

4.3 Mechanical inclusion

In this approach, all the interactions between the atoms of the region QM and MM are described by the force field: the bonds between atoms QM and MM are modelled by a harmonic potential, the non-binding interactions are described by a potential of Lennard-Jones etc. With this approach, the position of the atoms QM is directly influenced by the environment MM, but the electronic cloud QM is not polarized by this same environment.

To calculate the electrostatic interactions between the subsystems, it is possible either to attribute the charges from the force field to the QM atoms or to recalculate the charges of the QM atoms at each simulation stage.

Concerning the van der Waals interactions, these are simply described by the force field, van der Waals parameters are therefore necessary for the active region (the QM atoms). When they do not exist, these parameters can be borrowed from similar atoms. And even when they exist, the state of the QM atoms may change during the reaction, and one may wonder whether these parameters should not be recalculated. In practice, the Lennard Jones parameters are usually not updated during the calculation, but this probably does not involve big problems since the Lennard Jones potential is a short-range function and only near-border atoms QM / MM are important. Riccardi et al.⁹⁴ show that the van der Waals parameters have almost no influence on the thermodynamic parameters calculated from QM / MM simulation but that they slightly influence certain geometrical details around the QM region.

4.4 Electrostatic inclusion

A first improvement of the mechanical inclusion is to include the polarization effects of the electron density in the QM part by the MM [68] environment by incorporating, for example, the point electric charges of the MM cores. In the electrostatic inclusion scheme, the electrostatic interactions between the two regions are considered during the ab-initio calculation. The charged MM atoms are considered in the QM Hamiltonian as a mono-electronic operator:

$$H_i^{QM-MM} = H_i^{QM} - \sum_A^M \frac{e^2 Q_A}{4\pi\epsilon_0 |r_i - R_A|} \quad (79)$$

r_i et R_A are the positions of the electron i and the atom MM A of partial charge Q_A . H_i^{QM} is the original mono-electronic operator for kinetic energy and nuclear attraction energy.

In this approach, the MM atoms polarize the electron density according to their partial loads Q_A from the force field. However, these partial loads are set to give correct MM simulations rather than a physically correct load distribution. One can then wonder if the polarization induced by these loads MM is realistic or not. Electrostatic inclusion is, nevertheless, the most popular method.

4.5 Polarizable inclusion

The next step in sophistication is to include the polarizability of the MM atoms allowing the polarization of the two regions (QM and MM). For this, different approaches have been developed [69, 70], but there are still no well-established polarizable force fields for biomolecules. Moreover, the inclusion of the polarization requires the calculation of the polarization of the region MM at each step of the self-consistent iterative calculations QM. In addition, the polarization itself is determined in a self-consistent manner, making the calculations particularly heavy.

4.6 Saturation of broken bonds at the QM / MM boundary

If the QM and MM subsystems are connected by chemical bonds, particular attention should be paid when evaluating the QM energy. Indeed, if the boundary between region QM and MM cuts a chemical bond, it would leave in the QM region one or more unbound electrons which are actually bound with MM atoms. The simplest solution to this problem is to introduce a monovalent atom at an appropriate position along the cut bond and which would then be covalently bound to the QM atom whose bond has been broken. Hydrogen can fulfil this role, but it is not forbidden to use a group such as the methyl group. A popular alternative to this solution is to replace the chemical bond with saturated orbitals [71, 72].

4.7 QM/MM program packages used in this work

4.7.1 *deMon2k*

deMon2k is a program specialized in Kohn-Sham ^[56] Density Functional Theory (DFT) within the Auxiliary DFT framework [57-59] (*vide supra*). *deMon2k* allows QM/MM simulations with either non-polarizable or polarizable force fields and includes a polarizable continuum model for remote environments. Link atoms and capping potentials are incorporated to deal with QM/MM frontiers that cut covalent bonds. Both Born-Oppenheimer molecular dynamics (MD) simulations and Ehrenfest non-adiabatic MD simulations can be performed with this software. *deMon2k* also provides one of the very few implementations for conducting attosecond electron dynamics within polarizable MM environments. Further details can be found in ref. [73].

Molecular dynamics simulations at QM/MM level (BOMD/MD) with demon2k

Born Oppenheimer Molecular Dynamics (BOMD)/MD simulations can be useful to gain a better understanding of the stability solutes in a solvent medium [74, 75] [76]. In this hybrid approach, the solute of interest is included in the QM layer, whereas the solvents atoms are treated classically with the OPLS-AA force field [13]. A BOMD step in deMon2k consists of solving the time-independent electronic Schrödinger equation to get the forces on the nuclei, followed by a classical velocity Verlet [77, 78] propagation of the nuclei according to those forces. For each BOMD step, the step time, molecular energies, atomic coordinates, and velocities are written to the trajectory file. deMon2k records the time and the instantaneous temperature as well as the instantaneous kinetic, potential, and total energies. It also records averages of temperature and (total) energy. Furthermore, the mechanical interaction between the QM and MM regions is expressed by a Lennard-Jones potential. Therefore, Lennard-Jones parameters must be assigned to the QM atoms, and they are taken from the force field. These atom types serve only for the assignment of van der Waals radii and potential depths [73].

4.7.2 COBRAMM

COBRAMM is a software interface for tailoring molecular electronic structure calculations and running nanoscale simulations [79]. It is designed for hybrid QM/MM computations and accurate non-adiabatic dynamics in the excited state. It is well coupled with the Molcas suite of programs [80] and, along with the development of approaches to speed-up multi-reference perturbative calculations [81], it allows optimizations and non-adiabatic on-the-fly excited states dynamics at a full correlated level. Moreover, COBRAMM has recently integrated tools for accurate modelling of non-linear and bi-dimensional electronic spectroscopy in multichromophoric (bio)organic systems (e.g., DNA, proteins) [82, 83].

Ab Initio Simulations of Two-Dimensional Electronic Spectra

Two-dimensional (2D) UV electronic spectroscopy (2DES) [84-88] is a recently developed technique, which can disentangle signals arising from different intermolecular interactions that remain hidden in the 1D pump-probe spectra. 2DUV electronic spectroscopy can target the π - π^* transitions of aromatic residues, and can be employed for tracking all sorts of localised, delocalised, energy and charge-transfer photo-induced phenomena; however, their interpretation proves to be challenging. Alternatively, computational simulation of 2DES can be used to separate the different peaks into their contributing specific electronic transitions, thus providing a tool to recognise, understand, and interpret the observed primary fingerprints [89]. In this approach, an explicit mixed quantum mechanics/molecular mechanics (QM/MM) methodology is employed for the evaluation of electronic excited state energies and transition dipole moments, in which different photochemical outcomes are associated with specific molecular conformations.

The Sum-Over-States (SOS) approach [90] and the QM/MM scheme (SOS//QM/MM) [89, 91] have shown great potential to simulate 2DES and to characterise different conformations in DNA and proteins. It has been shown that 2DES obtained with this approach can distinguish between two configurations of a peptide containing both unstacked (non-interacting) and π or T-stacked (interacting) peptide motifs containing phenolic side chains in proteins or DNA/RNA [82, 92-94]. Multiconfigurational and multireference perturbation methods, such as complete active space self-consistent field (CASSCF) and second-order multireference perturbation theory (CASPT2) techniques, can be applied to calculate the electronic properties of multichromophoric system reliably. Hybrid QM/MM methods and molecular dynamics techniques can be used to assess environmental and conformational effects, respectively, that shape the 2D electronic spectra. Further details of the theory, implementation, and generation of 2D-ES can be found in refs. [82, 89, 92-94]

REFERENCES

1. Frenkel, D. and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*. 2nd ed. Vol. 1. 2002, San Diego: Academic Press.
2. Rapaport, D.C., *The art of molecular dynamics simulation*. 2nd ed. 2004, Cambridge, UK; New York: Cambridge University Press.
3. van der Spoel, D., et al., *Gromacs User Manual version 4.5.4*, www.gromacs.org. 2010.
4. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers*. SoftwareX, 2015. **1–2**: p. 19-25.
5. Bekker, H., et al. *Gromacs: A parallel computer for molecular dynamics simulations*. in *Physics computing*. 1993.
6. Berendsen, H.J.C., D. van der Spoel, and R. van Drunen, *GROMACS: A message-passing parallel molecular dynamics implementation*. Computer Physics Communications, 1995. **91**(1): p. 43-56.
7. Van Der Spoel, D., et al., *GROMACS: Fast, flexible, and free*. Journal of Computational Chemistry, 2005. **26**(16): p. 1701-1718.
8. Páll, S., et al., *Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS*, in *Solving Software Challenges for Exascale: International Conference on Exascale Applications and Software, EASC 2014, Stockholm, Sweden, April 2-3, 2014, Revised Selected Papers*, S. Markidis and E. Laure, Editors. 2015, Springer International Publishing: Cham. p. 3-27.
9. Weiner, S.J., et al., *A new force field for molecular mechanical simulation of nucleic acids and proteins*. Journal of the American Chemical Society, 1984. **106**(3): p. 765-784.
10. MacKerell, A.D., et al., *All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins†*. The Journal of Physical Chemistry B, 1998. **102**(18): p. 3586-3616.
11. MacKerell, A.D., N. Banavali, and N. Foloppe, *Development and current status of the CHARMM force field for nucleic acids*. Biopolymers, 2000. **56**(4): p. 257-265.

12. Martin, M.G., *Comparison of the AMBER, CHARMM, COMPASS, GROMOS, OPLS, TraPPE and UFF force fields for prediction of vapor–liquid coexistence curves and liquid densities*. *Fluid Phase Equilibria*, 2006. **248**(1): p. 50-55.
13. Jorgensen, W.L., 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*. *Journal of the American Chemical Society*, 1996. **118**(45): p. 11225-11236.
14. Jorgensen, W.L. and J. Tirado-Rives, *The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin*. *Journal of the American Chemical Society*, 1988. **110**(6): p. 1657-1666.
15. Beveridge, D.L. and F.M. DiCapua, *Free Energy Via Molecular Simulation: Applications to Chemical and Biomolecular Systems*. *Annual Review of Biophysics and Biophysical Chemistry*, 1989. **18**(1): p. 431-492.
16. Oostenbrink, C., et al., *A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6*. *Journal of Computational Chemistry*, 2004. **25**(13): p. 1656-1676.
17. Guvench, O., et al., *Additive empirical force field for hexopyranose monosaccharides*. *J Comput Chem*, 2008. **29**.
18. Guvench, O., et al., *CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses*. *J Chem Theory Comput*, 2009. **5**.
19. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water*. *The Journal of Chemical Physics*, 1983. **79**(2): p. 926-935.
20. Dick, T.J. and J.D. Madura, *Chapter 5 A Review of the TIP4P, TIP4P-Ew, TIP5P, and TIP5P-E Water Models*, in *Annual Reports in Computational Chemistry*. 2005, Elsevier. p. 59-74.
21. Mark, P. and L. Nilsson, *Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K*. *The Journal of Physical Chemistry A*, 2001. **105**(43): p. 9954-9960.
22. Bjelkmar, P., et al., *Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models*. *Journal of Chemical Theory and Computation*, 2010. **6**(2): p. 459-466.
23. Ostlund, N.S. and A. Szabo, *Modern Quantum Chemistry: Introduction to advanced electronic structure theory*. 1982: Macmillan.
24. McWeeny, R., *Methods of molecular quantum mechanics*. 1992: Academic press.
25. Hehre, W.J., et al., *Ab Initio Molecular Orbital Theory*, ed. Wiley. 1986.
26. Simons, J., *An experimental chemist's guide to ab initio quantum chemistry*. *The Journal of Physical Chemistry*, 1991. **95**(3): p. 1017-1029.
27. Simons, J. and J.A. Nichols, *Quantum mechanics in chemistry*. 1997: Oxford University Press New York.
28. Helgaker, T., P. Jorgensen, and J. Olsen, *Molecular electronic-structure theory*. 2014: John Wiley & Sons.

29. Kolos, W. and L. Wolniewicz, *Accurate Adiabatic Treatment of the Ground State of the Hydrogen Molecule*. The Journal of Chemical Physics, 1964. **41**(12): p. 3663-3673.
30. Sutcliffe, B.T., *The Nuclear Motion Problem in Molecular Physics*, in *Advances in Quantum Chemistry*, P.-O. Löwdin, et al., Editors. 1997, Academic Press. p. 65-80.
31. Roothaan, C., L. Sachs, and A. Weiss, *Rev. Mod. Phys.* 1951.
32. Hall, G.G., *The molecular orbital theory of chemical valency VIII. A method of calculating ionization potentials*. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 1951. **205**(1083): p. 541-552.
33. Pople, J. and R.K. Nesbet, *Self-consistent orbitals for radicals*. The Journal of Chemical Physics, 1954. **22**(3): p. 571-572.
34. Fermi, E., *A statistical method for determining some properties of the atom. I. Atti della Accademia Nazionale dei Lincei, Classe di Scienze Fisiche, Matematiche e Naturali, Rendiconti (1927), 6 602-7. CODEN: AANLAW ISSN:0001-4435. CAN 22:14799 AN 1928:14799*
35. Yang, W., *Density-functional theory of atoms and molecules*. 1989, Oxford University Press, Oxford.
36. Hohenberg, P. and W. Kohn, *Inhomogeneous electron gas*. Physical review, 1964. **136**(3B): p. B864.
37. Kohn, W. and L. Sham, *Phys Rev A 140: 1131*. Sham LJ, Kohn W (1966) Phys Rev, 1965. **145**: p. 561.
38. Vosko, S.H., L. Wilk, and M. Nusair, *Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis*. Canadian Journal of physics, 1980. **58**(8): p. 1200-1211.
39. Perdew, J.P. and Y. Wang, *Accurate and simple analytic representation of the electron-gas correlation energy*. Physical Review B, 1992. **45**(23): p. 13244.
40. Dirac, P.A.M., *Note on Exchange Phenomena in the Thomas Atom*. Mathematical Proceedings of the Cambridge Philosophical Society, 1930. **26**(3): p. 376-385.
41. Slater, J.C., *A simplification of the Hartree-Fock method*. Physical review, 1951. **81**(3): p. 385.
42. Becke, A.D., *Density-functional exchange-energy approximation with correct asymptotic behavior*. Physical review A, 1988. **38**(6): p. 3098.
43. Handy, N.C. and A.J. Cohen, *Left-right correlation energy*. Molecular Physics, 2001. **99**(5): p. 403-412.
44. Becke, A.D., *Density-functional thermochemistry. III. The role of exact exchange*. The Journal of Chemical Physics, 1993. **98**(7): p. 5648-5652.
45. Stephens, P.J., et al., *Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields*. The Journal of Physical Chemistry, 1994. **98**(45): p. 11623-11627.
46. Kümmel, S. and L. Kronik, *Orbital-dependent density functionals: Theory and applications*. Reviews of Modern Physics, 2008. **80**(1): p. 3-60.

47. Burke, K., J. Werschnik, and E.K.U. Gross, *Time-dependent density functional theory: Past, present, and future*. The Journal of Chemical Physics, 2005. **123**(6): p. 062206.
48. March, N.H., *Electron density theory of atoms and molecules*. The Journal of Physical Chemistry, 1982. **86**(12): p. 2262-2267.
49. Townsend, J., J.K. Kirkland, and K.D. Vogiatzis, *Chapter 3 - Post-Hartree-Fock methods: configuration interaction, many-body perturbation theory, coupled-cluster theory*, in *Mathematical Physics in Theoretical Chemistry*, S.M. Blinder and J.E. House, Editors. 2019, Elsevier. p. 63-117.
50. Roos, B.O., P.R. Taylor, and P.E.M. Sigbahn, *A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach*. Chemical Physics, 1980. **48**(2): p. 157-173.
51. Roos, B.O., *The Multiconfigurational (MC) Self-Consistent Field (SCF) Theory*, in *Lecture Notes in Quantum Chemistry: European Summer School in Quantum Chemistry*, B.O. Roos, Editor. 1992, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 177-254.
52. Malmqvist, P.A., A. Rendell, and B.O. Roos, *The restricted active space self-consistent-field method, implemented with a split graph unitary group approach*. The Journal of Physical Chemistry, 1990. **94**(14): p. 5477-5482.
53. Andersson, K., et al., *Second-order perturbation theory with a CASSCF reference function*. The Journal of Physical Chemistry, 1990. **94**(14): p. 5483-5488.
54. Andersson, K., P.Å. Malmqvist, and B.O. Roos, *Second-order perturbation theory with a complete active space self-consistent field reference function*. The Journal of Chemical Physics, 1992. **96**(2): p. 1218-1226.
55. Pulay, P., *A perspective on the CASPT2 method*. International Journal of Quantum Chemistry, 2011. **111**(13): p. 3273-3279.
56. Kohn, W. and L.J. Sham, *Self-Consistent Equations Including Exchange and Correlation Effects*. Physical Review, 1965. **140**(4A): p. A1133-A1138.
57. Mintmire, J.W. and B.I. Dunlap, *Fitting the Coulomb potential variationally in linear-combination-of-atomic-orbitals density-functional calculations*. Physical Review A, 1982. **25**(1): p. 88-95.
58. Salahub, D.R., et al., *QM/MM Calculations with deMon2k*. Molecules, 2015. **20**(3): p. 4780-4812.
59. Koster, A.M., et al., *deMon2k Version 5, The deMon developers*. 2018: Cinvestav, Mexico City.
60. Veryazov, V., et al., *MOLCAS as a development platform for quantum chemistry software*. International Journal of Quantum Chemistry, 2004. **100**(4): p. 626-635.
61. Aquilante, F., et al., *Molcas 8: New capabilities for multiconfigurational quantum chemical calculations across the periodic table*. Journal of Computational Chemistry, 2016. **37**(5): p. 506-541.
62. Aquilante, F., et al., *MOLCAS 7: The Next Generation*. Journal of Computational Chemistry, 2010. **31**(1): p. 224-247.
63. Warshel, A. and M. Levitt, *Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme*. Journal of Molecular Biology, 1976. **103**(2): p. 227-249.

64. Field, M.J., *Simulating enzyme reactions: Challenges and perspectives*. Journal of Computational Chemistry, 2002. **23**(1): p. 48-58.
65. van der Kamp, M.W. and A.J. Mulholland, *Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology*. Biochemistry, 2013. **52**(16): p. 2708-2728.
66. Tongraar, A., T. Kerdcharoen, and S. Hannongbua, *Simulations of Liquid Ammonia Based on the Combined Quantum Mechanical/Molecular Mechanical (QM/MM) Approach*. The Journal of Physical Chemistry A, 2006. **110**(14): p. 4924-4929.
67. Acevedo, O. and W.L. Jorgensen, *Cope Elimination: Elucidation of Solvent Effects from QM/MM Simulations*. Journal of the American Chemical Society, 2006. **128**(18): p. 6141-6146.
68. Field, M.J., P.A. Bash, and M. Karplus, *A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations*. Journal of Computational Chemistry, 1990. **11**(6): p. 700-733.
69. Lamoureux, G. and B.t. Roux, *Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm*. The Journal of Chemical Physics, 2003. **119**(6): p. 3025-3039.
70. Warshel, A., M. Kato, and A.V. Pisliakov, *Polarizable Force Fields: History, Test Cases, and Prospects*. Journal of Chemical Theory and Computation, 2007. **3**(6): p. 2034-2045.
71. Assfeld, X. and J.-L. Rivail, *Quantum chemical computations on parts of large molecules: the ab initio local self consistent field method*. Chemical Physics Letters, 1996. **263**(1): p. 100-106.
72. Gao, J., et al., *A Generalized Hybrid Orbital (GHO) Method for the Treatment of Boundary Atoms in Combined QM/MM Calculations*. The Journal of Physical Chemistry A, 1998. **102**(24): p. 4714-4721.
73. de la Lande, A., et al., *Molecular Simulations with in-deMon2k QM/MM, a Tutorial-Review*. Molecules, 2019. **24**(9): p. 1653.
74. Gamboa, G.U., et al., *How Important are Temperature Effects for Cluster Polarizabilities?* The Journal of Physical Chemistry A, 2008. **112**(47): p. 11969-11971.
75. Vásquez-Pérez, J.M., et al., *The discovery of unexpected isomers in sodium heptamers by Born-Oppenheimer molecular dynamics*. The Journal of Chemical Physics, 2009. **131**(12): p. 124126.
76. Alvarez-Ibarra, A., et al., *First Principles Computational Biochemistry with deMon2k*, in *Frontiers in Computational Chemistry*. 2015, Elsevier. p. 281-325.
77. Verlet, L., *Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules*. Physical Review, 1967. **159**(1): p. 98-103.
78. Swope, W.C., et al., *A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters*. The Journal of Chemical Physics, 1982. **76**(1): p. 637-649.
79. Weingart, O., et al., *COBRAMM 2.0 — A software interface for tailoring molecular electronic structure calculations and running nanoscale (QM/MM) simulations*. Journal of Molecular Modeling, 2018. **24**(9): p. 271.
80. Pepino, A.J., et al., *Resolving Ultrafast Photoinduced Deactivations in Water-Solvated Pyrimidine Nucleosides*. The Journal of Physical Chemistry Letters, 2017. **8**(8): p. 1777-1783.

81. Martínez-Fernández, L., et al., *Photophysics of Deoxycytidine and 5-Methyldeoxycytidine in Solution: A Comprehensive Picture by Quantum Mechanical Calculations and Femtosecond Fluorescence Spectroscopy*. Journal of the American Chemical Society, 2017. **139**(23): p. 7780-7791.
82. Nenov, A., et al., *Disentangling Peptide Configurations via Two-Dimensional Electronic Spectroscopy: Ab Initio Simulations Beyond the Frenkel Exciton Hamiltonian*. The Journal of Physical Chemistry Letters, 2014. **5**(4): p. 767-771.
83. Segatta, F., et al., *A Quantum Chemical Interpretation of Two-Dimensional Electronic Spectroscopy of Light-Harvesting Complexes*. Journal of the American Chemical Society, 2017. **139**(22): p. 7558-7567.
84. Selig, U., et al., *Coherent two-dimensional ultraviolet spectroscopy in fully noncollinear geometry*. Optics Letters, 2010. **35**(24): p. 4178-4180.
85. Tseng, C.-h., S. Matsika, and T.C. Weinacht, *Two-Dimensional Ultrafast Fourier Transform Spectroscopy in the Deep Ultraviolet*. Optics Express, 2009. **17**(21): p. 18788-18793.
86. Tseng, C.-h., et al., *Two-Dimensional Fourier Transform Spectroscopy of Adenine and Uracil Using Shaped Ultrafast Laser Pulses in the Deep UV*. The Journal of Physical Chemistry A, 2012. **116**(11): p. 2654-2661.
87. West, B.A., J.M. Womick, and A.M. Moran, *Probing Ultrafast Dynamics in Adenine With Mid-UV Four-Wave Mixing Spectroscopies*. The Journal of Physical Chemistry A, 2011. **115**(31): p. 8630-8637.
88. West, B.A. and A.M. Moran, *Two-Dimensional Electronic Spectroscopy in the Ultraviolet Wavelength Range*. The Journal of Physical Chemistry Letters, 2012. **3**(18): p. 2575-2581.
89. Rivalta, I., et al., *Ab initio simulations of two-dimensional electronic spectra: The SOS//QM/MM approach*. International Journal of Quantum Chemistry, 2014. **114**(2): p. 85-93.
90. Mukamel, S., R. Oszwaldowski, and D. Abramavicius, *Sum-over-states versus quasiparticle pictures of coherent correlation spectroscopy of excitons in semiconductors: Femtosecond analogs of multidimensional NMR*. Physical Review B, 2007. **75**(24): p. 245305.
91. Abramavicius, D., et al., *Coherent Multidimensional Optical Spectroscopy of Excitons in Molecular Aggregates; Quasiparticle versus Supermolecule Perspectives*. Chemical Reviews, 2009. **109**(6): p. 2350-2408.
92. Segarra-Martí, J., et al., *Two-dimensional electronic spectroscopy as a tool for tracking molecular conformations in DNA/RNA aggregates*. Faraday Discussions, 2018. **207**(0): p. 233-250.
93. Nenov, A., et al., *Tracking Conformational Dynamics of Polypeptides by Nonlinear Electronic Spectroscopy of Aromatic Residues: A First-Principles Simulation Study*. ChemPhysChem, 2014. **15**(15): p. 3282-3290.
94. Nenov, A., et al., *Two-Dimensional Electronic Spectroscopy of Benzene, Phenol, and Their Dimer: An Efficient First-Principles Simulation Protocol*. Journal of Chemical Theory and Computation, 2015. **11**(8): p. 3755-3771.

Chapter 3. Structures, intermolecular interactions and chemical hardness of binary water- organic solvents

Contents

Abstract	102
1. Introduction.....	102
2. Methods and computational details	104
2.1 Molecular dynamics simulations	104
2.2 Dynamic total hardness	106
3. Results and discussion.....	107
3.1 Thermodynamic properties: validation of the MD simulations.....	107
3.2 Structures of the aqueous binary solvents	109
3.3 Analysis of the intermolecular interaction types	113
3.4 Averaged hardness from MD simulations	117
4. Conclusions.....	119
5. References.....	120

Abstract

The evolution of structural properties, thermodynamics and averaged (dynamic) total hardness values as a function of the composition of binary water-organic solvents, is rationalized in view of the intermolecular interactions. The considered organic solvents are ethanol, acetonitrile, and isopropanol at 0.25, 0.5, 0.75, and 1 mass fractions, and the results are obtained using molecular dynamics simulations. The site-to-site radial distribution functions reveal a well-defined peak for the first coordination shell in all the solvents. The characteristic peak of the second coordination shell exists in aqueous mixtures of acetonitrile, whereas in the water-alcohol solvents, a second peak develops with the increase of the alcohol content. From the computed coordination numbers, averaged hydrogen bonds and their lifetimes, we found that water mixed with acetonitrile largely preserves its structural features and promotes the acetonitrile structuring. Both the water and alcohol structures in their mixtures are disturbed and form hydrogen bonds between molecules of different kinds. The dynamic hardness values are obtained as the average over the total hardness values of 1200 snapshots per solvent type, extracted from the equilibrium dynamics. The dynamic hardness profile has a non-linear evolution with the liquid compositions, similarly to the thermodynamic properties of these non-ideal solvents.

1. Introduction

The intermolecular interactions between water and organic solvents are driving the microstructuring in aqueous mixtures of simple organic solvents, which determines their thermodynamic properties [1]. In recent years, the studies of structural organization and the mixing behavior at the molecular level in binary water-simple organic solvents has gained a revived interest, because of their use in the biomass fractionation processes [2-6], in the formation of membranes for CO₂/flue gases (N₂) separation [7], in addition to the plethora of well-established applications in electrochemistry, organic synthesis, chromatography, and solvent extractions. Moreover, the individual properties of the organic components confer them distinctive features when mixed with water at various contents, temperature and pressure conditions.

In this work, we will focus on aqueous mixtures of ethanol (EtOH), isopropanol (2PrOH), and acetonitrile (AN). A variety of physical [8-12], thermodynamic [13-16], transport [9, 17-19], electronic [12, 20-23], and structure properties [24-28] are available from the experimental and computational studies. Simple alcohols, such as EtOH and 2PrOH, are miscible in water, form strong hydrogen bonds with water molecules, and are thought to be inhibitors for clathrate hydrate formation. Thermodynamics of short (methanol and ethanol) alcohol-water mixtures have been shown to depend on the solvent composition in a very complex way [29-38]. Study of ¹H NMR measurements [33] at -10° C found that the decrease of the alcohol mole fraction to ~0.2, strengthens the H-bonds between water and hydroxyl OH groups in ethanol while enhancing the water structure through H-bonding. Pozar et al. 2016 [34] used NMR techniques to understand the

microheterogeneity of EtOH-water mixtures [34] and suggested that ethanol is entirely structured by hydrogen bonding, whereas this is not true for water. The calculated structure factors pointed to the conclusion that water is less hydrogen bonded when EtOH concentration increases, even if the change in intensities is smooth. Petong et al. [35] observed non-linear changes of ethanol dipole moment with the changes of the mixture concentration. From the molecular dynamics simulations by Saiz et al. [1, 36] of pure liquid ethanol, the role of H-bonding on the structural and dynamical properties was evidenced. The hydrogen bonding study in water-alcohol mixtures by X-ray absorption spectroscopy of Lam et al. [37] suggests that the additional hydrogen bonding interactions generated from the interaction of water-alcohol would result in superior ordering in the liquid structures, leading to a reduction in entropy and a negative enthalpy of mixing. This would be true for EtOH. However, the spectra of the 2PrOH-water mixtures exhibit an increase in the number of broken alcohol hydrogen bonds for mixtures containing up to 0.5 water mole fraction. Therefore, this explanation does not account for the negative excess entropy of the 2PrOH-water mixtures, and still needs to be addressed. Finneran et al. [38] succeeded to register the rotational spectrum of an isolated EtOH-dimer using Fourier transform microwave spectroscopy and identified the presence of only gauche EtOH conformer, stabilized by H-bonding with water. These measurements confirm that ethanol is a better hydrogen-bond acceptor than donor. Ab initio calculations [38] identified water and EtOH as donor and acceptor, respectively.

Acetonitrile, on the other hand, although miscible with water, is not a hydrogen bond donor and the water-acetonitrile mixture is known to present a microheterogeneity [39]. The term microheterogeneity has been introduced to define the structural organizations of multi-component liquids, where the molecules of each component are surrounded by the molecules of the same kind. The various patterns, hitherto proposed [16, 26, 40-47] for microstructural organization in water-AN mixtures revealed that the structuring of this binary solvent is a complex function of the liquid compositions, similarly to the short alcohols - water mixtures. Starting with the earlier thermodynamics study of Robertson and Sugamori [41] on the temperature dependence of the enthalpy (ΔH) and of the specific heat capacity at constant pressure (C_p), the water structure was found broken, or partially molt, due to the presence of the acetonitrile component in the liquid. Moreau and Douheret [42] established the existence of three interaction regions with different structural patterns in water-AN mixtures. In the water-rich region, below 0.2 mol fraction of AN, the voids of the aqueous structure are progressively filled by AN, without enhancement of the water cluster. The geometry of the acetonitrile molecules could not easily fit into the void, which led to the disruption of the H-bond network. By increasing the AN concentration, an intermediate region between ~0.2 to ~0.8-mole fraction was characterized by a progressive break of the H-bonds, which resulted in a reduction of the water aggregates and gave rise to a region of microheterogeneity. In this region water molecules were mostly surrounded by water, and AN molecules were surrounded by AN. At AN concentrations, above ~0.8 mol, the domains of predominantly AN molecules are progressively disrupted by the addition of water molecules. Nevertheless, the aqueous mixtures of AN are characterized by microheterogeneity, the formation of water clusters has been

excluded from the analysis of IR spectra for the mixtures above 0.15 mol concentration [45]. Moreover, above the equimolar mixture, water solvated preferably AN than other water molecules, which suggested that water molecules form short water chains and rings [45]. In mixtures with high AN content, water molecules can gain a motional freedom, whereas at lower AN concentration water dimers and trimers are observed [44]. Formation of true H-bonds between water and AN molecules at low AN concentration have been concluded from X-ray Diffraction and IR spectroscopy [46] and radial distribution functions [26]. The H-bond lifetime is found to increase with the decrease of water content from the molecular dynamics simulations [47]. The dipole-dipole dimers between water and AN are suggested to co-exists with the first-shell H-bonds in the mixtures with AN content between 0.2 - 0.6 mol [46]. This correlates with the ab-initio results [43] that established two types of water-AN dimers, one favouring the hydrogen bonding with water and another favoring the dipole-dipole interaction. However, the dipole-dipole interactions in pure AN clusters is reduced in the presence of water because hydrogen-bonded AN dimers appear [48].

This work aims to provide a systematic comparative study of the structural, dynamic and thermodynamic properties of aqueous mixtures of three organic solvents (called cosolvent throughout this paper), and namely ethanol (EtOH), isopropanol (2PrOH) and acetonitrile (AN) at 0.25, 0.50, and 0.75 mass fraction concentrations (X_{Cosol}), using molecular dynamics simulations. The results for the four pure liquids obtained at the same theoretical level are also included. In addition, for the first time, it is proposed to evaluate the chemical hardness at the equilibrium dynamics and to analyze its behaviour as a function of cosolvent mass concentration.

2. Methods and computational details

2.1 Molecular dynamics simulations

All-atom molecular dynamics (MD) simulations were carried out using the GROMACS [49-53] package version 2016.3, along with the CHARMM36 additive force field [54, 55] and the TIP4P model for liquid water [56-58], which better reproduces the electrostatic distribution around water molecules in comparison with other 3-sites models. Solvent structures for EtOH, 2PrOH, and AN are available in the GROMACS molecule and liquid database [59]. The compositions of the cubic boxes of 4 nm, used in our simulations, are presented in Table 3.1 with the calculated number of molecules to reproduce the experimental densities at each concentration.

For each simulation box, energy minimization was performed using the steepest descent algorithm until convergence to a tolerance of $100 \text{ kJ.mol}^{-1}.\text{nm}^{-1}$. After minimization, unconstrained simulations to stabilize and distribute the solvent molecules were performed for 200 ps at 298.15 K and 1 bar with a 0.5-fs time step, and a frame-saving rate (for analysis) of 1ps. Temperature and pressure coupling was handled using the leap-frog stochastic dynamics integrator and the Parrinello-Rahman method [60, 61], respectively. Initial velocities

were generated from a Maxwell distribution at 298.15 K, and 15-ns MD simulations were performed. The isothermal-isobaric (NPT) ensemble was considered for data collection. Neighbour searching and short-range nonbonded interactions were handled with the Verlet cut-off scheme.

Table 3.2. Compositions of the simulated systems in 4 nm cubic boxes for the modelled solvent mixtures, studied in this work. The cosolvent mass fraction (X_{Cosol}) and the cosolvent molar fraction (χ_{Cosol}), the number of cosolvent molecules (N_{Cosol}) and the number of water molecules (N_{water}) are reported.

Solvent system	X_{Cosol}	χ_{Cosol}	N_{Cosol}	N_{water}
Pure water	0	0	0	2133 ^a
Pure EtOH	1	1	657 ^b	0
Pure 2PrOH	1	1	501 ^c	0
Pure AN	1	1	730 ^d	0
EtOH-water mixture	0.25	0.115	201	1539
	0.50	0.281	381	973
	0.75	0.540	534	455
2PrOH-water mixture	0.25	0.910	153	1531
	0.50	0.231	288	960
	0.75	0.473	448	851
AN-water mixture	0.25	0.128	224	1534
	0.50	0.305	417	950
	0.75	0.568	582	442

^{a)} 6399 atoms + 2133 virtual sites. ^{b)} 5913 atoms. ^{c)} 6012 atoms. ^{d)} 4380 atoms.

Electrostatics were treated with the Fast, smooth Particle-Mesh Ewald (SPME) method, with a Coulomb cut-off of 1.2 nm, a fourth-order interpolation and Fourier spacing of 0.12 nm. Van der Waals (vdW) interactions were treated using the Lennard-Jones potential with a cut-off distance of 1.2 nm. The potential was decreased over the whole range, whereas the forces were decayed smoothly to zero between 1.0 nm to the cut-off distance. A LINCS algorithm was used to constrain the bonds when constraints applied, 12 being the highest order in the expansion of the constraint coupling matrix.

The structure and dynamics of the water-cosolvent mixtures, using the GROMACS analysis tools, were characterized by the site-to-site radial distribution functions (RDFs), coordination numbers (CN), the average number of hydrogen bonds (HBs), HB lifetimes, and intermolecular energies. Each of these descriptors was calculated for all the water-cosolvent concentrations considered in this study, using the last 5-ns trajectory for analysis. Site-to-site RDFs were computed for the water-water, cosolvent-cosolvent, and cosolvent-water pairs using the water oxygen, EtOH oxygen, 2PrOH oxygen, and AN nitrogen. The number of neighbour molecules was obtained by integration of the RDFs up to a 2-nm correlation distance, from where the CN can be

extracted at the first minimum of RDF. HBs were calculated using a geometrical criterion with a maximum donor-acceptor distance of 0.35 nm and a hydrogen-donor-acceptor angle of 30°. The forward HB lifetimes were computed from the autocorrelation function of each H-bond type, using the theory of Luzar and Chandler [62, 63], as described by van der Spoel et al. [64].

2.2 Dynamic total hardness

The concepts of chemical hardness and softness, introduced by Pearson [65], have been proven useful to classify the acids and bases as hard or soft in many chemical reactions with the aim to predict the product's stability prior to the laboratory work. The hard (soft) chemical species are characterized as hardly (easily) polarizable. Since the formulation of the hard-soft-acid-base principle of Pearson [65], the chemical hardness and softness concepts have been widely explored in organic chemistry and homogeneous catalysis [66]. The theoretical formulation of the chemical hardness and softness, was derived in the frame of Density Functional Theory (DFT) by Parr et al. [67], followed by developments and implementations of various algorithms for their numerical calculations [68-72]. Various reactivity indices within DFT were derived from the chemical hardness/softness giving rise to the entire research field called conceptual DFT (see, for example, the exhaustive recent review in ref. [73]).

Among these algorithms, the reactivity indices, computed with the orbitally resolved hardness tensor approach (ORHT) [74] were successfully applied in isomerization [75, 76] and protonation [77] reactions, as selectivity descriptors [78], and in the framework of Pearson's hard-soft-acid-base principle [77, 78]. A simplification of the ORHT method was proposed as well, where the diagonal orbital hardness elements are replaced by the atomic hardnesses, either computed at the DFT level from the energies of orbitals with fractional electron occupations [69, 74-80], or, obtained from the experiments as the sum of the measured ionization potential (I) and the electronic affinity (A) [66]. This approach, which is called the atomically resolved hardness tensor (ARHT), provides an efficient scheme for easy and relatively fast calculations of the reactivity indices in large scale systems. Here, for the first time we apply the ARHT approach to large scale dynamic systems, as are the 4 nm boxes, used in the present MD simulations, in order to compute the averaged total hardness. In the present ARHT calculations the hardness value (η_A) of each constituent atom A is the experimental atomic hardness, taken from ref. [66]. In a molecular system at equilibrium the electronegativity equalization principle (EEP) holds, which allows to estimate the interactions between every AB atom pair from their interatomic distances in a molecule. Keeping in mind that EEP holds for the molecular structures (snapshots), extracted from the equilibrium dynamics simulations, the interatomic hardness (η_{AB}) can be obtained from the atomic hardness values, η_A and η_B , and the interatomic distances R_{AB} . R_{AB} distances are provided by MD simulations and η_{AB} are computed from the empirical relation proposed by Ohno [81]:

$$\eta_{AB} = \frac{1}{\sqrt{b_{AB}^2 + R_{AB}^2}}$$

$$b_{AB} = \frac{2}{\eta_A + \eta_B}$$
Eq. (1)

The elements of the inverse hardness tensor are the atomic softness elements s_{AB} , [69, 74-80], their sum gives the total softness (S), from which the total hardness can be computed as follows:

$$S = \sum_{AB} s_{AB}$$

$$\eta = \frac{1}{S} = \frac{1}{\sum_{AB} s_{AB}}$$
Eq. (2)

The dynamic hardness is then easily derived as the average over the total hardnesses η_i , $i=1, \dots, N$ computed from ARHT approach for every snapshot, extracted from the equilibrium dynamics trajectories. Each snapshot, thus, provides the R_{AB} distances, used in Eq. (1). For N trajectories, the dynamic total hardness is $\langle \eta \rangle = (\sum_{i=1}^N \eta_i) / N$. In this first study, we use only the experimental atomic hardnesses, taken from ref [66]. The total hardness for each solvent model and content was calculated for a set of 1200 snapshots extracted from the last 300 ps of the simulation trajectories, recorded at every 0.25 ps. The number of molecules in each box, used for the ARHT computations, is provided in Table 3.1. The experimental atomic hardness values, η_A in eV, are 6.43 for H, 5.00 for C, 7.23 for N, and 6.08 for O [66].

3. Results and discussion

3.1 Thermodynamic properties: validation of the MD simulations

The models used in the MD simulations were first validated for the pure water, EtOH, 2PrOH and AN solvents from a comparison between the computed and experimental density (ρ), diffusivity (D), enthalpy of vaporization (ΔH_{vap}), and the averaged number of H-bonds. These thermodynamic properties were obtained over 5 ns of the equilibrium MD trajectories. The theoretical results, collected in Table S3.1 in the Supporting Information (SI) section, reproduce closely the reference experimental data. This demonstrates the very good performance of TIP4P water model and CHARMM36 additive force field [82]. The thermodynamic properties, calculated for the three aqueous mixtures of EtOH, 2PrOH, and AN are collected in Table S3.2 (SI Materials). All the simulated solvent mixtures reproduced closely the reference experimental conditions, which are $T=298.15$ K and $P=1$, similarly to the pure solvents (see Table S3.1).

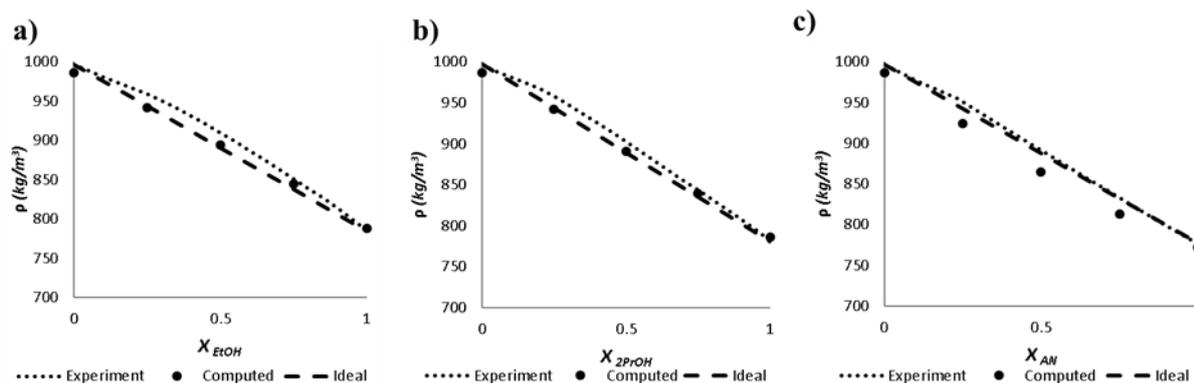


Figure 3.1. Computed densities for solvents and their aqueous mixtures. Experimental curves are included for comparison and were taken from ref. [8, 9], ref. [10], and ref. [11, 12] for EtOH, 2PrOH, and AN, respectively.

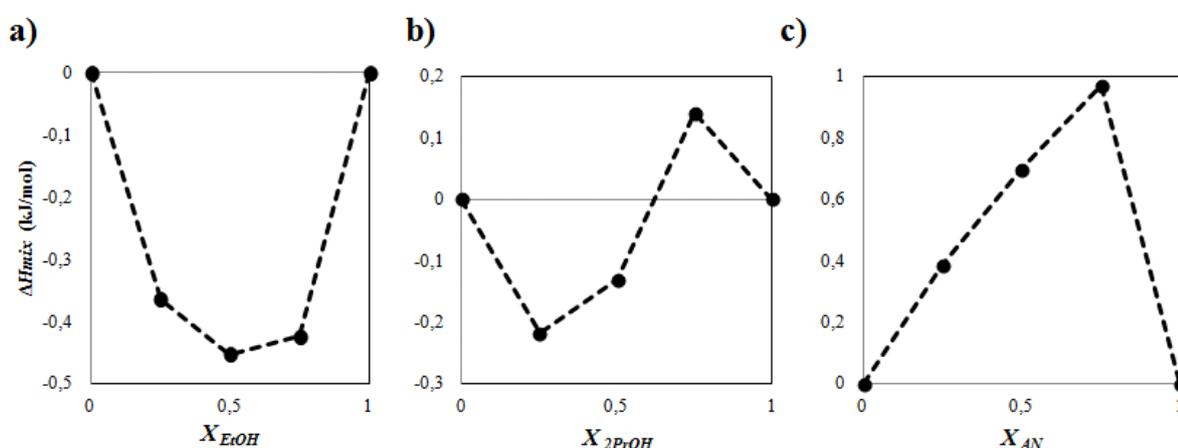


Figure 3.2. Computed enthalpy of mixing, ΔH_{vap} , for the water- a) EtOH, b) 2PrOH, and c) AN solvent mixtures at different cosolvent mass fractions, X_{cosol}

The density profile as a function of the binary mixture composition plotted in Fig. 3.1, demonstrates the non-ideality of the water - cosolvent mixtures. Again a good reproduction (with errors below 3% for all mixed solvents) of the experimental density profile in Fig. 3.1 is obtained. Similar are the conclusions from the mixing enthalpy (Fig. 3.2), calculated as

$$\Delta H_{mix} = H_{liq,mix} - H_{liq,1} * y_1 - H_{liq,2} * y_2 = (U + PV)_{liq,mix} - (U + PV)_{liq,1} * y_1 - (U + PV)_{liq,2} * y_2 ,$$

The internal energy, U , of the liquid mixture is obtained directly from the potential energy (see Tables S3.1 and S2) during the simulation and y_i is the molar fraction of component i . The ΔH_{mix} profiles reported in Fig. 3.2, display correctly the expected non-ideal behaviour for all the solvent mixtures, in agreement with the experiment [14-16]. The curves show that the mixing of EtOH and water is exothermic at all concentrations, whereas 2PrOH-water mixing becomes endothermic with increasing the cosolvent content. The AN mixing with water is endothermic, since the penetration of AN molecules into the water H-bond network requires additional energy to break the water-water H-bonds. In conclusion, we note that this excellent comparison

between the calculated and known experimental thermodynamic features demonstrate the very good performance of the chosen force-fields and numerical details.

3.2 Structures of the aqueous binary solvents

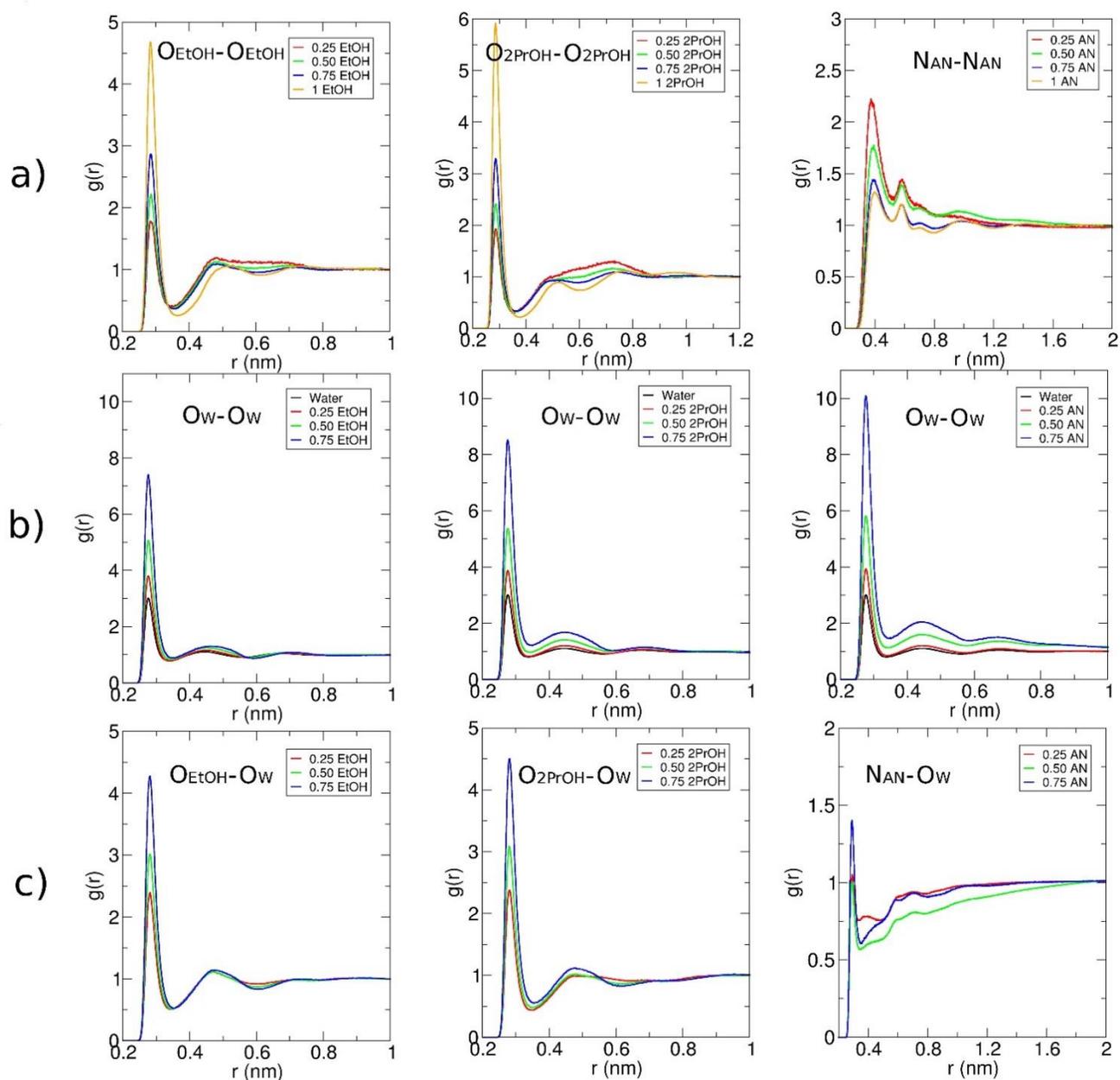


Figure 3.3 Radial distribution functions, $g(r)$, of the a) cosolvent-cosolvent, b) water-water, and c) cosolvent-water pairs. The water oxygen, EtOH oxygen, 2PrOH oxygen, and AN nitrogen are the reference sites.

In order to analyse the structural properties, we computed the RDFs of the cosolvent-cosolvent, water-water, and cosolvent-water pairs at each solvent composition. The radial distribution functions of the cosolvent-cosolvent pairs displayed in Fig. 3.3. Fig. 3.3a compare well with previously studied RDFs in pure and mixed solvents [36, 83-86]. The EtOH O-O radial distribution curve (in Fig. 3.3a, left panel) attains a maximum of

4.7 at a distance of 0.284 nm, thus agreeing with the results of Saiz et al. [36]. The 2PrOH O-O RDF (in Fig. 3.3a, middle panel) displays a peak of 5.9 at a distance of 0.286 nm and compares well with the results of Anisimov et al. [83]. In the acetonitrile N-N RDF profile in Fig. 3.3a, right panel, the first peak has a height of 1.36 and appears at 0.398 nm, followed by a second peak of height 1.2 at 0.58 nm, similarly to the N-N RDF obtained from 3 and 6-sites model potentials [84, 85]. In Fig. 3.3b, the RDF of the distances between oxygen atoms in H₂O molecules (O_{water}) displays a sharp first peak with an intensity of 3.00 at 0.276 nm, which reproduces well the experimental 3.09 peak probability [86]. Moreover, the second and third peaks, experimentally observed around 0.452 nm and 0.677 can be appreciated near the same values. The $O_{\text{water}}-O_{\text{cosolv}}$ RDFs are presented in Fig. 3.3c.

From all the RDFs in Fig. 3.3 we conclude that the solvent content does not affect the positions of the maximum peaks of the mixed solvents, but only their heights and depths. On the contrary, the secondary peaks in the RFFs are significantly affected by the presence of water. It is evident that the ordering in the solvents disappears beyond the first coordination shell with the increase of the water content. In Fig. 3.3a, left panel, the second shell structure of EtOH is gradually recovered as the content of EtOH increases. Similarly, the second shell structure in 2PrOH (Fig. 3.1 a, middle) rises with the 2PrOH concentration. Moreover, the 2PrOH shell is found to be more extended in comparison to EtOH, which indicates that the 2PrOH aggregates formed in the mixture have a more complex organisation than in EtOH. Following $N_{\text{AN}}-N_{\text{AN}}$ RDFs, we note that the second coordination sphere is preserved with the increase of the water content. Moreover, a visual inspection of the trajectories (see Fig. S3.1) showed predominantly an anti-parallel mutual displacement between the CH₃CN molecules in the first shell. Perpendicular mutual orientations are also observed in the first coordination shell. Parallel orientations between the units dominate in the second shell. The RDF profiles in the alcohol-water mixtures display significantly sharper peaks than the RDF profile in the water-AN solvents.

With the increase of the alcohol contents, the first-coordination peak develops and has a well distinguishable high intensity in the pure alcohol solvents. The opposite behaviour of RDF peaks is established for the acetonitrile co-solvent. In addition, the RDFs profiles in the AN-water mixtures are not characterised by sharp peaks (Fig 3.3a and c), and the intensity ratio between the first and second peak decreases with the amount of AN up to the pure solvent. It seems that the presence of water promotes the structural ordering in acetonitrile at variance to the alcohols. The large overlapping area between the first and second peaks in the acetonitrile RDF indicate high numbers of AN and water molecules between the first and second coordination shells, increasing with the AN content.

The cumulative numbers of molecules as a function of distances are calculated from the integration of RDFs and plotted in Fig. 3.4 in the range of $r = 0.5$ nm for the cosolvents and $r = 0.35$ nm for $O_{\text{w}}-O_{\text{w}}$ pairs. At the first coordination sphere distances, the cumulative number corresponds to the coordination number within the first coordination sphere (Table 3.2).

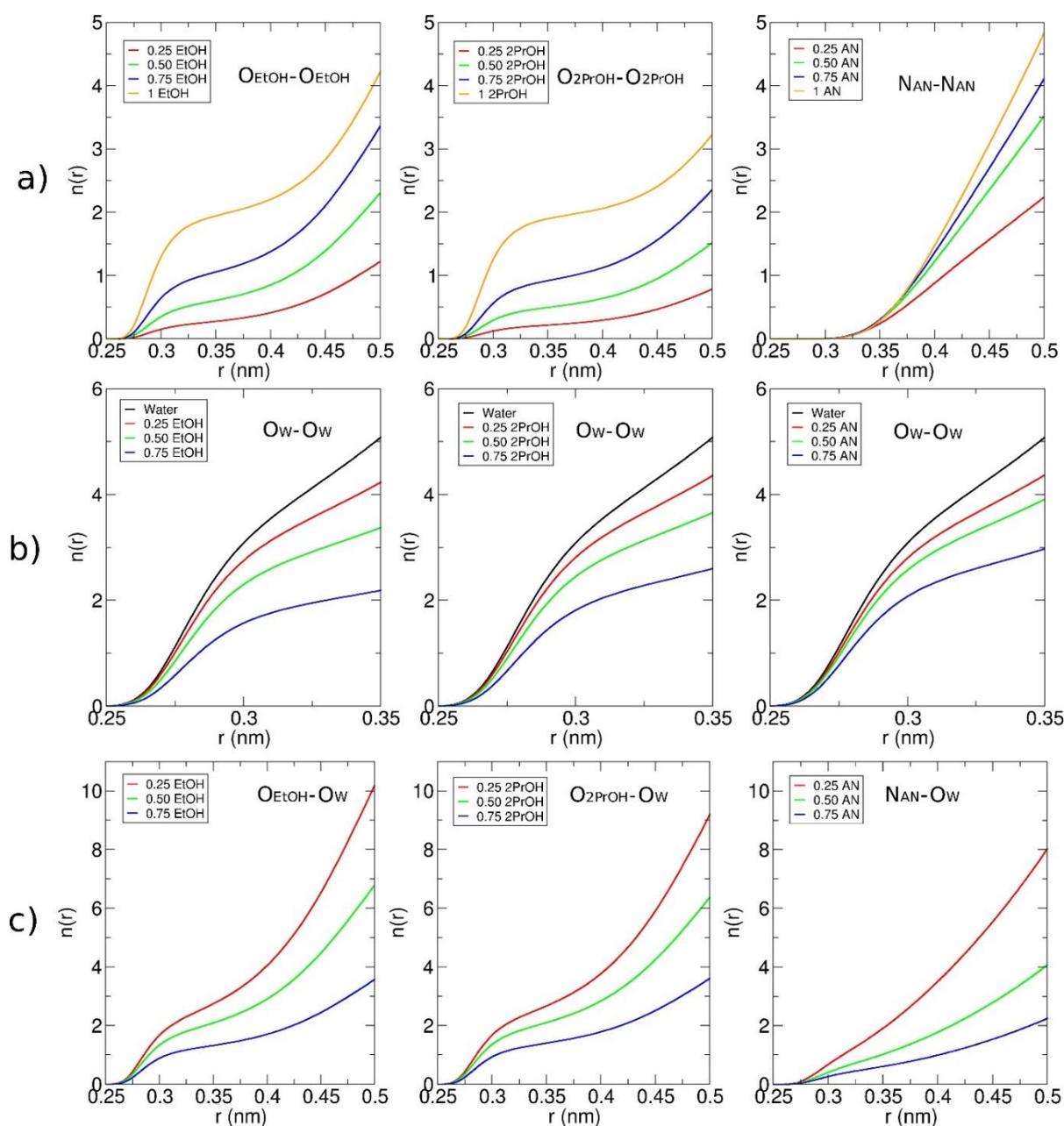


Fig. 3.4 Cumulative numbers, $n(r)$, in the short-range distance ($r < 0.5$ nm) of the a) cosolvent-cosolvent, b) water-water, and c) cosolvent-water pairs. The water oxygen, EtOH oxygen, 2PrOH oxygen, and AN nitrogen are the reference sites

The cumulative numbers of AN - AN neighbours increase with their concentration, whereas the number of water-water neighbours decreases. For acetonitrile, this finding is in a very good agreement with acoustic and positron annihilation measurements of aqueous solutions of acetonitrile reporting a destabilization of AN-clathrates with the increase of the AN concentration in water [87]. It also points to microheterogeneity, hitherto largely accepted in the literature (see Introduction section). Moreover, the number of water-water neighbour molecules is higher in AN mixture than in the alcohol mixtures. We attribute this behaviour to the

chemical nature of the intermolecular interactions of the solvents that define the local homogeneity of the mixture. In the AN liquid, the dipole-dipole interactions dominate, whereas in EtOH and in 2PrOH the network of hydrogen bonds is formed. Since acetonitrile is less polar than the other solvents, does not make hydrogen bonds with itself, and can interact with water only as an H-bond acceptor, the individual molecules will repel each other as the mixture becomes more concentrated, leading therefore to a local phase separation, as exemplified in Fig. S3.1. As observed by mass spectrometry [16], several aprotic solvents exhibited ‘additional mixing’, in which AN molecules cannot substitute water molecules inside their clusters; thus they interact with the water cluster as an external agent.

Table 3.2. Cutoff distance of the first coordination peak (r_{min}) and coordination numbers (CN) of each interaction pair, calculated from the cumulative numbers, $c(r)$, at r_{min} .

	Cosolvent- Cosolvent		Water-Water		Cosolvent- Water	
	r_{min} (nm)	CN	r_{min} (nm)	CN	r_{min} (nm)	CN
X_{EtOH}						
0	-	-	0.338	4.6	-	-
0.25	0.352	0.3	0.338	3.9	0.344	2.6
0.50	0.346	0.6	0.346	3.3	0.348	2.1
0.75	0.358	1.1	0.35	2.2	0.35	1.3
1	0.364	2.0	-	-	-	-
X_{2PrOH}						
0	-	-	0.338	4.6	-	-
0.25	0.348	0.2	0.338	4.0	0.354	2.7
0.50	0.356	0.5	0.344	3.5	0.346	2.1
0.75	0.364	1.0	0.35	2.6	0.354	1.4
1	0.378	2.0	-	-	-	-
X_{AN}						
0	-	-	0.338	4.6	-	-
0.25	0.53	2.7	0.34	4.1	0.338	1.6
0.50	0.524	4.1	0.344	3.8	0.342	0.9
0.75	0.516	4.6	0.352	3.0	0.35	0.6
1	0.51	5.2	-	-	-	-

The cumulative numbers of AN - AN neighbours increase with the AN concentration, whereas the number of water-water neighbours decreases. This suggests an increase in water-acetonitrile interactions. For acetonitrile, this finding is in a very good agreement with acoustic and positron annihilation measurements of aqueous solutions of acetonitrile reporting a destabilization of AN- clathrates with the increase of the AN concentration in water [87]. Moreover, the number of water-water neighbour molecules is higher in AN mixture than in the alcohol mixtures. It, therefore, points to microheterogeneity of the water-AN solvents, hitherto largely accepted in the literature (see Introduction section). This behaviour is attributed to the chemical nature of the intermolecular interactions of the solvents that define the microstructure organisation in the mixtures. In the AN liquid, the dipole-dipole interactions dominate, whereas in EtOH and in 2PrOH the network of hydrogen bonds is formed. Since acetonitrile is less polar than the other solvents, does not make hydrogen bonds with itself, and can interact with water only as an H-bond acceptor, the individual molecules will repel each other as the mixture becomes more concentrated, leading therefore to a local phase separation, as exemplified in Fig. S3.1. As observed by mass spectrometry [16], several aprotic solvents exhibited 'additional mixing', in which AN molecules cannot substitute water molecules inside their clusters; thus they interact with the water cluster as an external agent.

The number of water to water neighbours (Fig. 3.4b) decreases as the cosolvent concentration increases. This means that water structures are experiencing rupture due to the presence of the cosolvents. Alcohols, known to be strong structure breakers, particularly EtOH, interact with water molecules by substitution, contrary to the AN-water interactions. At all the concentrations, AN molecules display the lowest structure breaker power. Also, from the values of the coordination numbers, it can be inferred that water is a strong structure breaker for alcohol structures, but promotes partially the ordering of the AN liquid component. On the other hand, the number of cosolvent molecules surrounding water molecules increases with the cosolvent concentration. At short distances, < 0.35 nm, the coordination of cosolvent to water neighbours is controlled by hydrogen bonds. At this distance, the number of alcohol neighbours to the water molecules increases if the alcohol content > 0.5 mass concentration (see Fig. 3.4c), which shows that the alcohols are H-bonded with water. The number of AN neighbouring the water molecules is significantly smaller at $r < 0.35$ nm and increases faster than the alcohol-water neighbours at larger distances, $r \geq 0.7$ nm, as follows from the cumulative numbers in the range of 2 nm, presented in Fig. S3.2. This indicates that water-AN interactions are not dominated by H-bonding. The much faster increase of $N_{AN}-N_{AN}$ cumulative number in comparison to $N_{EtOH}-N_{EtOH}$ and $N_{2PrOH}-N_{2PrOH}$ in the interval of 0.35 - 0.5 nm, suggests an enhanced microheterogeneity of the water-AN mixture in this region.

3.3 Analysis of the intermolecular interaction types

The variations of the Coulomb and Lenard-Jones (LJ) energies and H-bonds with the kind and content of the cosolvent are used to infer the effect of the mixture compositions on the intermolecular interactions. In Fig. 3.5 the water-water, cosolvent-cosolvent, and cosolvent-water Coulomb and LJ energies are plotted for the

considered mixed and pure liquids. The cosolvent-cosolvent energies (Fig. 3.5a) decrease fast (meaning the electrostatic interactions increased) with the content of the organic solvents, which is expected. In ethanol, the water-EtOH coulombic interactions (Fig. 3.5b) attain their maximum at $X=0.75$ mass fraction content. In isopropanol, the maximum of the electrostatic and LJ interactions (minimum in Fig. 3.5a) is observed near $0.5 X_{\text{Cosol}}$. This suggests a decrease of the 2PrOH-water interactions occurring already at concentrations of about $0.5 X_{\text{Cosol}}$ most probably because of predominant 2PrOH-2PrOH interactions. This leads to conclude that 2PrOH aggregations start to occur between $0.5-0.75 X_{\text{cosol}}$. The water-acetonitrile interactions find a minimum of the LJ interactions at the same concentration; however, the Coulomb interactions increases up to $0.75 X_{\text{Cosol}}$. This points to the conclusion that the electrostatic interactions between the water and the organic component have not a linear behaviour with the liquid composition, and they are cosolvent specific. The electrostatic and LJ energies (Fig. 3.5c) computed for the water-water component of the solvents do not experience any significant variations in the presence of cosolvents.

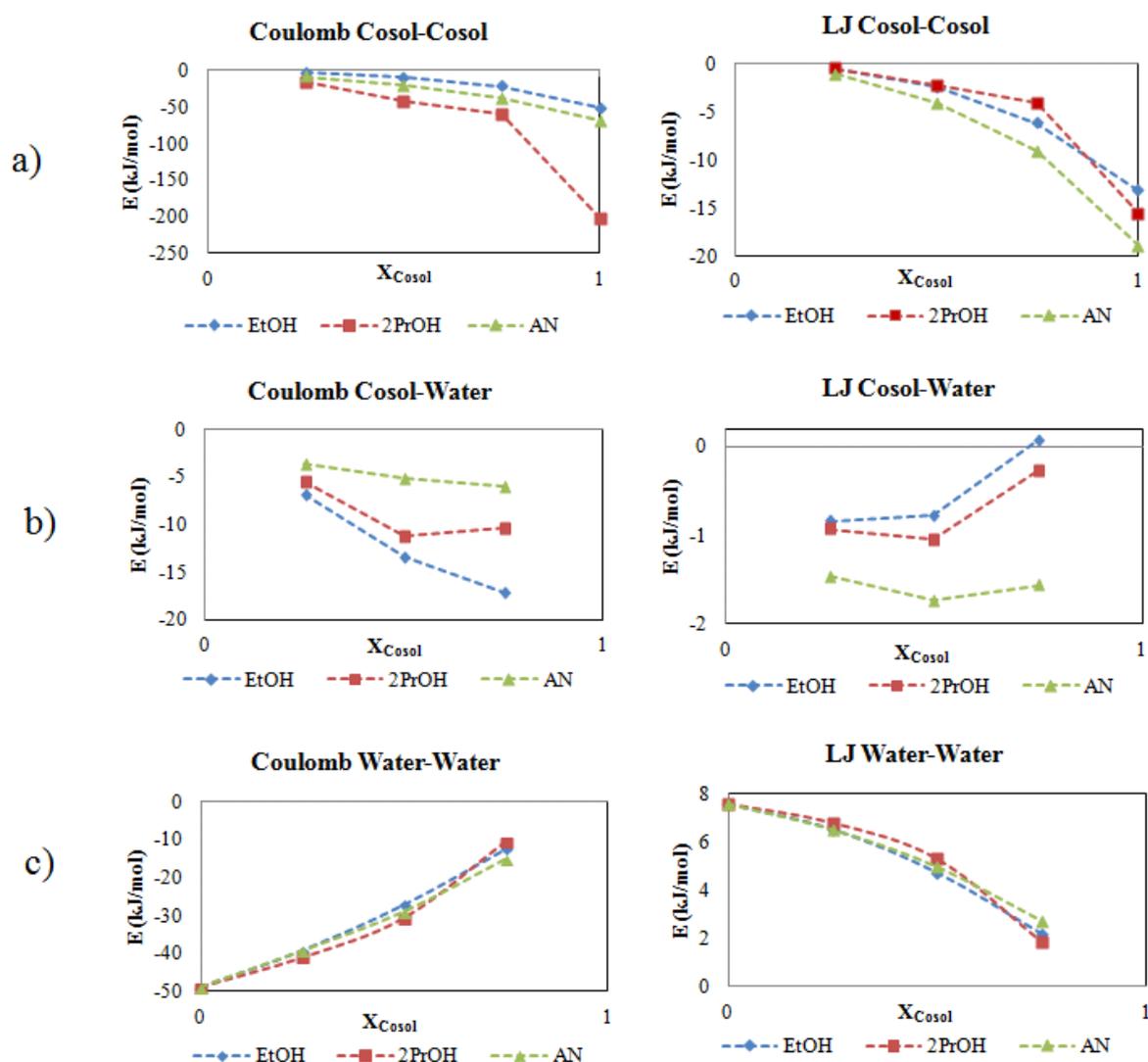


Figure 3.5. Coulomb and Lennard-Jones (LJ) interaction energies of the a) cosolvent-cosolvent, b) cosolvent-water, and c) water-water types.

The H-bonds evolution with the liquid compositions can be analysed from the computed averaged number of H-bonds reported in Table 3.3. Our simulations reproduce well the average H-bonds per molecule obtained by previous MD simulations for the same water potential [64, 88, 89]. For example, Noskov et al. [90] obtained an average H-bonds per water molecule of 3.03 but using the SWM4-DP polarizable water model. Likewise, our results for the calculated H-bond numbers (HB) in EtOH agreed closely with the results of Noskov et al. [90], reporting HB = 1.65 using a polarizable potential, and of Saiz et al. [36], who found a value of 1.9 using the united atom OPLS force field. We did not find reference values for HB in 2PrOH; however, this value should be similar to EtOH since the two compounds have the same number of OH groups, even though 2PrOH is less polar. Acetonitrile does not make any hydrogen bonds with itself since no hydrogen is bonded to any N atom.

The H-bonds between water molecules dominate in the three binary solvents up to 0.50 X_{cosol} . The averaged number of alcohol-water H-bonds becomes greater than HB_{w-w} only when alcohols attain a concentration of 0.75 X_{cosol} . Interestingly, the presence of water, even at its minimal concentration, disturbs the H-bonding network in the alcohols. This follows from the significantly smaller values of $HB_{EtOH-EtOH}$ and $HB_{2PrOH-2PrOH}$ in the mixed solvents than in the pure alcohol liquids. Both alcohols interact in a similar manner with water; that is, they preferably form H-bonds with H_2O than between themselves for cosolvent concentrations ≤ 0.75 . This leads to conclude that a molecular aggregation of the short alcohols in water is not evidenced by the present MD simulations. We believe that in mixtures above 0.50 X_{cosol} , the water molecules optimise the number of H-bonding sites available in order to reduce the non-favourable interactions with the alcohol solvents. This is also an indication of the more hydrophobic character of EtOH and 2PrOH alcohols.

Table 3.3. Calculated average numbers of hydrogen bonds per molecule (HB) from the MD simulations of cosolvent-water mixtures at 0.25, 0.50, 0.75, and 1 cosolvent mass fraction (X_{coso}).

	EtOH			
X_{cosol}	0.25	0.50	0.75	1
HB_{sys}	3.35	3.06	2.60	1.84
HB_{W-W}	2.83	1.93	0.85	-
$HB_{EtOH-EtOH}$	0.03	0.14	0.50	1.84
HB_{EtOH-W}	0.50	0.98	1.25	-
	2POH			
X_{cosol}	0.25	0.50	0.75	1
HB_{sys}	3.38	3.12	2.67	1.83
HB_{W-W}	2.97	2.20	1.12	-
$HB_{2PrOH-2PrOH}$	0.02	0.10	0.39	1.83
$HB_{2PrOH-W}$	0.39	0.82	1.16	-

	AN			
X_{cosol}	0.25	0.50	0.75	1
HB _{sys}	3.08	2.44	1.50	-
HB _{W-W}	2.83	2.07	1.05	-
HB _{AN-AN}	-	-	-	-
HB _{AN-W}	0.24	0.37	0.45	-

The acetonitrile, as stated above, does not form H-bonds with itself. Note that the AN-AN HBs were not taken into consideration. The AN-water HB grow with the increase of AN content, which confirms experimental studies, reporting on the existence of H-bonds between water and AN molecules [46]. The H-bonded water-alcohol molecules are however significantly fewer in comparison to H-bonded water-alcohol molecules, thus confirming the microheterogeneity of this mixture and the predominant dipole-dipole interactions.

Furthermore, we calculated the HB autocorrelation functions and hydrogen bond lifetimes, as previously described in the computational details. The HBs autocorrelation functions (see Fig. S3.3) display a fast decay of the hydrogen bonds studied here, and the velocity of this decay is influenced by the cosolvent concentration. The lifetimes of the H-bonds are summarised in Table 3.4.

Table 3.4. Hydrogen bond (HB) lifetimes for the water-water (w-w), water-cosolvent (w-cosol), and cosolvent-cosolvent (cosol-cosol) interaction types, obtained from the HB autocorrelation functions of each HB type.

	HB lifetimes (ps)								
Solvent mixture	Water-EtOH			Water-2PrOH			Water-AN		
	X_{cosol}	W-W	Cosol- W	Cosol- Cosol	W-W	Cosol- W	Cosol- Cosol	W-W	Cosol- W
0		4.3	-	-	4.3	-	-	4.3	-
0.25		7.1	8.3	8.5	7.5	11.3	16.7	6.0	2.3
0.50		12.4	13.1	13.2	11.1	14.5	26.3	7.3	3.1
0.75		21.7	21.3	20.4	21.4	28.6	44.3	10.2	3.9
1		-	-	40.3	-	-	119.1	-	-

In all the cases, we note an increase of the H-bond lifetime with the cosolvent content in line with the results in ref. [4, 47]. This tendency is well pronounced for the water and alcohol liquids and their mixtures at variance to water-AN solvents. Our computed H-bond lifetimes display lengths in line with those obtained by other authors [4, 91]. Very recent 2D infrared spectroscopy studies of diluted alcohol solutions (≤ 0.09 mol

fraction) [91] revealed a vibrational lifetime of ~ 3 ps of the OH stretching mode, comparable with the expected values of our alcohol-alcohol HB lifetimes if extrapolated for diluted alcohol concentrations. Furthermore, we obtained water-AN and water-water HB lifetimes comparable to those obtained for water-tetrahydrofuran mixtures, another aprotic solvent [4], reporting HB lifetimes in the range between ~ 2 and ~ 4.5 ps. As inferred from the water diffusivities in THF mixtures [4], the increase of the water-water H-bond lifetimes obeys to a reduction in the free movement of water molecules in the solvent mixtures.

The significantly longer H-bond lifetimes in the pure EtOH (40.3 ps) and 2PrOH (119.1 ps) solvents, in comparison to those in their water mixtures, points to conclude a strong perturbation of the alcohols structures when mixed with water. Again, relatively short lifetimes of water-AN H-bonds corroborates the microheterogeneity. Moreover, the H-bond lifetimes vary only little with the increase of the AN content. This is in agreement with the observations of X-ray Diffraction and IR spectroscopy [46]. This authors established that the water and acetonitrile HB interactions play an important role in the formation of a large interface between water agglomeration domains and AN agglomeration domains. In summary, the water structure mixed with acetonitrile appears to be only slightly perturbed, whereas this structure perturbation is stronger in alcohol mixtures. Furthermore, the alcohol structure experiences a significant change in their H-bonding networks.

3.4 Averaged hardness from MD simulations

The characterisation of mixed liquids as hard and soft might find its interest in providing a qualitative prediction of the reactivity and product stability of numerous chemical reactions that take place in solvents. The main problem faced in using the gas-phase reactivity indices (including the chemical hardness) for the reactivity predictions of reactions in solvents, was associated with the modulation of the gas-phase reactivity because of the hydration energies [66]. This effect causes failures of the hard-soft-acid-base principle [65], stating that hard (soft) acids (bases) preferably reacts with hard (soft) bases (acids). Examples are described in ref. [66] together with the difficulties associated with the experimental estimations of the chemical hardness (or, equivalently, the chemical softness). It is therefore of interest to establish a methodology for theoretical computations of the total chemical hardness, or, total chemical softness, of solvents by explicitly considering their atomic structures along the equilibrium dynamics.

In Fig. 3.6 and Table S3.3, the computed dynamic total hardness values, $\langle \eta \rangle$, as a function of the liquid composition, are presented. As follows, the variations of $\langle \eta \rangle$ in the three binary mixtures are nonlinear with the increase of the cosolvent content, similar to the nonlinear behaviour of the enthalpy of mixing in Fig. 3.2. Therefore, the dynamic hardness can also be considered as a measure for the non-ideal behaviour of the binary solvents, analogously to the thermodynamic properties (*vide supra*).

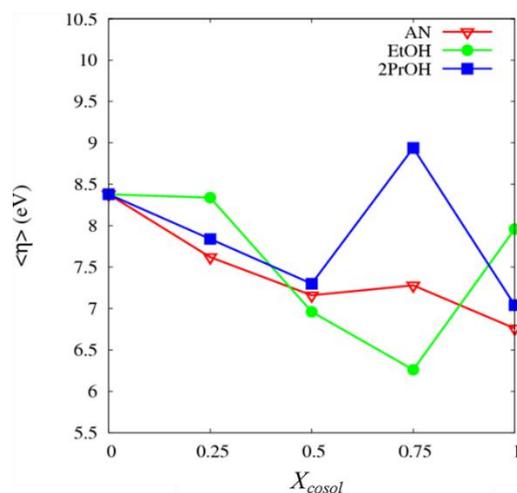


Figure 3.6 Computed dynamic total hardnesses, $\langle \eta \rangle$, as a function of the cosolvent composition, X_{cosol} .

We first note that the dynamic hardnesses of the pure solvents remain close to the available experimental values [66], which are obtained from the measured ionisation potential and electron affinity of the molecules of water (9.4 eV), 2PrOH (8.0 eV) and AN (7.5 eV). In comparison to the pure water solvents, the addition of the organic component causes a decrease of $\langle \eta \rangle$ up to 0.50 X_{cosol} of the three cosolvents, but the water-EtOH mixture becomes softer also at 0.75 X_{EtOH} . The behaviour of the other two cosolvents is the opposite - the addition of $X=0.75$ 2PrOH and AN into water liquid increases the hardnesses of these mixtures. The mixed solvents have systematically smaller $\langle \eta \rangle$ values, compared to $\langle \eta \rangle = 8.38$ eV of the pure water, with the exception of the dynamic hardness of water-2PrOH at 0.75 X_{cosol} . In the latter case, $\langle \eta \rangle$ increases notably up to 8.94 eV.

With an attempt to better understand this behaviour, we computed with density-functional theory (DFT) approach the total hardness for the two simplest molecular patterns that appear in the mixed solvents, namely the water-cosolvent and cosolvent - cosolvent dimers. Their total hardness values, reported in Table 3.5, are obtained from the HOMO-LUMO energy differences and labelled as η_{H-L} . For the DFT calculations, we used PBE exchange-correlation functional with DZVP and TZVP basis sets. For these simple dimers, exactly the same trend of η_{H-L} was found with both bases.

Table 3.5. DFT total hardness (η_{H-L}) in eV computed as HOMO-LUMO energy difference using DFT-PBE and TZVP basis for the water-water (w-w), water-cosolvent, and cosolvent-cosolvent dimers. The cosolvent molecules are ethanol (EtOH), isopropanol (2PrOH) and acetonitrile (AN).

Dimer	w-w	w-EtOH	EtOH-EtOH	w-2PrOH	2PrOH-2PrOH	w-AN	AN-AN
η_{H-L}	6.38	5.28	5.63	6.21	5.71	5.18	6.94

The water-EtOH dimer is found to be softer than EtOH-EtOH dimer. The observed softening of the w-EtOH solvent mixture with the increase of EtOH concentration (see Fig. 3.6) can be thus explained with a continues

increase of water-EtOH interactions when increasing the EtOH concentration. In the pure EtOH solvent, the hardness increases because of the harder EtOH-EtOH pair interactions. Contrary to the ethanol cosolvent, η_{H-L} reveals that the dimer of water and isopropanol (w-2PrOH) is harder than the dimer of 2PrOH-2PrOH. This correlates with the strong increase of $\langle\eta\rangle = 8.94$ eV at $X_{2PrOH} = 0.75$, which suggests that mixed water-isopropanol interactions are predominant at this concentration and the isopropanol solvent structure is significantly disrupted by H₂O molecules. At the smaller $X_{2PrOH} = 0.25$ and 0.5 concentrations, the 2PrOH-2PrOH interactions dominate, because these mixtures are softer and their $\langle\eta\rangle$ values are close to the hardness of the pure isopropanol solvent. In the case of acetonitrile, the relation $\eta_{H-L}(w-AN) < \eta_{H-L}(AN-AN)$ generally correlates with the decrease of the dynamic hardness profile in Fig. 3.6, but does not explain the softer pure AN-solvent. For the acetonitrile solvent with dipole-dipole interactions (not H-bonding as in the case of alcohol-containing solvents) the total hardness seems to be dependent on the size of the cluster. We computed η_{H-L} for a cluster of 20 AN molecules, extracted from MD simulations and optimized at DFT level, and found $\eta_{H-L} = 6.24$ eV, being smaller than η_{H-L} of the dimer. Therefore, a decrease of the total hardness as a function of the acetonitrile solvent structure and size of the box seems plausible.

Furthermore, comparing $\langle\eta\rangle$ between the pure solvents, the following trend holds: water \geq 2PrOH $>$ EtOH $>$ AN. It, therefore, appears that mixtures with predominant HB interactions between the constituent molecules are harder than those with predominant dipole-dipole interactions. The $\langle\eta\rangle$ tendency in the pure solvents respects the molecular hardness trend [66], which is in line with the fact that the individual molecular properties in the studied liquids are well preserved. However, mixing with water alters non-negligibly this $\langle\eta\rangle$ tendency, moreover in a non-linear manner. Therefore, the variation of the averaged dynamic hardness appears to be a complex function of the mixed solvents compositions, analogously to the structural and thermodynamical features. The latter conclusion justifies the need of tools for relatively easy, but robust, estimations of hardnesses (respectively softnesses) in complex solvents, which will enable a rational reactivity predictions of various chemical processes in liquids within the HSAB principle, providing the chemical hardness (or, equivalently, the softnesses) of the solute systems are also computed.

4. Conclusions

This work covered a systematic comparative study of structural and thermodynamic properties of selected pure solvents (water, ethanol, acetonitrile, and isopropanol) and their aqueous mixtures at different concentrations, using the tools of the molecular dynamics simulations. In addition, a simple method for the averaged, dynamic, total hardness computations is suggested. The obtained thermodynamic properties are in a good agreement with the experimental and previous theoretical studies (see the text), demonstrating the excellent performance of the TIP4P (water) and CHARMM36 (organic solvents) model potentials for these types of binary mixtures.

The site-to-site radial distribution functions reveal that the concentrations of the mixed solvents do not affect the positions in the peaks and valleys, but only their heights and depths. The latter effect is particularly pronounced for the peaks, characteristics for the second coordination shells. The analysis of the average numbers of HBs with the liquid composition shows that the maximum of the interactions between the water and the organic solvents occurs at the mass fractions of 0.75. At this concentration, alcohol molecules tend to substitute water molecules, allowing compensating for the loss of H-bonds in the water solvent domains. The alcohol structures experience significant changes at all the concentrations. Acetonitrile is not an H-bond donor, and it is inferred that the presence of water and the formation of H-bonds helps it to reduce the strong dipole-dipole interactions while preserving the microheterogeneity of the mixture.

The averaged hardness, similar to the other dynamic and thermodynamic properties, has a nonlinear profile with the solvent compositions. As a general trend, a softening of the water solvent by introducing the organic cosolvents is obtained with the only exception for the isopropanol-water mixture at $X = 0.75$. The proposed method could be applicable within the HSAB principle to rationalise the behaviour of complex solute systems in mixed solvents, providing dynamic hardness values of the solutes are also assessed.

5. References

1. Saiz, L., J.A. Padro, and E. Guardia, *Dynamics and hydrogen bonding in liquid ethanol*. Molecular Physics, 1999. **97**(7): p. 897-905.
2. Chang, S., et al., *Solvent polarity and internal stresses control the swelling behaviour of green wood during dehydration in organic solution*. Bioresources, 2012. **7**(2).
3. Smith, M.D., et al., *Cosolvent pretreatment in cellulosic biofuel production: effect of tetrahydrofuran-water on lignin structure and dynamics*. Green Chemistry, 2016. **18**(5): p. 1268-1277.
4. Smith, M.D., et al., *Molecular Driving Forces behind the Tetrahydrofuran–Water Miscibility Gap*. The Journal of Physical Chemistry B, 2016. **120**(4): p. 740-747.
5. Li, M.-F., S. Yang, and R.-C. Sun, *Recent advances in alcohol and organic acid fractionation of lignocellulosic biomass*. Bioresource Technology, 2016. **200**: p. 971-980.
6. Bossu, J., et al., *Sorption of water–ethanol mixtures by poplar wood: swelling and viscoelastic behaviour*. Wood Science and Technology, 2018. **52**: p. 987–1008.
7. Jawad, Z.A., et al., *The Role of Solvent Mixture, Acetic Acid and Water in the Formation of CA Membrane for CO₂/N₂ Separation*. Procedia Engineering, 2016. **148**: p. 327-332.
8. Pires, R.M., et al., *Viscosity and Density of Water + Ethyl Acetate + Ethanol Mixtures at 298.15 and 318.15 K and Atmospheric Pressure*. Journal of Chemical & Engineering Data, 2007. **52**(4): p. 1240-1245.
9. Hurle, R.L., A.J. Easteal, and L.A. Woolf, *Self-diffusion in monohydric alcohols under pressure. Methanol, methan (2 H) ol and ethanol*. Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases, 1985. **81**(3): p. 769-779.

10. Pang, F.-M., et al., *Densities and viscosities of aqueous solutions of 1-propanol and 2-propanol at temperatures from 293.15 K to 333.15 K*. Journal of Molecular Liquids, 2007. **136**(1): p. 71-78.
11. Grande, M.d.C., H.L. Bianchi, and C.M. Marschoff, *On the density of pure acetonitrile*. Anales de la Asociación Química Argentina, 2004. **92**: p. 109-114.
12. Cunningham, G.P., G.A. Vidulich, and R.L. Kay, *Several properties of acetonitrile-water, acetonitrile-methanol, and ethylene carbonate-water systems*. Journal of Chemical & Engineering Data, 1967. **12**(3): p. 336-337.
13. Lide, D.R., *CRC handbook of chemistry and physics*. 12J204, 2004.
14. Peeters, D. and P. Huyskens, *Endothermicity or exothermicity of water/alcohol mixtures*. Journal of Molecular Structure, 1993. **300**(Supplement C): p. 539-550.
15. Stokes, R.H., *Excess partial molar enthalpies for (acetonitrile + water) from 278 to 318 K*. The Journal of Chemical Thermodynamics, 1987. **19**(9): p. 977-983.
16. Wakisaka, A., et al., *Non-ideality of binary mixtures Water-methanol and water-acetonitrile from the viewpoint of clustering structure*. Journal of the Chemical Society, Faraday Transactions, 1998. **94**(3): p. 369-374.
17. Shaker-Gaafar, N., et al., *p,T-Dependence of Self-diffusion in Liquid Ethanol and the Propanols*. Berichte der Bunsengesellschaft für physikalische Chemie, 1993. **97**(6): p. 805-811.
18. Hurle, R.L. and L.A. Woolf, *Self-diffusion in liquid acetonitrile under pressure*. Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases, 1982. **78**(7): p. 2233-2238.
19. Li, Z., et al., *Molecular dynamics simulation of self-diffusion coefficients for several alkanols*. Russian Journal of Physical Chemistry A, 2017. **91**(7): p. 1260-1269.
20. Akerlof, G., *Dielectric constants of some organic solvent-water mixtures at various temperatures*. Journal of the American Chemical Society, 1932. **54**(11): p. 4125-4139.
21. Wyman, J., *The dielectric constant of mixtures of ethyl alcohol and water from -5 to 40*. Journal of the American Chemical Society, 1931. **53**(9): p. 3292-3301.
22. Venables, D.S. and C.A. Schmuttenmaer, *Far-infrared spectra and associated dynamics in acetonitrile-water mixtures measured with femtosecond THz pulse spectroscopy*. The Journal of Chemical Physics, 1998. **108**(12): p. 4935-4944.
23. Rivelino, R., et al., *Electronic polarization in liquid acetonitrile: A sequential Monte Carlo/quantum mechanics investigation*. Chemical Physics Letters, 2005. **407**(1): p. 13-17.
24. Jorgensen, W.L. and J.M. Briggs, *Monte Carlo simulations of liquid acetonitrile with a three-site model*. Molecular Physics, 1988. **63**(4): p. 547-558.
25. Evans, M., *Molecular dynamics and structure of liquid acetonitrile - a review and computer simulation*. Journal of Molecular Liquids, 1983. **25**(3): p. 149-175.
26. Oldiges, C., et al., *MD Calculated Structural Properties of Clusters in Liquid Acetonitrile/Water Mixtures with Various Contents of Acetonitrile*. The Journal of Physical Chemistry A, 2002. **106**(31): p. 7147-7154.

27. M., N.A. and L.A. P., *New six-site acetonitrile model for simulations of liquid acetonitrile and its aqueous mixtures*. *Journal of Computational Chemistry*, 2007. **28**(12): p. 2020-2026.
28. Lindquist, B.A., R.T. Haws, and S.A. Corcelli, *Optimized Quantum Mechanics/Molecular Mechanics Strategies for Nitrile Vibrational Probes: Acetonitrile and para-Tolunitrile in Water and Tetrahydrofuran*. *The Journal of Physical Chemistry B*, 2008. **112**(44): p. 13991-14001.
29. Sato, T., A. Chiba, and R. Nozaki, *Hydrophobic hydration and molecular association in methanol–water mixtures studied by microwave dielectric analysis*. *The Journal of Chemical Physics*, 2000. **112**(6): p. 2924-2932.
30. Sato, T., A. Chiba, and R. Nozaki, *Dynamical aspects of mixing schemes in ethanol–water mixtures in terms of the excess partial molar activation free energy, enthalpy, and entropy of the dielectric relaxation process*. *The Journal of Chemical Physics*, 1999. **110**(5): p. 2508-2521.
31. Sato, T., A. Chiba, and R. Nozaki, *Composition-dependent dynamical structures of 1-propanol–water mixtures determined by dynamical dielectric properties*. *The Journal of Chemical Physics*, 2000. **113**(21): p. 9748-9758.
32. Sato, T. and R. Buchner, *Dielectric relaxation spectroscopy of 2-propanol–water mixtures*. *The Journal of Chemical Physics*, 2003. **118**(10): p. 4606-4613.
33. Matsugami, M., et al., *Hydrogen bonding in ethanol–water and trifluoroethanol–water mixtures studied by NMR and molecular dynamics simulation*. *Journal of Molecular Liquids*, 2016. **217**: p. 3-11.
34. Pozar, M., et al., *Micro-heterogeneity versus clustering in binary mixtures of ethanol with water or alkanes*. *Physical Chemistry Chemical Physics*, 2016. **18**(34): p. 23971-23979.
35. Petong, P., R. Pottel, and U. Kaatze, *Water–Ethanol Mixtures at Different Compositions and Temperatures. A Dielectric Relaxation Study*. *The Journal of Physical Chemistry A*, 2000. **104**(32): p. 7420-7428.
36. Saiz, L., J.A. Padró, and E. Guàrdia, *Structure and Dynamics of Liquid Ethanol*. *The Journal of Physical Chemistry B*, 1997. **101**(1): p. 78-86.
37. Lam, R.K., J.W. Smith, and R.J. Saykally, *Communication: Hydrogen bonding interactions in water-alcohol mixtures from X-ray absorption spectroscopy*. *The Journal of Chemical Physics*, 2016. **144**(19): p. 191103.
38. Finneran, I.A., et al., *Hydrogen bonding in the ethanol-water dimer*. *Physical Chemistry Chemical Physics*, 2015. **17**(37): p. 24210-24214.
39. Alavi, S., et al., *Hydrogen-bonding alcohol-water interactions in binary ethanol, 1-propanol, and 2-propanol+methane structure II clathrate hydrates*. *The Journal of Chemical Physics*, 2010. **133**(7): p. 074505.
40. Pothoczki, S. and L. Pusztai, *Intermolecular orientations in liquid acetonitrile: New insights based on diffraction measurements and all-atom simulations*. *Journal of Molecular Liquids*, 2017. **225**: p. 160-166.
41. Robertson, R.E. and S.E. Sugamori, *The Hydrolysis of t-Butyl Chloride in Aquo-Organic Mixtures: Heat Capacity of Activation and Solvent Structure*. *Canadian Journal of Chemistry*, 1972. **50**(9): p. 1353-1360.

42. Moreau, C. and G. Douhéret, *Thermodynamic behavior of water-acetonitrile mixtures excess volumes and viscosities*. *Thermochimica Acta*, 1975. **13**(4): p. 385-392.
43. Damewood, J.R. and R.A. Kumpf, *Hydration of polar organic molecules: the interaction of acetonitrile with water*. *The Journal of Physical Chemistry*, 1987. **91**(12): p. 3449-3452.
44. Kovacs, H. and A. Laaksonen, *Molecular dynamics simulation and NMR study of water-acetonitrile mixtures*. *Journal of the American Chemical Society*, 1991. **113**(15): p. 5596-5605.
45. Bertie, J.E. and Z. Lan, *Liquid Water–Acetonitrile Mixtures at 25 °C: The Hydrogen-Bonded Structure Studied through Infrared Absolute Integrated Absorption Intensities*. *The Journal of Physical Chemistry B*, 1997. **101**(20): p. 4111-4119.
46. Takamuku, T., et al., *Liquid Structure of Acetonitrile–Water Mixtures by X-ray Diffraction and Infrared Spectroscopy*. *The Journal of Physical Chemistry B*, 1998. **102**(44): p. 8880-8888.
47. Bakó, I., T. Megyes, and G. Pálinkás, *Structural investigation of water–acetonitrile mixtures: An ab initio, molecular dynamics and X-ray diffraction study*. *Chemical Physics*, 2005. **316**(1): p. 235-244.
48. Rowlen, K.L. and J.M. Harris, *Raman spectroscopic study of solvation structure in acetonitrile/water mixtures*. *Analytical Chemistry*, 1991. **63**(10): p. 964-969.
49. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers*. *SoftwareX*, 2015. **1–2**: p. 19-25.
50. Bekker, H., et al. *Gromacs: A parallel computer for molecular dynamics simulations*. in *Physics computing*. 1993.
51. Berendsen, H.J.C., D. van der Spoel, and R. van Drunen, *GROMACS: A message-passing parallel molecular dynamics implementation*. *Computer Physics Communications*, 1995. **91**(1): p. 43-56.
52. Van Der Spoel, D., et al., *GROMACS: Fast, flexible, and free*. *Journal of Computational Chemistry*, 2005. **26**(16): p. 1701-1718.
53. Páll, S., et al., *Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS*, in *Solving Software Challenges for Exascale: International Conference on Exascale Applications and Software, EASC 2014, Stockholm, Sweden, April 2-3, 2014, Revised Selected Papers*, S. Markidis and E. Laure, Editors. 2015, Springer International Publishing: Cham. p. 3-27.
54. Guvench, O., et al., *Additive empirical force field for hexopyranose monosaccharides*. *J Comput Chem*, 2008. **29**.
55. Guvench, O., et al., *CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses*. *J Chem Theory Comput*, 2009. **5**.
56. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water*. *The Journal of Chemical Physics*, 1983. **79**(2): p. 926-935.
57. Dick, T.J. and J.D. Madura, *Chapter 5 A Review of the TIP4P, TIP4P-Ew, TIP5P, and TIP5P-E Water Models*, in *Annual Reports in Computational Chemistry*. 2005, Elsevier. p. 59-74.
58. Mark, P. and L. Nilsson, *Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K*. *The Journal of Physical Chemistry A*, 2001. **105**(43): p. 9954-9960.

59. van der Spoel, D., P.J. van Maaren, and C. Caleman, *GROMACS molecule & liquid database*. Bioinformatics, 2012. **28**(5): p. 752-753.
60. Nosé, S. and M.L. Klein, *Constant pressure molecular dynamics for molecular systems*. Molecular Physics, 1983. **50**(5): p. 1055-1076.
61. Parrinello, M. and A. Rahman, *Polymorphic transitions in single crystals: A new molecular dynamics method*. Journal of Applied Physics, 1981. **52**(12): p. 7182-7190.
62. Luzar, A. and D. Chandler, *Hydrogen-bond kinetics in liquid water*. Nature, 1996. **379**: p. 55.
63. Luzar, A., *Resolving the hydrogen bond dynamics conundrum*. The Journal of Chemical Physics, 2000. **113**(23): p. 10663-10675.
64. van der Spoel, D., et al., *Thermodynamics of Hydrogen Bonding in Hydrophilic and Hydrophobic Media*. The Journal of Physical Chemistry B, 2006. **110**(9): p. 4393-4398.
65. Pearson, R.G., *Hard and Soft Acids and Bases*. Journal of the American Chemical Society, 1963. **85**(22): p. 3533-3539.
66. Pearson, R.G., *Chemical hardness*. 1997: Wiley Online Library.
67. Parr, R.G., et al., *Electronegativity: The density functional viewpoint*. The Journal of Chemical Physics, 1978. **68**(8): p. 3801-3807.
68. Nalewajski, R.F., *The hardness based molecular charge sensitivities and their use in the theory of chemical reactivity*, in *Chemical Hardness*, K.D. Sen, Editor. 1993, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 115-186.
69. Neshev, N. and T.Z. Mineva, *The Role of Interelectronic Interaction in Transition Metal Oxide Catalysts*, in *Metal-Ligand Interactions: Structure and Reactivity*, N. Russo and D.R. Salahub, Editors. 1996, Springer Netherlands: Dordrecht. p. 361-405.
70. Senet, P., *Nonlinear electronic responses, Fukui functions and hardnesses as functionals of the ground-state electronic density*. The Journal of Chemical Physics, 1996. **105**(15): p. 6471-6489.
71. Senet, P., *Kohn-Sham orbital formulation of the chemical electronic responses, including the hardness*. The Journal of Chemical Physics, 1997. **107**(7): p. 2516-2524.
72. Chandrakumar, K.R.S. and S. Pal, *A Systematic Study on the Reactivity of Lewis Acid-Base Complexes through the Local Hard-Soft Acid-Base Principle*. The Journal of Physical Chemistry A, 2002. **106**(48): p. 11775-11781.
73. Chattaraj, P.K. and D.R. Roy, *Update 1 of: Electrophilicity Index*. Chemical Reviews, 2007. **107**(9): p. PR46-PR74.
74. Mineva, T. and T. Heine, *Efficient Computation of Density-Functional Orbitally Resolved Reactivity Indices*. The Journal of Physical Chemistry A, 2004. **108**(50): p. 11086-11091.
75. Mineva, T., et al., *Density functional potential-energy hypersurface and reactivity indices in the isomerization of X3H+ (X = O, S, Se, Te)*. Journal of the Chemical Society, Faraday Transactions, 1997. **93**(18): p. 3309-3312.

76. Mineva, T., E. Sicilia, and N. Russo, *Density-Functional Approach to Hardness Evaluation and Its Use in the Study of the Maximum Hardness Principle*. Journal of the American Chemical Society, 1998. **120**(35): p. 9053-9058.
77. Mineva, T., et al., *Gas-phase properties and Fukui indices of sulfine (CH₂SO). Potential energy surface and maximum hardness principle for its protonation process. A density functional study*. Theoretical Chemistry Accounts, 1999. **101**(6): p. 388-395.
78. Madjarova, G., et al., *Selectivity Descriptors for the Michael Addition Reaction as Obtained from Density Functional Based Approaches*. The Journal of Physical Chemistry A, 2005. **109**(2): p. 387-393.
79. Mineva, T., *Selectivity study from the density functional local reactivity indices*. Journal of Molecular Structure: THEOCHEM, 2006. **762**(1): p. 79-86.
80. Tzonka, M. and H. Thomas, *Orbital hardness tensors from hydrogen through xenon from Kohn–Sham perturbed orbitals*. International Journal of Quantum Chemistry, 2006. **106**(6): p. 1396-1405.
81. Ohno, K., *Molecular Orbital Calculations of Φ Electron Systems*, in *Advances in Quantum Chemistry*, P.-O. Löwdin, Editor. 1967, Academic Press. p. 239-322.
82. Bjelkmar, P., et al., *Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models*. Journal of Chemical Theory and Computation, 2010. **6**(2): p. 459-466.
83. Anisimov, V.M., et al., *Polarizable Empirical Force Field for the Primary and Secondary Alcohol Series Based on the Classical Drude Model*. Journal of Chemical Theory and Computation, 2007. **3**(6): p. 1927-1946.
84. Nikitin, A.M. and A.P. Lyubartsev, *New six-site acetonitrile model for simulations of liquid acetonitrile and its aqueous mixtures*. Journal of Computational Chemistry, 2007. **28**(12): p. 2020-2026.
85. Guàrdia, E., et al., *Comparison of Different Three-site Interaction Potentials for Liquid Acetonitrile*. Molecular Simulation, 2001. **26**(4): p. 287-306.
86. Soper, A.K. and M.G. Phillips, *A new determination of the structure of water at 25°C*. Chemical Physics, 1986. **107**(1): p. 47-60.
87. Jerie, K., et al., *Structure of aqueous solutions of acetonitrile investigated by acoustic and positron annihilation measurements*. Acta Physica Polonica-Series A General Physics, 2005. **107**(5): p. 826-831.
88. Xu, H., H.A. Stern, and B.J. Berne, *Can Water Polarizability Be Ignored in Hydrogen Bond Kinetics?* The Journal of Physical Chemistry B, 2002. **106**(8): p. 2054-2060.
89. Soper, A.K., F. Bruni, and M.A. Ricci, *Site–site pair correlation functions of water from 25 to 400 °C: Revised analysis of new and old diffraction data*. The Journal of Chemical Physics, 1997. **106**(1): p. 247-254.
90. Noskov, S.Y., G. Lamoureux, and B. Roux, *Molecular Dynamics Study of Hydration in Ethanol–Water Mixtures Using a Polarizable Force Field*. The Journal of Physical Chemistry B, 2005. **109**(14): p. 6705-6713.

91. Salamatova, E., et al., *Hydrogen bond and lifetime dynamics in diluted alcohols*. Physical Chemistry Chemical Physics, 2017. **19**(41): p. 27960-27967.

Chapter 4. Solvent effect on the structure and dynamics of lignocellulose model compounds

Contents

Abstract	128
1. Introduction.....	128
2. Computational details	130
2.1 Models	130
2.2 Classical Molecular Dynamics simulations.....	131
2.3 Density Functional Theory calculations	136
2.4 Molecular dynamics simulations at the QM/MM level (BOMD/MD).....	136
3. Results and discussion.....	137
3.1 Modelling of the interaction of water-organic solvents with cellulose nanocrystal	137
3.2 Dynamics of xylan tetramer model in water-organic solvents	147
3.3 Dynamics of lignin in water-organic solvents	153
3.4 Polarization effects on lignocellulose polymers, evaluated at the QM (DFT) level.....	160
3.5 MD and BOMD/MD evaluation of the stability of lignin dimer structures	164
3 Conclusions	167
References	169

Abstract

The effect of binary water-organic solvents on the conformations of cellulose, hemicellulose, and lignin was studied with molecular dynamics simulations (MM), quantum chemistry calculations (QM), and hybrid QM/MM simulations considering ethanol and acetonitrile as organic co-solvents in binary water-organic solvent mixtures. The polymers interactions were examined in the pure water, ethanol and acetonitrile solvents as well. At the classical level, we assessed the evolution of the polymers using the following metrics: (i) interaction energies; (ii) number of H-bonds; (iii) site-to-site radial distribution functions (RDFs) and cumulative numbers (cn); and (iv) torsion angles analysis and solvent accessible surface areas (SASA). Furthermore, at the QM and QM/MM level, we analysed the DFT potential energies and the polarization effect on cellulose and lignin atoms. The computed cumulative numbers used to estimate the number of different types of solvent molecules in the polymer's coordination shell are found to correlate with the profiles of the polymer-solvent and polymer-polymer interaction energies. The results suggest that energy gain due to H-bond interactions with the solvent greatly compensates the loss of intramolecular H-bonding, allowing the existence of low energy conformers in the presence of water. Moreover, the lignin-solvent interactions pass through an optimization, and the most stable and more solvent-interacting structure is found in ethanol-water mixtures, i.e., 75 wt% EtOH.

1. Introduction

The variety of the organic polymers present in wood –i.e., lignocellulosic biomass–, makes it desirable for a broad variety of applications [1, 2]. However, the complex and versatile nature of lignocellulose presents a challenge and limits the understanding of the factors responsible for its properties and behaviour when in contact with a solvent medium, such as shrinking and swelling [3]. Molecular-scale studies achieved by computational methods often in combination with experimental investigations have been promising to understand the conformations of the principal wood components –i.e. cellulose [4-16], hemicellulose [17-21], and lignin [22-31]-, but the solvent effect on lignocellulosic biomass remains widely unexplored [32-40].

The ordered and crystalline structure of cellulose has facilitated to some extent studies of the bulk and the solvent-accessible surfaces of this material. Molecular dynamics (MD) simulations along with experimental studies have provided valuable information about the cellulose bulk structure [4, 5], the orientation of the hydroxymethyl group of pure crystalline, amorphous cellulose, and the hydrophilic /hydrophobic cellulose surface site interactions with water [6-16], and the degree of water structuring due to hydrogen bonding with the cellulose surfaces [14, 15]. Both QM [4] and MD [5] have shown that the orientation of the hydroxymethyl group is determined by the distinct cellulose-solvent interactions. It has been observed that in amorphous cellulose [16] the torsion angles explore a very large angular range and hydroxymethyl groups explore mostly the gg and gt orientations, in contrast with the crystal structure.

Although the study of lignin and hemicellulose has proven to be a considerable challenge due to the lack of a regular and ordered structure, theoretical studies have provided valuable information about the most common units and linkages present in these polymer structures [17-31, 41, 42]. The most abundant hemicellulose backbone components have served as a model to study the conformation and flexibility of hemicelluloses [17-21]. Simulations of low molecular weight of polysaccharides [17] have shown that conformation of the glycosidic linkage is highly flexible in xylose and can undergo conformational changes that depend on the surrounding environment. It has been reported that free xylan adopts a left-hand threefold helical (3_2) conformation in water, whereas a twofold helical (2_1) conformation is stabilized in the presence of cellulose [18-20]. This conformation emerges from the distortion of the OH3 rotamer allowing water molecules to move freely around the glycosidic region and exchange with bulk water [19]. Moreover, simulations of polysaccharides [21] have shown lower solute-water and solute-solute hydrogen bonds and less hydrated surfaces in xylose when compared to glucose polymers.

Meanwhile, the work on lignin includes studies of its structure [22-24], temperature effects [25, 26], and lignin aggregation [27]. The β -O-4 linkage in lignin represents the predominant inter-unit linkage, and it has been used as a model for studying major conformation features such as H-bonding and flexibility [22, 23]. Both experimental and theoretical studies suggest that β -O-4 structures are flexible molecules that can adopt a large number of conformations [23, 28-30]. Durbeej and Eriksson [31] have shown that intramolecular hydrogen bonding has a stabilizing effect on different dimeric structures in the absence of external hydrogen bonds. Moreover, it has been observed the tendency of softwood lignin polymers (degree of polymerization from 6 to 41) to self-aggregate in aqueous medium, even at high temperature. [24].

Whereas an extensive work on lignocellulosic components (mostly cellulose) in water has been carried out, only a hand of computational studies approaches the interactions of the above-mentioned wood components with non-water or mixed solvents [32-40]. Simulations of a methyl β -cellobioside [32] have shown that the intramolecular O3-H...O5' H-bond over the glycosidic linkage is present in apolar solvents but absent in polar solvents. More recently, simulation of cellulose with ethylenediamine (EDA) [33] shows that the solvent strongly disrupts the naturally occurring hydrogen bond pattern, including the well-preserved O3-H...O5' H-bond, as O3 prefers to donate to EDA. Simulation of glucose solvation in water, THF, DMSO, and DMF [37] showed that the organic solvents compete with water to be in the first solvation shell of glucose and a significant amount of water is pushed to the second solvation shell. Moreover, a phase separation of binary water-organic solvents (THF, acetone, ethanol, and γ -valerolactone) on cellulose surfaces was reported [38, 39]. Hemicelluloses have been reported experimentally to turn into a gel with the addition of ethanol [43, 44], and previous MD have also reported a decrease of xylan's diffusion coefficient in ethanol-water mixture [40]. Recently, all-atom MD simulations [35, 36] showed that tetrahydrofuran (THF) - a polar aprotic ether - preferentially solvates lignin, which shifts the equilibrium conformational distribution from a crumpled

globule to coil lignin conformer. Whereas pure water is a bad solvent for lignin, the THF-water cosolvent acts as a ϵ -solvent, thus providing a mechanism for the observed lignin solubilisation [35].

Thus, the aim of this chapter is to study the conformation and dynamics of wood components models – cellulose, hemicellulose, and lignin- in pure and mixed solvents. In the present chapter, we assess the evolution of cellulose, xylan, and lignin models in the presence of water-ethanol and water-acetonitrile binary solvents at 25, 50, and 75 wt% concentrations. The results are compared to those obtained for the respective pure mono-component solvents. To achieve the scope we carried out a detailed analysis of the MD dynamics of the polymers and the solvents using the following metrics: (i) torsion angles analysis and solvent accessible surface areas (SASA); (ii) number of H-bonds and H-bond lifetimes; (iii) site-to-site radial distribution functions (RDFs) and cumulative numbers (cn); and (iv) interaction energies. The distance and angle between the lignin monomers has been measured additionally to describe the stacking, and the xylan's diffusivities have been computed to measure its mobility.

2. Computational details

2.1 Models

A model of truncated cellulose crystal (cellulose nano-crystallite) was built with seven cellulose chains, eight-monomers long (56 anhydroglucose units, Figure 4.1A), using the crystallographic structure of cellulose I β reported by Nishiyama et al.[45] and built with the cellulose-builder tool of Gomes and Skaf [46]. Oxygen terminal residues were capped with hydrogen atoms, and carbon terminal residues were capped with OH groups to obtain a finite chain. A single cellulose chain with eight monomers was also studied (Figure 4.1B), but only the QM results are presented for the sake of concise presentation. The model reproduces correctly the intramolecular and intermolecular hydrogen bond patterns expected for cellulose I β chains[47].

Hemicellulose is regarded as a tetramer of xylose, the most abundant hemicellulose in plants [48]. The backbone of xylan consists of poly β -(1 \rightarrow 4) linked D-xylopyranosyl units, to which are attached a variety of side group. Here, we consider a xylan tetramer (Figure 4.1C) and its structure was obtained with the GLYCAM Carbohydrate builder [49]. Lignin is represented by a dimer and a tetramer of guaiacyl (G) monomer units with oxygen terminal residues capped with hydrogen atoms (Figure 4.1D-E, respectively). The G monomers are β -O-4 linked, this being the most frequent linkage found in natural lignin, connecting the β -carbon in one unit with the oxygen of the C4 on the phenyl ring of the other unit. The polymers were centred in cubic boxes, leaving at least one nm to the longest side of the model to avoid interactions between the cellulose or lignin models and their images in the neighbouring boxes. We used this rim spacing to determine the size of the box. Each system was further solvated with the number of solvent molecules needed to fill the box size, as summarized in Table 4.1. Each organic solvent, i.e. ethanol and acetonitrile, was examined at four

different concentrations: pure, 75 wt%, 50 wt%, and 25% wt% of organic solvent, and compared with the simulations in pure water.

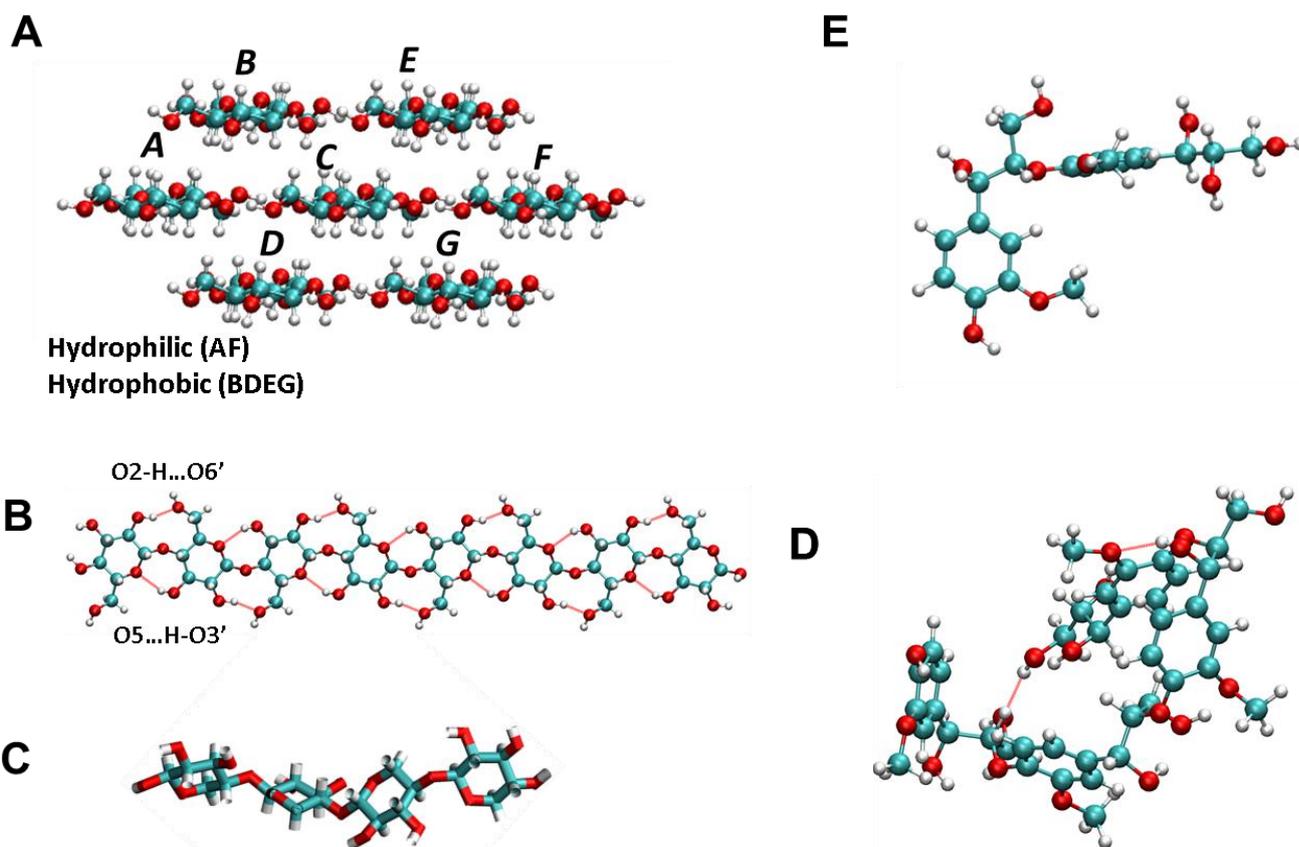


Figure 4.2 (A) Cellulose nanocrystal model with 7 chains (ABCDEFG). The solvent-accessible surface (SASA) of chains BDEG is referred to as the hydrophobic surface because the exposure to the solvent of their less polar aliphatic hydrogen atoms of the glucose rings and the glycosidic bonds between monomers. On the other hand, SASA of chains AF is referred to as the hydrophilic surface because the polar hydroxyl groups of the glucose monomers are exposed to the solvent. (B) Individual cellulose chain with 8 glucose monomers. Intramolecular hydrogen bonds O2-H...O6' and O5...H-O3' cover the glycosidic bond and keep the chain planar. (C) xylose tetramer formed with a β -(1 \rightarrow 4) bonds. (D) Optimized structure in gas phase from quantum-chemistry DFT-PBE calculations of a lignin dimer model formed by two guaiacyl (G-G) monomers, linked with a β -O-4 bond. (E) Structure of a guaiacyl lignin tetramer obtained from a MD simulation in the gas phase. Colour code CPK and Bonds representations: oxygen (red), hydrogen (white), and carbon (cyan). Hydrogen bonds are shown in dashed red lines.

2.2 Classical Molecular Dynamics simulations

All-atom Molecular Dynamics (MD) simulations of each system described in Table 4.1 were carried out using the GROMACS package version 2016.3[50-54], along with the 4-sites Transferable Intermolecular Potential (TIP4) for liquid water[55-57], the CHARMM36 additive force field[58, 59], and the CHARMM-compatible force field for lignin[60, 61]. Solvent structure for the organic solvents was available at the GROMACS molecule and liquid database[62]. For each simulation box, energy minimization was performed using the steepest descent algorithm until convergence to a tolerance of $100 \text{ kJ mol}^{-1} \text{ nm}^{-1}$. After minimization, restrained simulations were performed for 200 ps at 298.15 K to allow solvent equilibration around the

polymers. Afterwards, 20-ns MD simulations were performed with a frame-saving rate (for analysis) of 1 ps, in order to study the interaction of cellulose and the cellulose-lignin complex in the solvent mixtures. 10-ns MD simulations were performed for the lignin, xylan, and the xylan-lignin complex due to the reduced size of the models in comparison with the cellulose nanocrystal. Temperature and pressure coupling was handled using the leap-frog stochastic dynamics integrator and the Parrinello-Rahman method, respectively. Initial velocities were generated from a Maxwell distribution at 298.15 K, and the isothermal-isobaric (NPT) ensemble was considered for data collection. Neighbour searching and short-range nonbonded interactions were handled with the Verlet cut-off scheme. Electrostatics were treated with the Fast smooth Particle-Mesh Ewald (SPME) method, with a Coulomb cut-off of 1.2 nm, a fourth-order interpolation and Fourier spacing of 0.12 nm. Vander Waals (vdW) interactions were treated using the Lennard-Jones potential with a cut-off distance of 1.2 nm. Simulations were carried out in an Intel Xeon CPU with 2.10Gz with 32 logical cores.

Table 4.3. Configuration of simulated systems and equilibrium size of simulation boxes for the polymers studied in this work:(i) cellulose nanocrystal (7 chains, 8 glucose monomers each), (ii) cellulose chain (8 glucose monomers), (iii) xylan tetramer, (iv) lignin dimer and (v) lignin tetramer. Solvents studied include water, ethanol (EtOH), acetonitrile (ACN), and their binary aqueous mixtures.

Model wood component	Solvent system	Cosolvent fraction (wt%)	Number of cosolvent molecules	Number of water molecules	Cubic box side length (nm)	Volume (nm ³)
Cellulose nanocrystal (7 chains, 56 glucose monomers)	water	0	0	9764	6.73	306.03
	EtOH-water	25	1373	10559	7.69	456.32
		50	2607	6674	7.69	456.34
		75	3693	3151	7.69	456.40
	ACN-water	25	1512	10304	7.69	456.11
		50	2829	6451	7.69	455.88
		75	3990	3033	7.69	455.79
	EtOH	100	4321	0	7.54	429.92
	ACN	100	4874	0	7.60	440.43
	Cellulose chain (8 monomers)	water	0	0	7872	6.20
EtOH-water		25	1156	8877	7.23	377.42
		50	2196	5621	7.23	377.48
		75	3111	2654	7.23	377.64

Model wood component	Solvent system	Cosolvent fraction (wt%)	Number of cosolvent molecules	Number of water molecules	Cubic box side length (nm)	Volume (nm ³)
	ACN-water	25	1273	8712	7.23	377.24
		50	2382	5433	7.22	377.04
		75	3360	2555	7.23	377.17
	EtOH	100	2314	0	6.05	211.76
	ACN	100	2633	0	6.12	229.20
Xylose tetramer	water	0	0	2469	4.22	75.47
	EtOH-water	25	224	1717	4.19	73.37
		50	425	1087	4.19	73.41
		75	602	513	4.19	73.48
	ACN-water	25	246	1685	4.19	73.48
		50	461	1051	4.19	73.35
		75	650	494	4.19	73.37
	EtOH	100	0	648	3.99	63.77
ACN	100	0	737	4.04	65.9	
Lignin dimer	pure water	0	0	1448	3.54	44.55
	EtOH-water	25	132	1012	3.51	43.45
		50	250	641	3.51	43.41
		75	355	302	3.51	43.55
	ACN-water	25	145	993	3.51	43.38
		50	272	619	3.51	43.46
		75	383	291	3.51	43.46
	EtOH	100	399	0	3.4	39.48
ACN	100	456	0	3.45	41.01	
Lignin	pure water	0	0	3362	4.69	103.02

Model wood component	Solvent system	Cosolvent fraction (wt%)	Number of cosolvent molecules	Number of water molecules	Cubic box side length (nm)	Volume (nm ³)
tetramer	EtOH-water	25	307	2357	4.65	100.92
		50	583	1493	4.66	100.95
		75	826	705	4.66	101.03
	ACN-water	25	338	2313	4.65	100.83
		50	633	1443	4.65	100.91
		75	892	678	4.65	100.91
	EtOH	100	934	0	4.51	92.02
	ACN	100	1074	0	4.58	96.14

The structure, dynamics, and interactions of the wood component models with the pure and water-organic solvent mixtures were characterized using the incorporated tools within GROMACS. To quantify the solvent effect, we computed intermolecular energies, radial distribution functions $g(r)$, cumulative numbers $cn(r)$, and the average number of hydrogen bonds. The polymer structures were characterized by means of solvent accessible surface areas (SASA)[63], and characteristic torsion angles. For the lignin dimer, we computed the distance between the centres of mass of the two-chromophore rings, d_{COM} , as an indicator of stacking between the chromophores. The angle between the planes of the two chromophore rings, α , is also measured to differentiate between co-planarity ($\alpha \sim 0^\circ$, $\alpha \sim 180^\circ$) and T-stacking ($\alpha \sim 90^\circ$). Each of these descriptors was calculated for all the water-organic solvent concentrations considered in this study. Site-to-site $g(r)$ were computed for the water-polymer and organic solvent-polymer pairs (see Fig. 4.2 for atom numbering) considering the polymer sites as follows: in cellulose and xylan, we considered the hydroxyl (O2 and O3) and hydroxymethyl (O6, only cellulose), the glycosidic bond (O4), and the monosaccharide ring (O5)). In lignin, we studied the solvation around the β -O-4 bond (O α , O β , and O γ) and the methoxy groups (OMe). The solvent sites considered here are the water O, ethanol O, and acetonitrile N, along with the methyl C of ethanol and acetonitrile. Here, $g(r)$ is normalized by the number of reference points and the volume of the shell. Thus, $g(r)$ is expressed as number density (atoms/nm³) per monomer, and it tends to the bulk density of the particle as r increases.

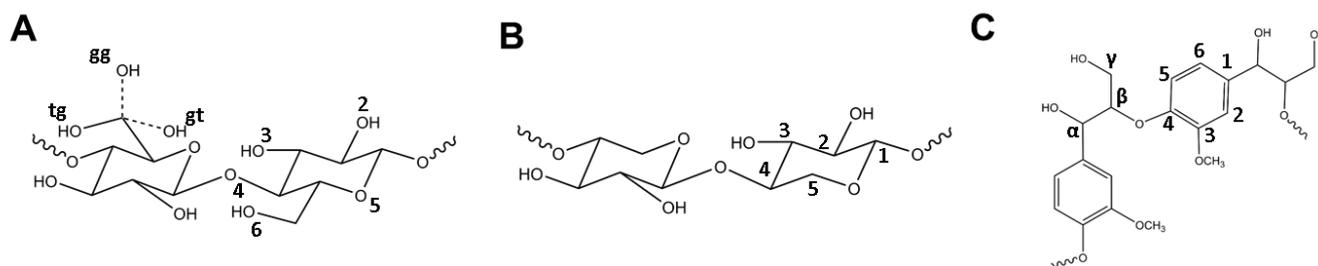


Figure 4.3. (A) Cellulose topology and oxygen notation. Three hydroxymethyl group conformers tg (bulk, plain line), gt, and gg (dashed lines), and their gauche or trans positions of O6 relative to both O5 and C4 are present in cellulose. Ideal tg, gt, and gg conformations are characterized by the hydroxymethyl torsion angle $\omega = 180^\circ$, 60° , and -60° with respect to O5, respectively[45]. (B) Topology and carbon notation of fundamental unit of xylan hemicellulose, a xylosyl dimer, bonded with a β -(1 \rightarrow 4) linkage. (C) Topology and carbon notation of a guaiacyl dimer (G-G) connected with a β -O²-4' linkage.

The integration of $g(r)$ from 0 to r , $cn(r)$, gives the cumulative number of particles within a distance r from the cellulose/lignin surface atoms. The cumulative numbers were obtained from the integration of the $g(r)$ up to a 1.5 nm correlation distance. H-bonds were calculated using a geometrical criterion with a maximum donor-acceptor distance of 0.35 nm and a hydrogen-donor-acceptor angle of 30° , and they were further normalized by the number of monomers. The forward H-bond lifetimes were computed from the autocorrelation function of each H-bond type, using the theory of Luzar and Chandler [64, 65], as described by van der Spoel et al. [66]. Solvent accessible surfaces were computed using a solvent probe radius of 0.14 nm. The angle between lignin rings planes was evaluated by means of the angle between the normal of the planes defined by the C1, C3, and C5 (Fig. 4.1B) atoms in each lignin monomer. Diffusivities were obtained from the evolution of the mean squared displacement and using the following equation

$$D = \lim_{t \rightarrow \infty} \frac{1}{6} (|r(t) - r(0)|^2),$$

where t and $r(t)$ denote the time and the position of the centre of mass of the studied polymer at time t , respectively.

Furthermore, MD snapshots containing the representative polymer structures, studied here in each solvent, were extracted from the trajectories for a graphical description of the results. For this purpose, we performed a root mean square deviation (RMSD)-based cluster analysis. First, the RMSD of atom positions between all pairs in the polymer structure was determined. For cellulose and hemicellulose, we varied the RMSD cut-off between 0.10 to 0.15 nm in 0.01 nm increments and performed cluster analysis for each RMSD value. For the lignin dimers and tetramers, we varied the RMSD cut-off between 0.10 to 0.40 nm in 0.1 increments, and we included the 0.15 nm cut-off. For all RMSD clusters, we found similar representative structures as centroids of the most populated clusters. In cases when more than one cluster was significantly populated, we selected the centroid of the cluster closest to the end of the simulation.

2.3 Density Functional Theory calculations

Classical MD (MM) simulations are a promising and reliable tool to study the statistically averaged physical properties and dynamics of many diverse molecular systems. However, the bigger drawback persists despite the efforts to improve the force fields description of molecular systems. As the force fields are parameterized, regular MD does not allow the simulation of covalent bond breaking or formation or the polarization effects of the single atoms interacting with neighbour molecules or solvents. Therefore, some valuable information is lost due to the classical approximation. Thus, quantum mechanics (QM) calculations are desirable in order to elucidate the origin of the specific interactions of the solute-solvent systems. However, the accuracy of the QM calculations comes with an increase of demanded computational resources, and this puts a limit on the number of atoms that can be studied in a QM calculation.

To study the polarization effects, we performed quantum chemical DFT calculations of cellulose single chains and lignin tetramers. The cluster analysis centroid of cellulose single chains and lignin tetramers were extracted, along with a 3.5-Å radius solvent shell from any atom of the polymer surface. Snapshot refinement was then performed at the QM level using DFT based methods with the deMon2k computer package (for the calculations we used the developers 4.4.5, based on the descriptions in ref. [67] and 6.0.2 versions, used as well for the QM/MM dynamics simulations (*vide infra*). The detailed description of the last implementation in this version can be found in ref. [68]. Geometry optimizations were performed using the revised version of the PBE functional [69-71] referred throughout this manuscript as revPBE and LYP correlation functional [72]. The electrons in all the atoms were described with double- ζ quality wave functions (DZVP) [73], with and without the solvent shell. In deMon2k code, automatically generated auxiliary functions up to $l = 2$ (for the metal atom) and 3 (for H, C and O atoms) were used for fitting the density with the GGA functionals, thus decreasing the computational time with a comparable accuracy to density calculations from the molecular orbitals [74]. We computed the partial Mulliken charges of all atoms, and we report only charges of the O2, O3, O4, O5, and O6 oxygens in cellulose and the OMe, O α , O β , and O γ oxygens in lignin. Furthermore, we compared the obtained charges against the oxygen charges in the MM force fields [58-61].

2.4 Molecular dynamics simulations at the QM/MM level (BOMD/MD)

We performed Born Oppenheimer Molecular Dynamics (BOMD)/MD simulations of the lignin dimer structures obtained from the classical MD simulations in water, ethanol, acetonitrile, and 75% water-cosolvent (ethanol and acetonitrile) mixtures for 12 ps simulation time for each solvent composition. MD snapshots were extracted containing the lignin dimers, along with a 15-Å radius solvent drop. The lignin dimer was included in the QM layer, whereas the solvents atoms were treated classically with the OPLS-AA force field [75]. Moreover, the Onsager reaction field model is applied to represent the solvent as a continuum medium outside the solvent drop within BOMD/MD/PCM(Onsager) approach [68]. BOMD/MD/PCM(Onsager)

simulations were carried out at 300 K in the canonical ensemble using a Nosé-Hoover chain of 5 thermostats with frequencies of 400 cm^{-1} . The integration time-step was set equal to 1 fs and geometries at every 3rd fs were registered. The linear and the angular momenta of the whole lignin-explicit solvent system were conserved with a threshold of 10^{-8} , and therefore the rotational and translational degrees of freedom of the whole QM/MM system were kept frozen to avoid spurious translation or rotation in the space. These simulations were carried out with deMon2k.6.0.2 developer's version.

3. Results and discussion

3.1 Modelling of the interaction of water-organic solvents with cellulose nanocrystal

Various physical and chemical descriptors have been hitherto proposed in the literature to describe the biopolymer-solvent interactions (*vide supra*). First, information about cellulose-solvent interactions can be gathered from the conformational changes of the polymer upon coordination to different solvent molecules in the mixtures. To follow the variations of cellulose conformations, we measured the evolution of the solvent-accessible surface area, SASA, along the dynamics, as illustrated in Figure 4.3.

3.1.1 Solvent accessible surface area

The SASA of the cellulose nanocrystal (Fig. 4.1A) is initially 42.5 nm^2 and it increases upon solvation to an area that amounts to $\sim 45 \text{ nm}^2$ (Fig. 4.3A,B), which means there is a 4-6% increase upon interacting with the solvent. The average cellulose SASA in solvents is found to vary in a very small range in the considered solvents and their aqueous mixtures, but a general trend following the order: mixed solvents > water > ethanol > acetonitrile can be observed. Furthermore, the cellulose SASA in concentrated water-acetonitrile mixtures is in an average greater (Fig. 4.3C) than in water-ethanol mixtures. In the pure organic solvents; however, inversion is observed. Moreover, we identify a maximum SASA near $\sim 46 \text{ nm}^2$ in 50% acetonitrile solution. In ethanol mixtures, the highest SASA is observed at 25-50% ethanol concentration.

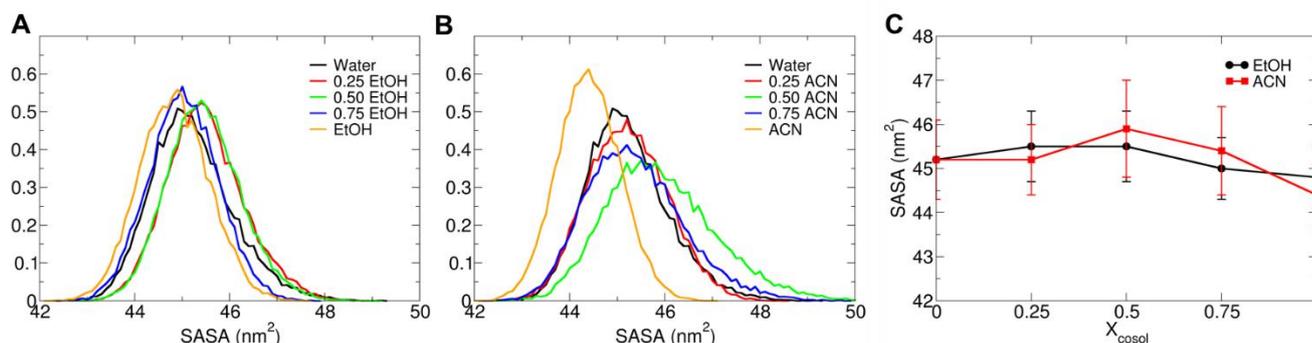


Figure 4.4. Solvent accessible surface area (SASA) distributions in (A) ethanol and (B) acetonitrile pure and mixed solvents, and their (C) averaged values along the dynamics.

3.1.2 Cellulose conformational changes upon solvent effect

We measured the characteristic torsion angles of cellulose to understand better the solvent effect on cellulose conformational changes. First, we measured the glycosidic torsion angle Φ (O5-C1-O4'-C4'), which describes the relative orientation of adjacent glycosyl residues in the same chain[45]. The maximum of the probability distributions of the glycosidic torsion angle Φ (not shown) falls near $\sim 88^\circ$ in water, whereas narrower distributions profiles (by $1-2^\circ$) are found for the organic pure and mixed water-organic solvents, but these small variation doesn't allow to draw any trend. These distributions are shifted by 10° in comparison to the Φ value for the crystal-like cellulose (-98.7°), and describe a twisting of the chain's backbone, as reported in previous MD simulations of cellulose [9-12, 14, 15]. It has been discussed that, in simulations of finite length crystals with explicit chain ends, the chirality amplification can result in a twist along the cellulose chain axis [11]. Nonetheless, Hadden et al. [12] suggest that the microfibril twisting is favoured by vdW interactions, and counteracted by both intra-chain hydrogen bonds and solvents effects at the surface. Furthermore, Matthews et al. [10] suggest that this event seems to be promoted by surface-mediated conformational changes, which result in an expanded unit cell upon conformational change of the hydroxymethyl group. It is expected that this twisting has an effect on the cellulose SASA; however, the backbone access to solvent is limited. We, therefore, suggest that changes of orientation of the hydroxymethyl and hydroxyl groups of the cellulose chain, driven by selective water-organosolv interactions are more likely to be responsible for conformational changes of the cellulose backbone in the surface.

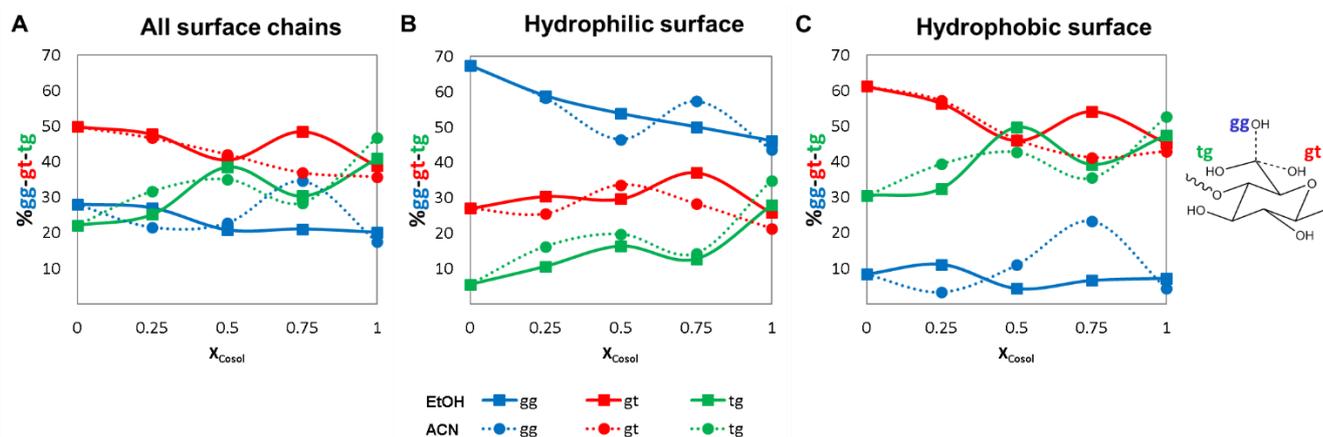


Figure 4.5. Hydroxymethyl dihedral conformer (gg-gt-tg) occurrence (in %) in (A) all surface cellulose residues, (B) glucose monomers in hydrophilic cellulose surfaces (chains AF) only, and (C) glucose monomers in hydrophobic cellulose surfaces (chains BDEG) only, in pure organic and mixed water-organic solvents. For the gg, gt and tg conformers the following colour code is used: gg (blue), gt (red), tg (green). The symbols used for the two organic Cosolvents are: ethanol, EtOH (square, full line); acetonitrile, ACN (circle, pointed line)

For this reason, we monitored the characteristic hydroxymethyl torsion angle ω , $\omega = \text{O5-C5-C6-O6}$ [45], as illustrated in Figure 4.2A. Three low-energy conformations are possible: tg, gt, and gg, referring to the gauche or trans positions of O6 relative to both O5 and C4 in cellulose. Ideal tg, gt, and gg conformations are characterized by $\omega = 180^\circ$, 60° , and -60° , respectively [45, 76, 77]. In the tg and gt conformations, the position of the C6-O6 bond is equatorial relative to the glucose ring. In the gg conformation, on the contrary, the C6-O6 bond is axial. In the bulk crystalline cellulose I β [45, 46], the hydroxymethyl groups adopt the tg conformation with an average value of $\omega = 169.4^\circ$.

We first computed the probability distributions of ω at each cosolvent concentration. Then, angle intervals were attributed to each conformation in order to compute the gg-gt-tg conformer occurrence frequency at each solvent concentration. These angle intervals were defined as following: gg= $(-120, 0)$, gt= $(0, 120)$, and tg= $(\pm 180, \pm 120)$. Figure 4.4A shows the gg-gt-tg dihedral occurrence of all surface monomers (in %) at each organic cosolvent concentration. As it follows, the crystal-like buried tg conformer is more preserved in the presence of organic solvents (acetonitrile > ethanol) which correlates well with the observed lower cellulose SASA in the pure organic solvents. A transition of ω from tg to the equatorial gt, or to the more solvent-exposed gg is enhanced in the presence of water, increasing the access of the solvent to the cellulose surface. In water, the gt conformer prevails, and an increase of the gg conformers is observed. A non-linear behaviour is observed in the mixed solvents: the gg conformer is significantly favoured in the 75 wt% water-acetonitrile mixture, whereas the gt conformer is mostly favoured in water-ethanol.

Furthermore, due to the amphiphilic nature of cellulose chains in our model, we examined separately the hydrophobic and hydrophilic surfaces labelled as BDEG and AF chains (Fig. 4.1A), respectively. The dihedral conformation of the hydrophilic AF surfaces in pure and mixed solvents (Fig. 4.4B) follows the trend $gg > gt > tg$, e.i. in the AF chains, the hydroxymethyl groups have predominantly gg and gt conformations. The

hydroxymethyls of the hydrophobic surface chains BDEG, on the contrary, display mostly an equatorial orientation, following the order $tg-gt > gg$ (Fig. 4.4C). Thus, the hydroxymethyl groups of cellulose chains BDEG are mainly buried inside the cellulose nanocrystal. Consequently, these chains display more hydrophobic behaviour than cellulose chains AF, whose hydroxymethyl groups stand exposed to the surface. Furthermore, the hydroxymethyl group follows the $tg \rightarrow gg$ transition in the hydrophilic surface (chains AF) with the decrease of the organic component (or, equivalently, increase of water) in the binary solvent mixtures. At the hydrophobic surface, the tg and gt conformations dominate; thus, the hydroxymethyl groups in BDGE chains remain mostly buried.

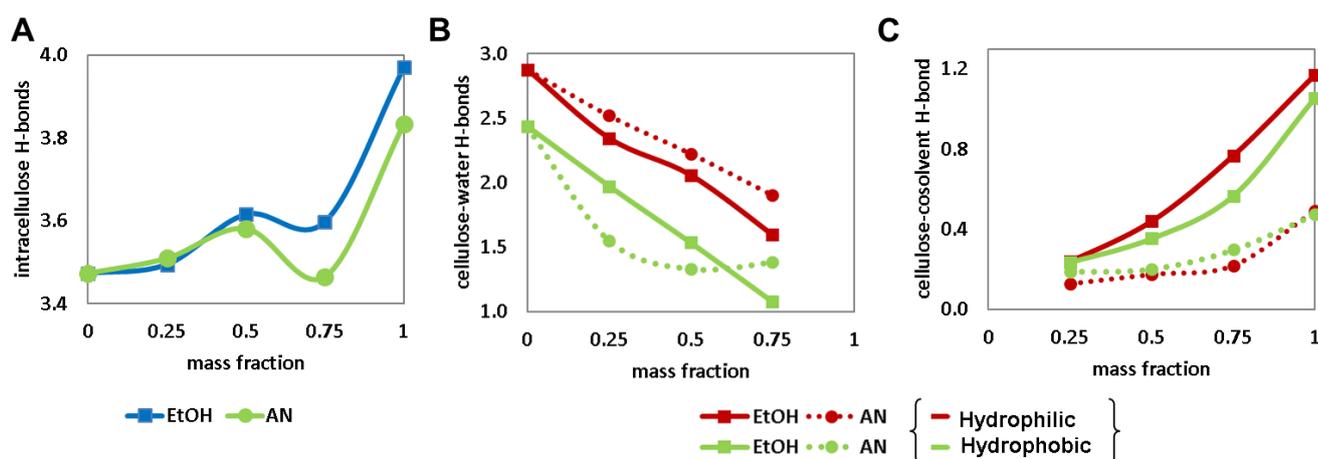


Figure 4.6. (A) Average number of cellulose intra-molecular H-bonds per anhydroglucose monomer at different organic solvent concentrations. Average number of hydrogen bonds per monomer between cellulose and water (B) or cellulose and organic solvent (C) at different organic solvent concentrations at both hydrophilic (chains AF, red series) and hydrophobic surfaces (BDEG), green series). Ethanol: squares and full lines; Acetonitrile: circles and pointed lines.

The increased cellulose SASA and the above-mentioned backbone twist are a result of the disruption of the H-bond network of the surface nanocrystal. In the tg conformation, the hydroxymethyl group participates in an intramolecular H-bond with the adjacent glucose monomer ($O2-H \dots O6'$); therefore, it is less available to participate in bridging hydrogen bonds with the solvent [13], thus keeping the cellulose chains planar [45, 46]. Because the tg conformer dominates in the two pure organic solvents (acetonitrile > ethanol), we expect that the intra-chain and inter-chain H-bond network is better preserved in ethanol and acetonitrile than in water to their binary aqueous mixtures, where gg and the gt conformers are more frequent. To better quantify this, we monitored the average number of (i) intramolecular (intra-chain and inter-chain) cellulose H-bonds and (2) the intermolecular cellulose-water, and (3) cellulose-organic solvent H-bonds, using a geometrical criterium as described in the computational details.

3.1.3 Intermolecular H-bonds in cellulose and between cellulose and solvents molecules

The results concerning the three types of intermolecular H-bonds are summarized in Figure 4.5. First, we found a general tendency showing that the frequency of cellulose intramolecular H-bonds (Fig. 4.5A) slightly

increases from water to binary aqueous mixtures, but increases strongly in pure ethanol and acetonitrile. The only exception is H₂O-75wt% ACN mixture, where a minimum of cellulose H-bonds is observed. Further on, comparing the two organic solvents, a higher number of intramolecular H-bonds is obtained in ethanol than in acetonitrile. This correlates with the slightly lower number of tg hydroxymethyl conformations in pure acetonitrile ($X_{\text{Cosol}}=1$) in Fig. 4.4A. The gt conformations are also slightly less present in acetonitrile in ethanol, which suggests that in the absence of strong interactions with water, the surface cellulose chains are more reoriented towards the core of the nanocrystal in ethanol than in acetonitrile. This leads to an increase of the cohesion of the sub-surface layer and their participation in intra-chain H-bonds.

Furthermore, the intermolecular cellulose-water and cellulose-cosolvent H-bonds in each water-organic solvent system are summarized in Figures 5B and 5C, respectively. The cellulose-water interactions have a major contribution to the intermolecular hydrogen bonding. More cellulose-water H-bonds are found on the hydrophilic surfaces (chains AF) than on the hydrophobic surfaces (chains BDEG), as expected. The cellulose-water H-bonds in ethanol-water systems linearly decrease with the ethanol concentration, as a response to water displacement due to alcohol competition for H-bonding sites, as can be inferred from the cellulose-ethanol H-bonds in Figure 4.5B. The cellulose-water H-bonds in acetonitrile-water mixtures display a similar behaviour with higher occurrence frequency in the hydrophilic surface (AF), whereas at the hydrophobic surfaces (BDEG), the cellulose-water H-bonds strongly decrease upon addition of acetonitrile for all the water-acetonitrile mixtures. The frequency of cellulose-acetonitrile H-bonds (in Fig. 4.5C) is significantly lower than the cellulose-ethanol H-bond frequency, as expected in regard to the aprotic nature of acetonitrile. The higher frequency of H-bonds in the hydrophobic than in the hydrophilic surface suggests that acetonitrile has a higher affinity towards the hydrophobic surface, whereas ethanol shows a higher affinity for the hydrophilic surface.

Table 4.4. H-bond lifetimes for the cellulose-water (C-W) and cellulose-cosolvent (C-Cosol), interaction types, obtained from the H-bond autocorrelation functions of each H-bond type, for the hydrophobic (chains BDEG) and the hydrophilic (chains AF) cellulose surfaces, respectively.

Solvent mixture	HB lifetimes (ps)							
	Water-EtOH				Water-ACN			
	Hydrophilic		Hydrophobic		Hydrophilic		Hydrophobic	
X_{cosol}	C-W	C-Cosol	C-W	C-Cosol	C-W	C-Cosol	C-W	C-Cosol
0	9.8	-	12.3	-	9.8	-	12.3	-
0.25	18.6	21.5	23.6	34.5	17.6	9.7	22.6	28.6
0.50	32.2	30.3	32.6	30.3	20.3	9.9	30.7	22.1
0.75	49.0	44.0	57.9	57.4	20.7	10.4	24.7	27.2

1	-	43.8	-	94.5	-	12.0	-	41.7
---	---	------	---	------	---	------	---	------

In addition to the information we could gather from the number of H-bonds between various solvent molecules and cellulose, the lifetime of H-bonds can also provide knowledge about the solvent mobility around the cellulose surface. We, therefore, analyzed the lifetimes of the H-bonds in Fig. 4.5 from the autocorrelation functions and the results are summarized in Table 4.2. The lifetime of cellulose-water H-bonds increase with the organic fraction and longer-lasting H-bonds are observed in water-ethanol than in water-acetonitrile mixtures. Moreover, shorter lifetimes are observed for the H-bonds between water and the hydrophilic surface in comparison to the hydrophobic surface, which is in line with the higher H-bonds frequency between water and hydrophilic AF chains. This suggests richer water –AF interface. As we showed in the previous chapter, the water H-bond lifetimes are shorter when richer water interfaces exist, since this allows faster water exchange between the bulk and surface water. Furthermore, we also observe that the cellulose-water H-bond lifetimes in water-ethanol mixtures are longer lasting than in water-acetonitrile mixtures. As we showed in the previous chapter, this is explained by the higher frequency of H-bond formation between water and ethanol than with acetonitrile. Thus, we believe that the cellulose-water H-bond lifetimes are longer-lasting in ethanol mixtures than in acetonitrile ones because of the continuous interactions between water and ethanol that constrains the water movements. Finally, the cellulose-ethanol H-bond lifetimes are always longer-lasting than the cellulose-acetonitrile ones due to their higher frequency. This can be explained by the ethanol's capacity to act as an H-bond donor and H-bond acceptor, whereas acetonitrile is only an H-bond acceptor.

3.1.4 Solvents organization around the cellulose surface.

Other information about cellulose-solvents can be gathered from the correlation of the solvent and specific cellulose sites. The cellulose solvation is described by the radial distribution function (RDF), that describes the number of particles at distance r from a reference site. The coordination of solvent molecules to O2, O3, O4, O5, and O6 sites in cellulose is identified from $O_{\text{cell}}-X_{\text{solv}}$ ($X_{\text{solv}}=O_{\text{water}}$; O_{EtOH} and N_{ACN}) radial distribution functions $g(r_{O_{\text{cell}}-X_{\text{solv}}})$. In this analysis, we distinguish between the hydrophilic AF and hydrophobic BDEG chains by computing $g(r_{O_{\text{cell}}-X_{\text{solv}}})$ for glucose monomers in AF and in BDEG chains separately. The $g(r_{O_{\text{cell}}-X_{\text{solv}}})$ results for the pure water, EtOH and ACN solvents are summarized in Fig. 4.6 A, C, and E for O_{cell} belonging to AF cellulose chains and in Fig. 4.6 B, D and F for O_{cell} belonging to BDEG chains. A well-structured water layer forms around both hydrophilic and hydrophobic surfaces in agreement with previous cellulose-water RDF studies,[7] mainly because of water structuring around O6, O2 and O3 sites. This follows from the sharp peak in O6, O2 and O3 $g(r_{O_{\text{cell}}-O_{\text{Water}}})$ profiles at ~ 0.28 nm and a minimum of the first coordination shell near 0.35 nm. The coordination order is $O6 > O2 > O3$. The other two oxygen sites (O4 and O5) do not display well-defined coordination to water oxygen, except a somewhat increased coordination

between O5 and O(H₂O) in AF chains. Moreover, chains AF display a somewhat higher hydration than chains BDEG.

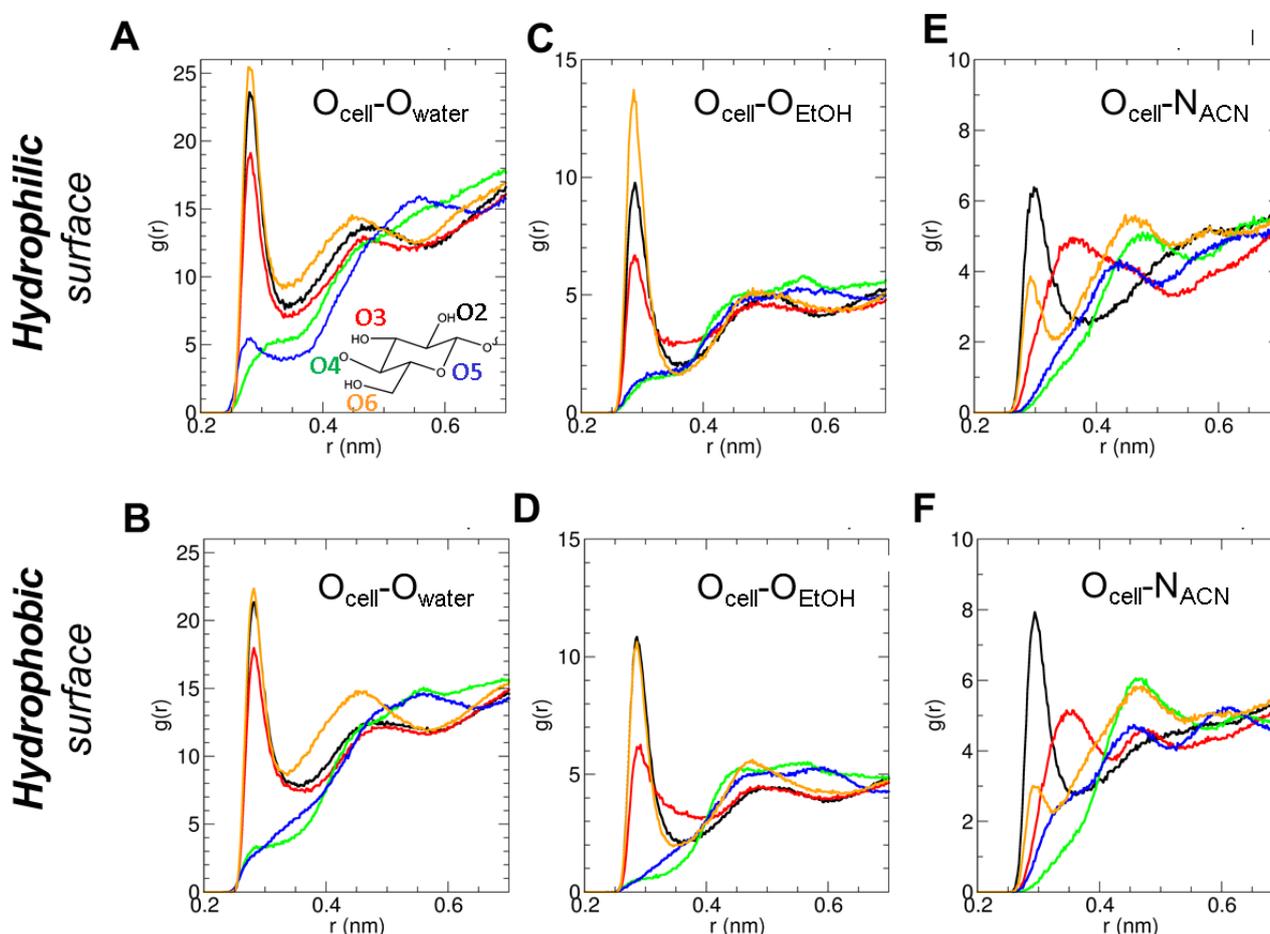


Figure 4.7 Solvation of cellulose in pure solvents. Radial distribution functions, $g(r)$, of the Ocellulose-Xsolvent pairs (Ocell-O_{water}, Ocell-O_{EtOH}, Ocell-N_{ACN}), for hydrophilic (chains AF) and hydrophobic (chains BDEG) surfaces. The reference site, X, in the solvents are water oxygen, EtOH oxygen, and ACN nitrogen. The colour code for the respective cellulose oxygen sites is O2 (black), O3 (red), O4 (green), O5 (blue), and O6 (orange), see the inset in A for oxygen numbering.

Similarly, O6, O2 and O3 cellulose sites play a role in the ethanol structuring around the cellulose surfaces, as concluded from the sharp peaks in $g(r_{\text{Ocell-OEtOH}})$ at ~ 0.29 nm and their minima of the first coordination shells between ~ 0.36 - 0.37 nm. As water, ethanol shows a slight preferential coordination to the hydrophilic AF cellulose surface. At the hydrophobic surface (chains BDEG), the peak intensity is decreased, and O6 coordination is reduced to the coordination with O2. Note that the number density between O_{cell} and O_{EtOH} decreases, as shown by the lower $g(r_{\text{Ocell-OEtOH}})$ intensity, in comparison with that with O_{water}, as expected from the larger ethanol molecular size. In acetonitrile solvent, the $g(r_{\text{Ocell-NACN}})$ profiles in Figure 4.6 C,F indicate that only cellulose oxygens O2 and O6 coordinate to N_{ACN} below 0.35 nm, at ~ 0.29 - 0.30 nm, whereas O3-N peak is at ~ 0.36 nm. The acetonitrile nitrogen prefers to coordinate to O2 instead to O6, contrary to water and ethanol solvents and has a higher preference to the hydrophobic cellulose surfaces, in which the coordination for O6 is significantly reduced.

The organization of the water-organic solvent mixtures at 25, 50, and 75 wt% is also described with $g(r)$ between $O_{\text{cell}}-X_{\text{solv}}$, and the results are reported in the Supporting Materials, Figures S4.1 and S4.2. These $g(r)$ display similar coordination-site trends in comparison with those found for the pure solvents, except for the $g(r_{O_{\text{cell}}-O_{\text{EtOH}}})$ in the cellulose hydrophobic surface (Fig S4.1B) and the hydrophilic surface at 25% ethanol. Particularly, the $O_{\text{cell}}-O_{\text{EtOH}}$ coordination is inverted and follows the order $O2>O6>O3$, in contrast with the ethanol coordination in the hydrophilic surfaces. This change of ethanol preference is most likely enhanced due to the presence of water, which competes with ethanol for the same coordination sites. Furthermore, we observed that the $O_{\text{cell}}-O_{\text{water}}$ $g(r)$ s (Fig S4.1-A and S4.1-A) display always higher water number densities in the hydrophilic surface than in the hydrophobic ones. While the water number density is expected to decrease with the addition of organic solvent, the water exclusion in acetonitrile mixtures (see Fig. S4.2-A) from the first cellulose solvation shell, especially in the hydrophobic surface, is stronger than for ethanol mixtures (See Fig. S4.1-A).

This observation is in agreement with the computed $cn(r)$ of water and organic solvent molecules within $r = 0.7$ nm from the glycosidic oxygen of cellulose, $O4$, as a function of organic solvent concentration, which is illustrated in Figure 4.7A. The number of water molecules in the first solvation shell (Figure 4.7A, top) in water-organic solvent mixtures decreases with the organic cosolvent concentration; however, there is a faster decrease of water molecules in the first cellulose solvation shell of AF (hydrophilic) chains when mixing water and acetonitrile, than in water-ethanol mixtures. Moreover, in water-acetonitrile mixtures, the water molecules coordinated to cellulose hydrophobic surfaces remain practically unchanged and even display a slight increase in the hydrophobic surface at 75 wt% cosolvent concentration. At this concentration, a somewhat large number of water molecules coordinate to the cellulose surfaces in the acetonitrile mixtures than in ethanol-water mixture (Fig. 4.8A, top). Also, we observe that only at 75% water-acetonitrile, the $O_{\text{cell}}-N_{\text{ACN}}$ coordination to $O6$ increases and it is equivalent to $O2$ (Figure S4.2B). On the other hand, in Figure 4.8A, bottom, the number of ethanol molecules surrounding the cellulose surfaces increases with the cosolvent concentration, whereas the numbers of ACN molecules surrounding the hydrophobic surface remains nearly the same for 50 and 75% cosolvent concentration. As evidenced by $cn(r)$, acetonitrile accumulates preferably at the hydrophobic cellulose sites, thus showing a more hydrophobic character. The $cn(r)$ values of ethanol coordinated to either AF or BDEG chains is similar; thus, ethanol has rather an amphiphilic character.

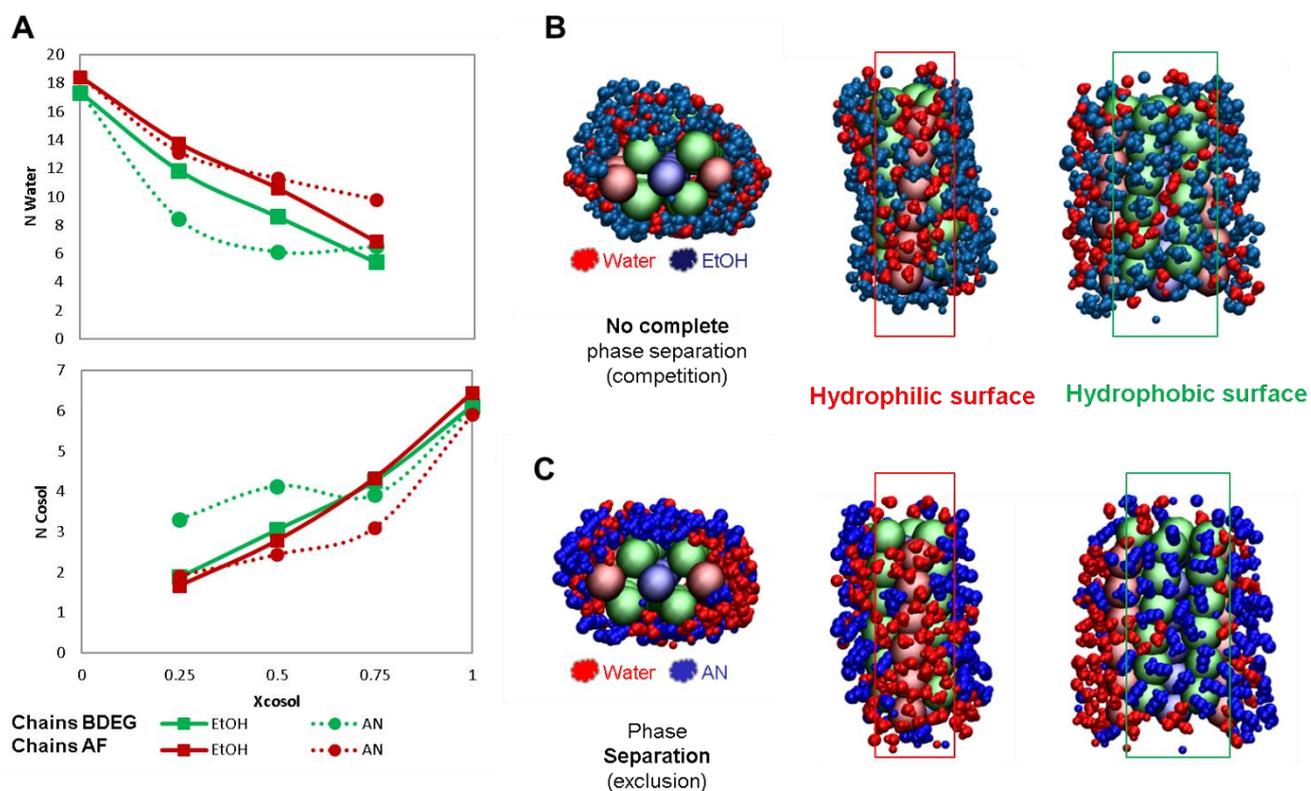


Figure 4.8.(A) Cumulative number (cn) of molecules of water (top) and organic cosolvent (Cosol, bottom) within 0.7 nm from the cellulose glycosidic oxygen, as a function of organic solvent concentration. Circles denote acetonitrile and squares denote ethanol cosolvents. Cumulative numbers at the hydrophobic surface sites (chains BDEG) are in green, and those at the hydrophilic surface (chains AF) are in red. Snapshots are illustrating 0.5 nm solvation shell around cellulose surfaces in 75 wt% ethanol (B) and 75 wt% acetonitrile (C) mixtures. Beads represent glucose monomers; water and cosolvents are represented by red and blue vdW surfaces, respectively.

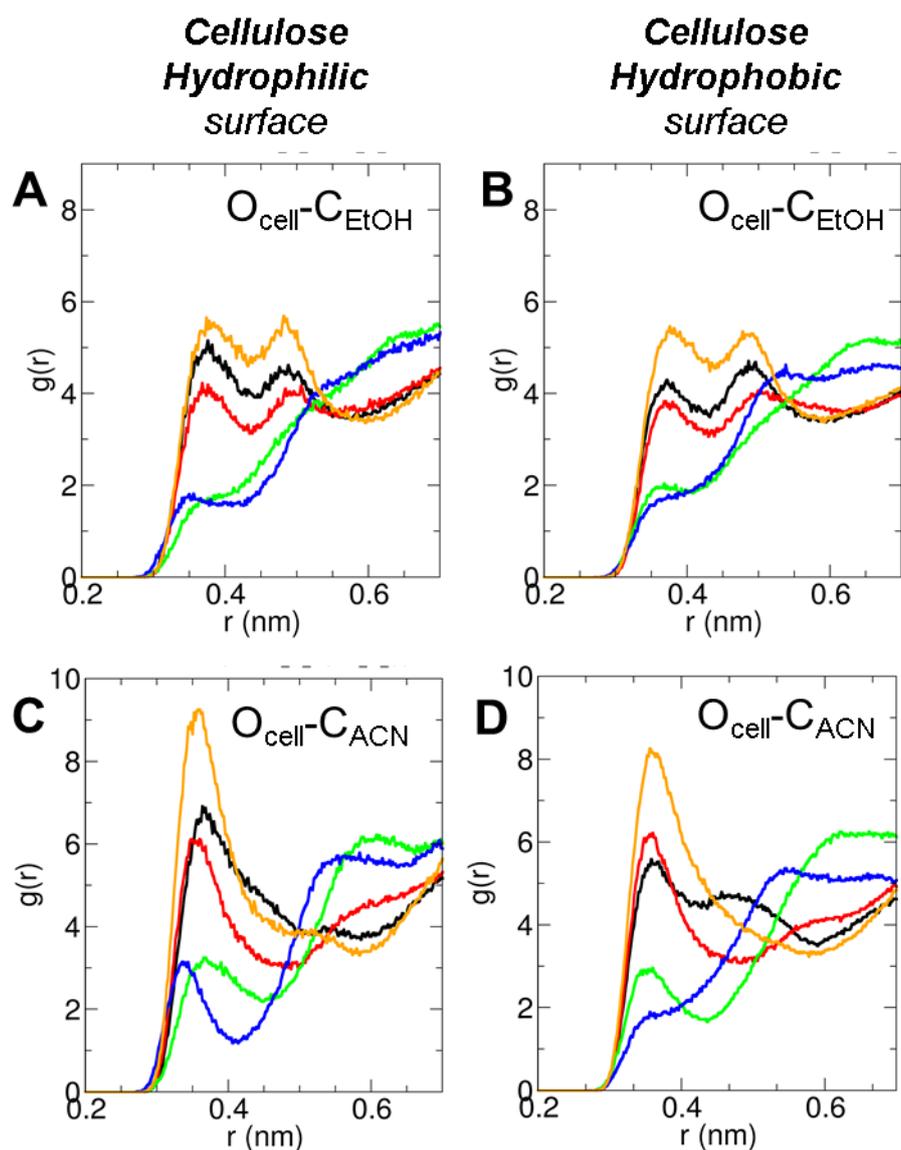


Figure 4.9. Ocellulose-Corganic solvent radial distribution functions, $g(r)$, between the cellulose oxygens (O_{cell}) in hydrophilic (chains AF) and hydrophobic (chains BDEG) surfaces, and methyl carbons of the ethanol, C_{EtOH} (A and B) and acetonitrile, C_{ACN} (C and D), respectively. Ocellulose colour code is O2 (black), O3 (red), O4 (green), O5 (blue), and O6 (orange), as in Figure 4.6.

The stronger water exclusion from the first solvation shell in cellulose hydrophobic surfaces for water-acetonitrile mixtures is explained by the coordination of the cellulose hydroxymethyl to the methyl group of acetonitrile, as illustrated by the $O_{\text{cell}}-C_{\text{Me}}$ $g(r)$ in Figure 4.8. This is possible due to the dipole-dipole interactions of acetonitrile molecules, which bring methyl groups closer to cellulose as acetonitrile nitrogens are H-bonded with cellulose hydroxyls. This is also enhanced by the local microheterogeneity of water-acetonitrile mixtures [78], which results in phase-separation when in contact with the different cellulose surfaces as illustrated in Figure 4.8B,C. This observation is consistent with the results of previous MD studies [38, 39]. Thus, an increase of the co-solvent phase could yield a more hydrated and more extended water shell

around cellulose hydrophilic surfaces. Ethanol mixtures, on the contrary, remain less separated as ethanol-water mixtures are more homogeneous and also share a similar affinity for a specific oxygen site in cellulose.

3.2 Dynamics of xylan tetramer model in water-organic solvents

3.2.1 Solvent effect on xylan conformational freedom

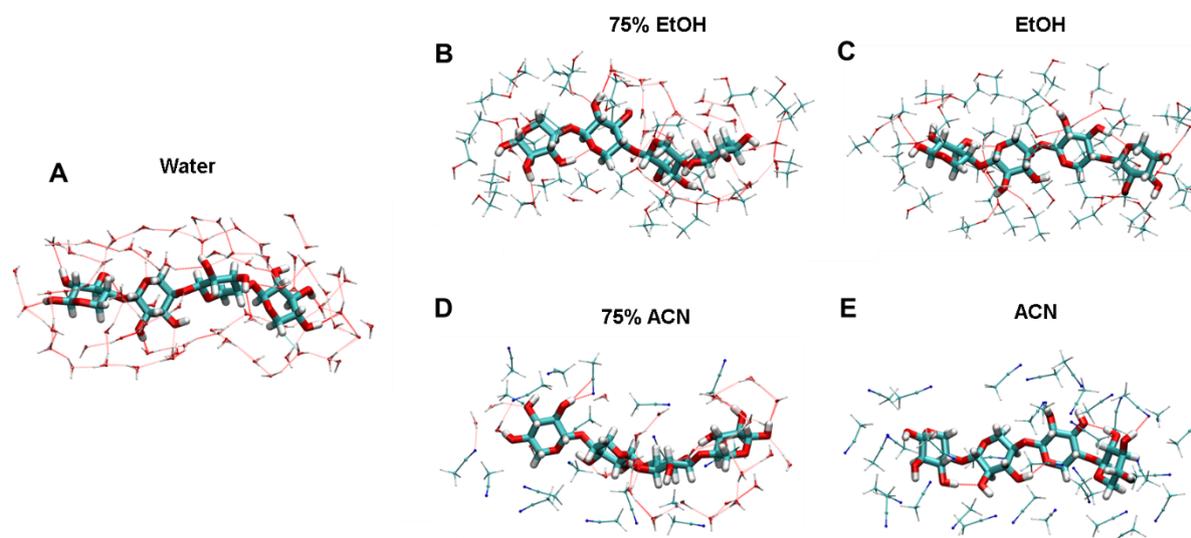


Figure 4.10. MD snapshots of representative xylose structures obtained from the cluster analysis in (A) water (B), 75 wt% ethanol-water, (C) ethanol, (D) 75% wt acetonitrile-water, and (E) acetonitrile, along with a 3.5 Å solvent shell. Bonds and CPK colour code: oxygen (red), hydrogen (white), carbon (cyan), and nitrogen (blue). Hydrogen bonds are shown as dashed red lines.

The simulated conformations of xylan in pure and binary mixtures (Figure 4.9) is described by the distribution of the SASA and the glycosidic torsion angles Φ (O5-C1-O4'-C4'), and Ψ (C1-O4'-C4'-C5'). The SASA distribution (Figure 4.10 A, D) falls at $\sim 8.0 \text{ nm}^2$ in all solvent mixtures and varies less than $\sim 0.1 \text{ nm}^2$ between all the ranges of solvent compositions. Despite this small variation, the hygroscopic nature of xylose results in a trend to find the lowest surface area in pure acetonitrile and ethanol, whereas a tendency of a more extended surface is observed in pure water and the diluted aqueous solvent mixtures. We tracked down the origin of this trend to the configuration of the glycosidic torsion angles Φ and Ψ , as shown in Figure 4.10 B,E and Figure 4.10 C,F, respectively. In comparison to cellulose, a hydroxymethyl group is absent in the xylose monomers, which participate in less intra-chain H-bonding and the chain's flexibility increases. The torsion angle Φ in the β -(1 \rightarrow 4) xylan tetramer has an average value of $-69^\circ \pm 4^\circ$ in the initial structure. During dynamics, the torsion angle Φ falls mostly near $\sim 74^\circ$ at all solvents and mixtures, thus preserving the β -(1 \rightarrow 4) linkage; however, there is a small amount of $\Phi \sim 55^\circ$ in pure and 75 wt% acetonitrile (Fig. 4.10 E), which reveals an occurrence of α -(1 \rightarrow 4) linkage in these two solvents.

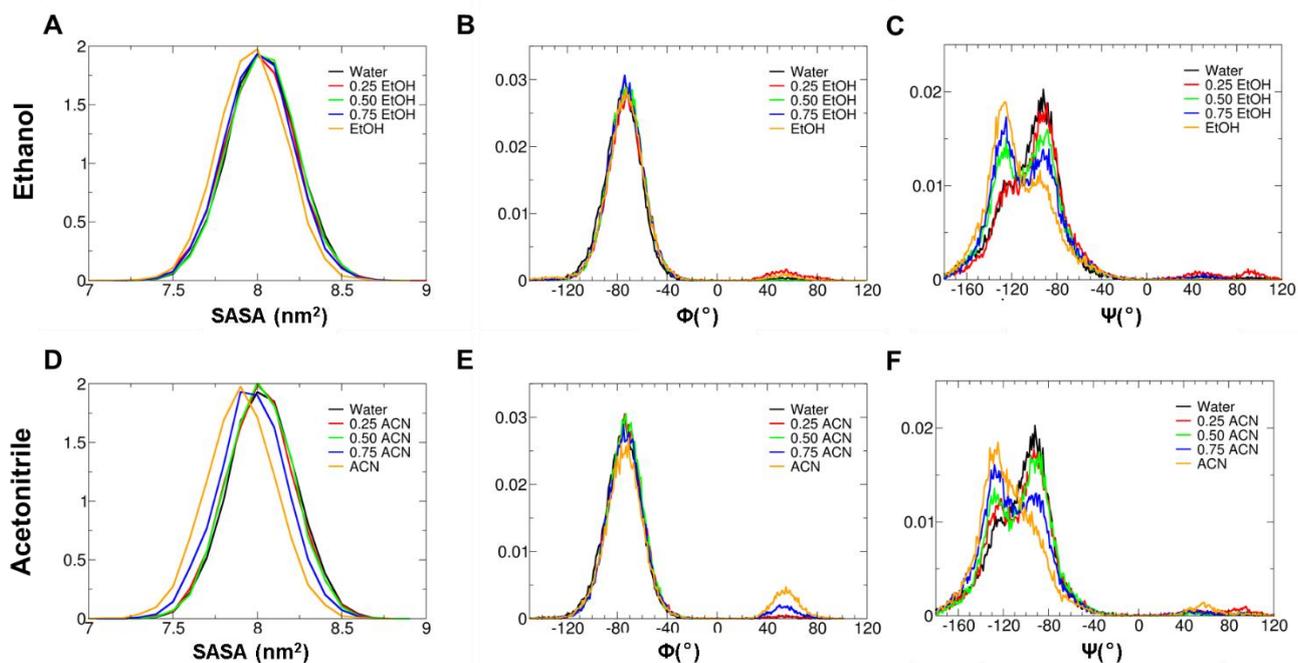


Figure 4.11. Distributions of solvent accessible surface area, SASA, of xylan tetramer (A), distribution of glycosidic torsion angle $\Phi = O5-C1-O4'-C4$ (B) and distribution of glycosidic torsion angle $\Psi = C1-O4'-C4'-C5'$ (C) at different ethanol concentrations. Distributions of SASA (D), the torsion angles Φ (E) and Ψ (F) of xylan tetramer at different acetonitrile concentrations

Furthermore, the torsion angle Ψ falls mainly in two distributions that describe the β -(1 \rightarrow 4) angle. In water and mixed solvents up to 50% concentration, the distribution falls near \sim -90°, corresponding to a left-hand threefold helical (3_2) conformation [18]; whereas in pure and concentrated organic solvents, the distribution mostly falls near \sim -130°, describing a twofold (2_1) conformation. In the latter solvents, xylan adopts a 2_1 conformation as a hydrophobic response to the pure and concentrated organic solvents, resulting in a more compact packing and slightly lower SASA. Our results of the simulated conformation of xylose in water are consistent with the previous MD results [19, 20], revealing that xylan adopts a 3_2 conformation, whereas the 2_1 conformation does not appear in water solvent.

3.2.1 Solvents organization around xylan tetramer.

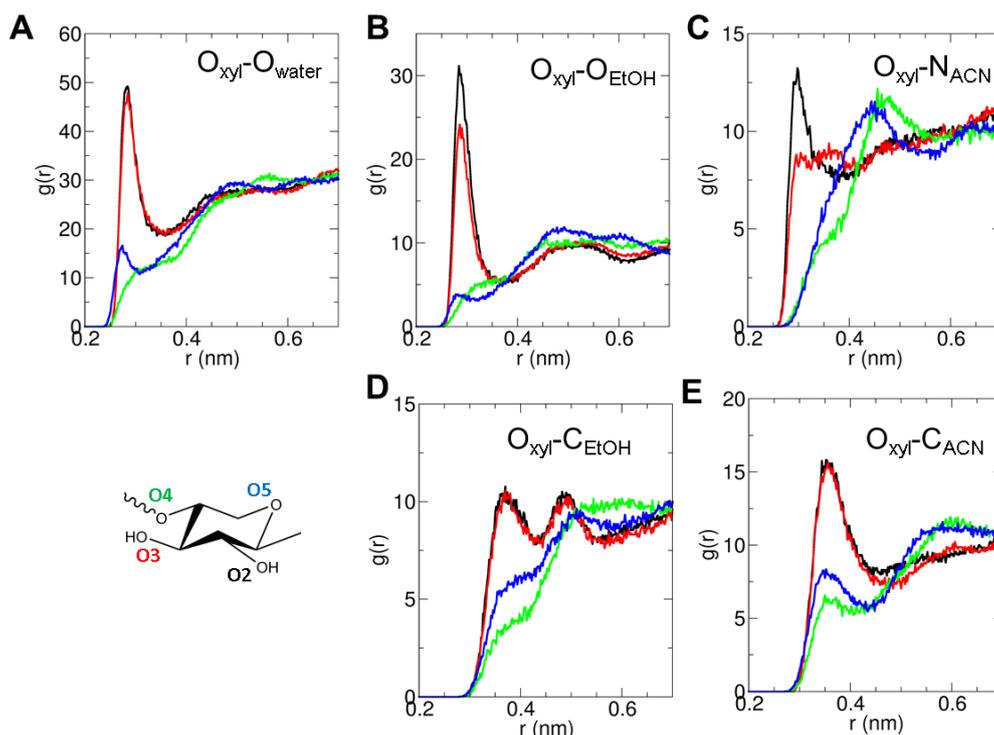


Figure 4.12. Oxylian-Xsolvent ($X_{\text{solvent}} = \text{O}_{\text{water}}, \text{O}_{\text{EtOH}}, \text{N}_{\text{ACN}}, \text{C}_{\text{EtOH}}$ and C_{ACN}) radial distribution functions, $g(r)$, between the xylan oxygens and (A) water oxygen, (B) ethanol oxygen, (C) acetonitrile nitrogen, and methyl carbons of the ethanol (D) and acetonitrile (E).. Oxylian colour code is O2 (black), O3 (red), O4 (green), and O5 (blue), as in Figure 4.2B.

The xylan solvation is described by the radial distribution function between the oxygens in xylan and X_{solvent} sites ($X_{\text{solvent}} = \text{O}_{\text{water}}, \text{O}_{\text{EtOH}}, \text{C}_{\text{EtOH}}, \text{N}_{\text{ACN}},$ and C_{ACN}). The $\text{O}_{\text{xylan}}\text{-}X_{\text{solvent}}$ $g(r)$ s (Figure 4.11A-E) describe well-structured solvation layers that follow a similar behaviour as that of $\text{O}_{\text{cell}}\text{-}X_{\text{solvent}}$ $g(r)$ s, presented in Figs 4.6 and 4.8. In the absence of the hydroxymethyl group, water coordinates almost equally O2 and O3, whereas O_{EtOH} and N_{ACN} show a higher preference towards O2. Moreover, the $\text{O}_{\text{xylan}}\text{-}\text{C}_{\text{EtOH/ACN}}$ $g(r)$ s in Figures 4.11D and 4.11E for pure ethanol and acetonitrile, respectively, show that the O2 and O3 are also equally coordinated. The organization of the water-organic solvent mixtures at 25, 50, and 75 wt% is also described with $g(r)$ between $\text{O}_{\text{xylan}}\text{-}X_{\text{solvent}}$, and the results are reported in the Supporting Materials, Figures S4.3 and S4.4. The $\text{O}_{\text{xylan}}\text{-}\text{O}_{\text{water}}$ $g(r)$ s display similar coordination-site trends in comparison with those found for the pure water, varying in the magnitude of the $g(r)$ as a function of solvent content. The $\text{O}_{\text{xylan}}\text{-}\text{O}_{\text{EtOH}}$ $g(r)$ and $\text{O}_{\text{xylan}}\text{-}\text{N}_{\text{ACN}}$ also display similar trends as in the pure organic solvents, but the $\text{O}_{\text{xylan}}\text{-}\text{C}_{\text{EtOH}}$ and the $\text{O}_{\text{xylan}}\text{-}\text{C}_{\text{ACN}}$ display a coordination-site inversion and follows the order $\text{O3} > \text{O2}$ in all water-organic solvent mixtures.

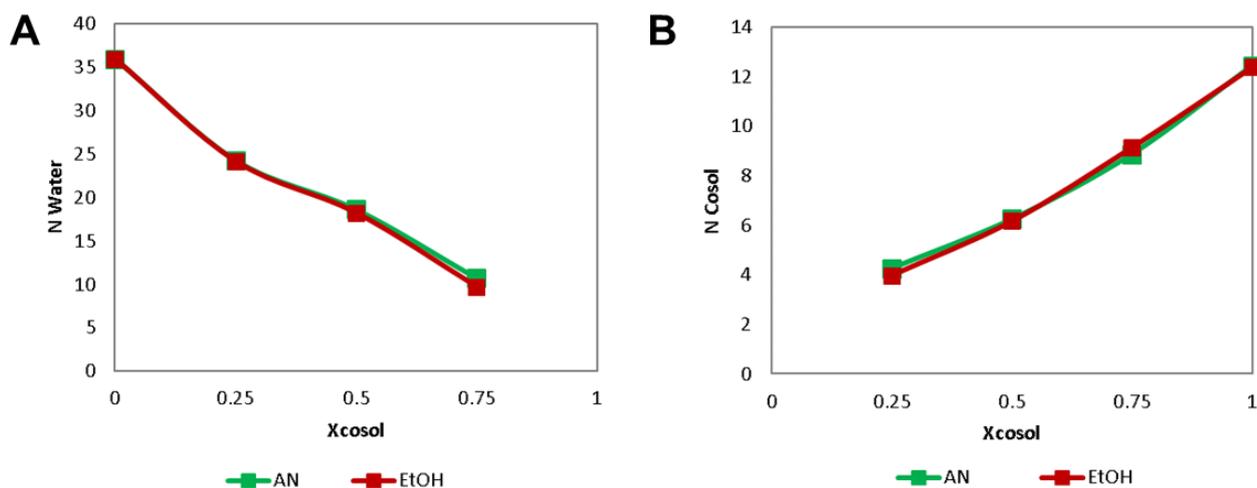


Figure 4.13. Cumulative number of molecules (N) of water (A) and organic cosolvent (B) within 0.7 nm from the xylose glycosidic oxygen, as a function of organic solvent concentration. Green: acetonitrile; red: ethanol.

The computed $cn(r)$ of water and organic solvent molecules within $r = 0.7$ nm from the glycosidic oxygen of xylose, O4, as a function of organic solvent concentration, are summarized in Figure 4.12. The number of water molecules surrounding the xylan surface (Figure 4.12A) decreases with cosolvent concentration, and show similar behaviour in ethanol and acetonitrile mixtures, being slightly more abundant in the presence of acetonitrile. Correspondingly, in Figure 4.12B, the number of organic solvent molecules surrounding the surface increases with the solvent concentration. At low organic solvent fraction, there is a slightly higher number of acetonitrile molecules than ethanol. However, at 75 % ethanol, an inversion of the $cn(r)$ is observed, in which more ethanol molecules are found near the xylose surface than acetonitrile molecules. When compared with the $cn(r)$ obtained for cellulose, we observe that approximately ~ 35 water molecules surround each xylose monomer, whereas this number is significantly reduced for each glucose monomer in cellulose. Due to the assembly of the cellulose chains in the nanocrystals, the solvent access in cellulose is more limited and only about ~ 18 water molecules are found in pure water in the most hydrophilic surface. Correspondently, ~ 12 organic solvent molecules are found surrounding each xylose monomers in comparison with ~ 6 ethanol or acetonitrile molecules surrounding glucose monomers.

3.2.3 H-bonds analysis

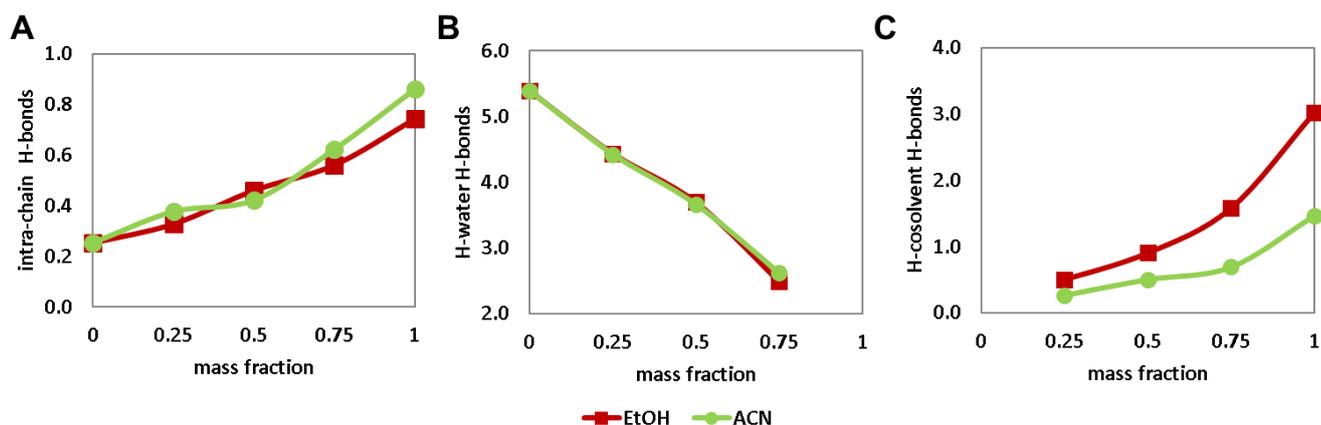


Figure 4.14(A) Average number of xylan intra-molecular H-bonds per xylose monomer at different ethanol and acetonitrile concentrations. Average number of H-bonds per xylose monomer between xylan and water (B) or xylan and organic solvent (C) at different organic solvent concentrations. Green circles: acetonitrile; red squares: ethanol

We monitored the average number of intramolecular hemicellulose (xylan) H-bonds and the intermolecular hemicellulose-water and hemicellulose-organic solvent H-bonds, and the results are summarized in Figure 4.13. As in the case of cellulose, the frequency of hemicellulose intramolecular H-bonds (Figure 4.13A) increases from water to pure ethanol and acetonitrile. Contrary to the cellulose nanocrystal, the intramolecular H-bonds frequency in acetonitrile is higher than in pure ethanol. The number of xylose-water H-bonds, in Figure 4.13B, show a decrease as the organic fraction increases and seem to be equivalent in the presence of ethanol or acetonitrile co-solvents. On the contrary, the difference between the number xylose-cosolvent H-bonds (Figure 4.13C) evidences the adhesiveness of ethanol to hemicellulose, due to its H-bond donor capacity. The aprotic nature of acetonitrile is reflected in the reduced intermolecular H-bonds.

We analyzed also the H-bond lifetimes of the xylan-water and xylan-organic solvent H-bond types from the autocorrelation functions (Table 4.3). As found for cellulose, the lifetime of xylan-water H-bonds increases with the organic fraction and longer-lasting H-bonds are observed in water-ethanol than in water-acetonitrile mixtures. Shorter lifetimes are observed for the xylan-water H-bonds than for the cellulose-water H-bonds, due to richer water interfaces with xylose. This allows water exchange between the bulk and surface water, along with the extra flexibility of the xylose tetramer chain and the access of the hydroxyl groups.

Table 4.5. H-bond lifetimes for the hemicellulose-water (H-W) and hemicellulose-cosolvent (H-Cosol), interaction types, obtained from the H-bond autocorrelation functions of each H-bond type.

	HB lifetimes (ps)	
Solvent mixture	Water-EtOH	Water-ACN

X_{cosol}	H-W	H-Cosol	H-W	H-Cosol
0	5.3	-	5.3	-
0.25	9.6	11.3	6.7	4.2
0.50	14.8	15.9	8.9	5.8
0.75	23.4	20.5	11.8	5.4
1	-	22.9	-	5.5

3.2.4. Solvent effect on diffusivity of xylan

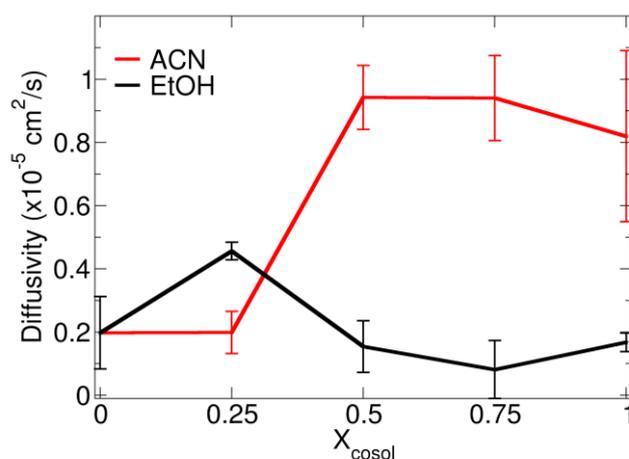


Figure 4.15-Computed diffusivity of xylan tetramers in water-organic solvent mixtures. Green: acetonitrile; red: ethanol

Finally, we described the dynamics of the xylose tetramer in the different solvents by computing the xylan diffusivities, shown in Figure 4.14. These were obtained from the evolution of the mean squared displacement, as described in the computational details. The evolution of the computed diffusivities as a function of the organic solvent fraction shows a non-linear behaviour and opposite trends are observed in ethanol and acetonitrile solvents. The mobility of xylan in water-ethanol mixtures increases at low concentration, i.e. 25 wt%, and then decreases below the water limit upon addition of ethanol. Hemicelluloses have been reported experimentally to turn into a gel with the addition of ethanol [43, 44]. Previous MD analysis have also reported a decrease of hemicellulose diffusion coefficient in the ethanol-water mixture [40]. In water-acetonitrile mixtures, on the other hand, the diffusivity of xylan remains almost unchanged at low acetonitrile content, but it strongly increases beyond 50 wt% thus showing an opposite behaviour than that in the presence of ethanol. In water, hemicelluloses are highly soluble, and no gelation has been experimentally reported [43, 44].

3.3 Dynamics of lignin in water-organic solvents

3.3.1 Solvent effect on lignin conformations

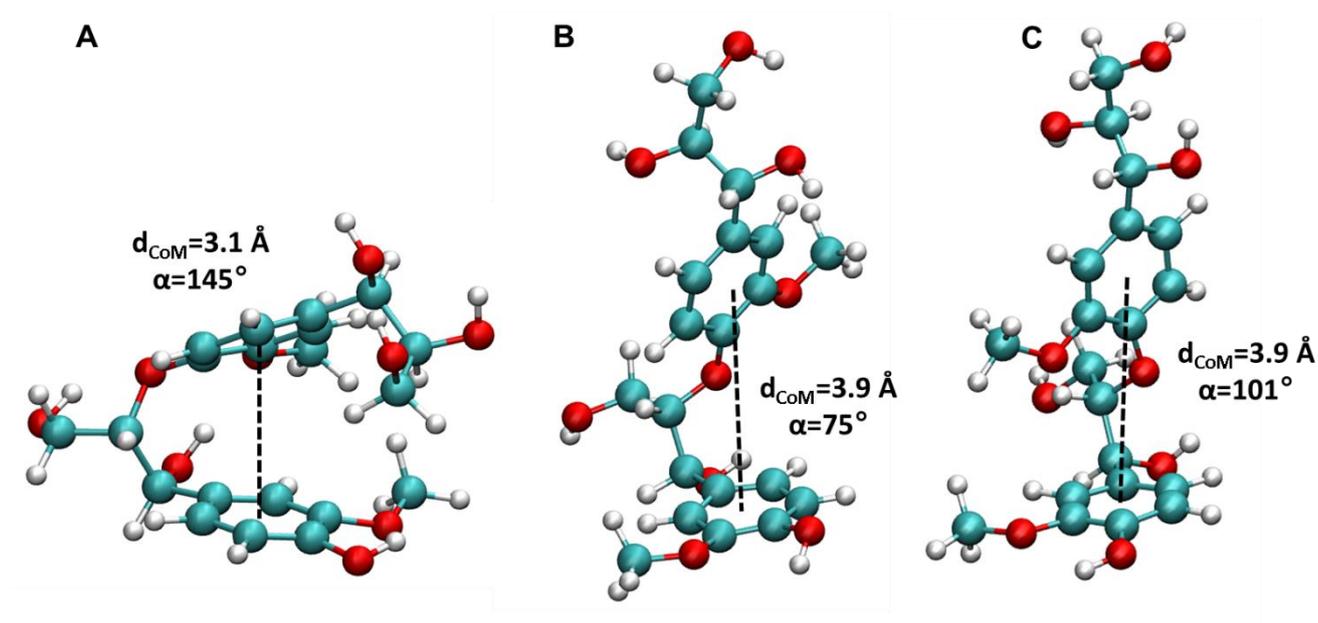


Figure 4.16. Lignin dimer conformers from the cluster analysis in (A) water, (B) 75% ethanol, and (C) 25% acetonitrile from MD simulations (solvent molecules are not shown). The distance between the centres of mass of the two-chromophore rings, d_{CoM} , is shown as an indicator of stacking between the chromophores. The angle between the planes of the two chromophore rings, α , is also measured to differentiate between co-planarity ($\alpha \sim 0^\circ$, $\alpha \sim 180^\circ$) and T-stacking ($\alpha \sim 90^\circ$).

The lignin dimer conformations in the solvents (Figure 4.15) were described considering the mutual displacement of the phenolic rings in lignin. This analysis was carried out by means of distributions of the solvent-accessible surface area (SASA), the ring-ring distance between the phenolic lignin rings (d_{CoM}), the angle between the planes of these rings (α), and the torsion angle of the dimer θ ($\theta = \text{C4}'\text{-O4-C}\beta\text{-C}\alpha$), summarized in Figure 4.16. In water, a bad solvent for lignin, the lignin dimer has the smallest average SASA (Fig 16 A, E) and the interactions with water remain highly hydrophobic as revealed from the most stacked lignin conformation (Figure 4.15A) with the closest ring-ring d_{CoM} of 0.32 nm (see Figure 4.16B, F) and an angle between the ring planes that falls at $\alpha \approx 150^\circ$ (10C, G). The torsion angle of the dimer θ ($\theta = \text{C4}'\text{-O4-C}\beta\text{-C}\alpha$) shows a dihedral conformation distribution near $\theta \approx -90^\circ$ in pure water and diluted ethanol (Figure 4.16C). For the pure ethanol and acetonitrile, and concentrated mixtures, the dihedral conformation is mostly found near $\theta \approx -150^\circ$.

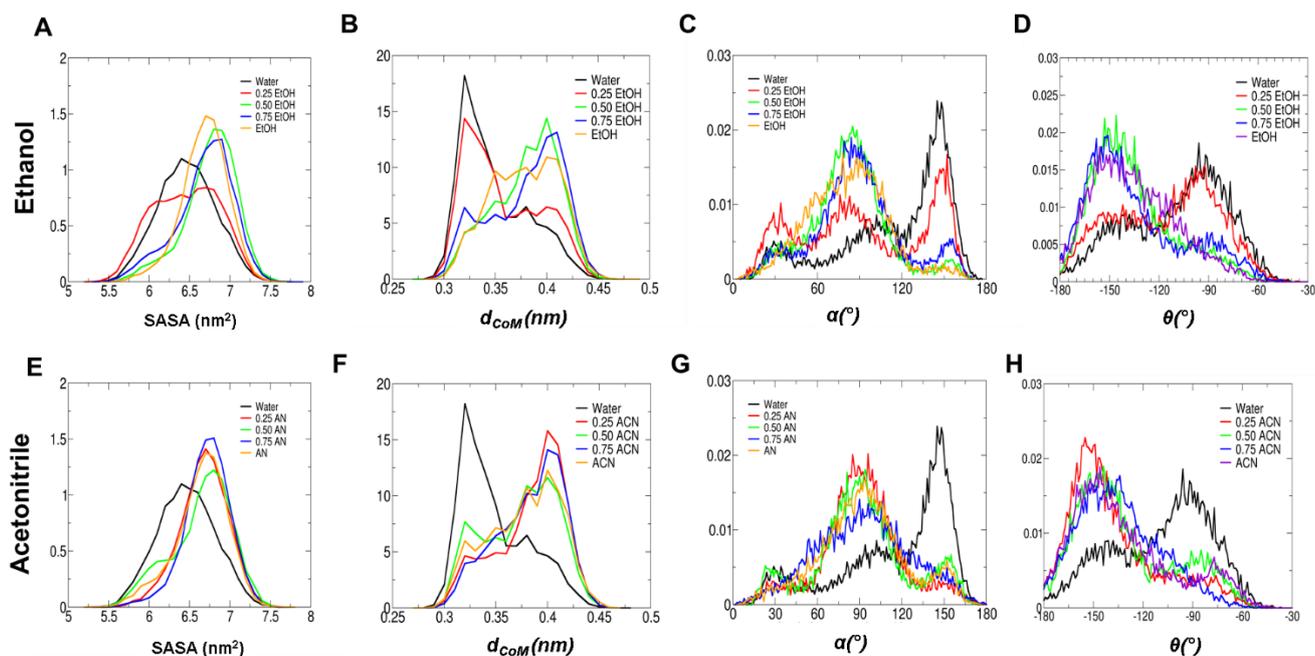


Figure 4.17 Lignin dimer distributions of (A,E) solvent accessible surface area, SASA, (B,F) average centre-of-mass ring-ring distance, d_{CoM} , (C,G) α angle between ring planes of lignin models (D,H) torsion angle θ ($\theta=C4'-O4-C\beta-C\alpha$). (A-D) in water-ethanol and (E-H) in water-acetonitrile, respectively, at different concentrations.

In the mixed solvents, the SASA increases, thus indicating more favourable interactions between the solvents and the lignin dimer. The lignin dimer adopts a less stacked conformation with a ring-ring d_{CoM} of ~ 0.4 nm (Fig. 4.16B, D), and the angle between the phenolic rings $\alpha \approx 90^\circ$, corresponding to a T-shaped stacking. Furthermore, the torsional angle distribution with a maximum of $\theta \approx -150^\circ$ (Fig. 4.16 D, H) is narrower in 75% ethanol and 25% acetonitrile solution than in pure organic solvent and the other solvent mixtures.

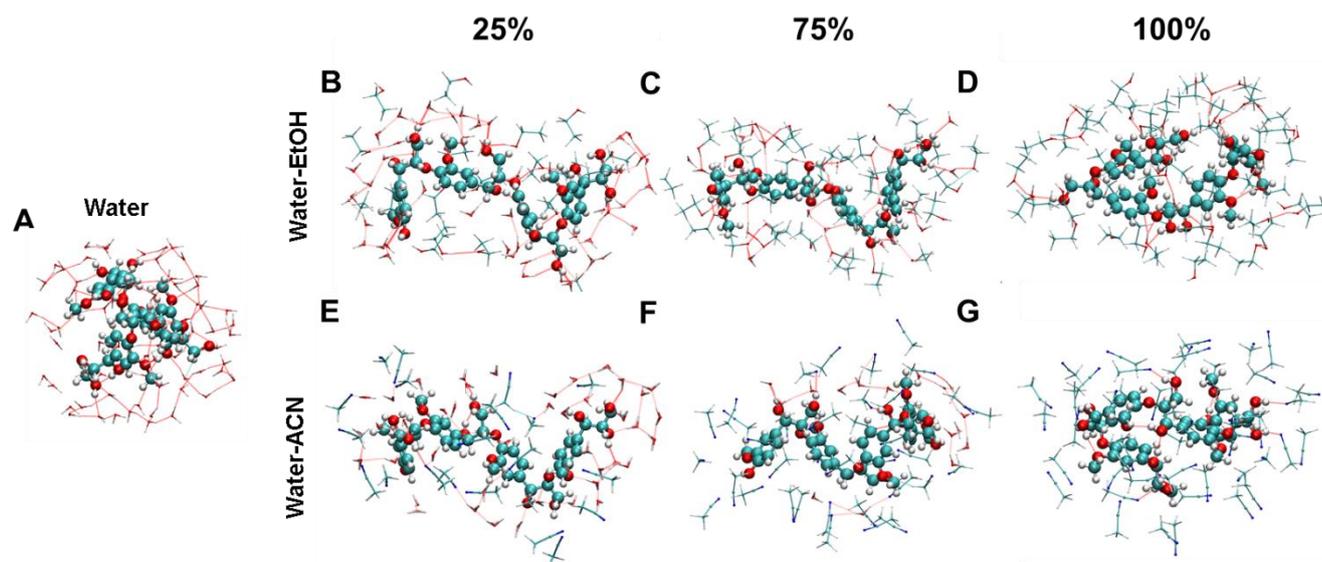


Figure 4.18. MD snapshots of representative lignin tetramer structures obtained from the cluster analysis in pure and mixed solvents, along with a 3.5 Å solvent shell. CPK colour code: oxygen (red), hydrogen (white), carbon (cyan), and nitrogen (blue). Hydrogen bonds are shown in dashed red lines.

It is worth noting that, despite the SASA distributions in Figures 4.16A,C fall in a very narrow range, the surface area of lignin tends to be larger in the mixed solvents than in the pure organic solvents. To follow this observation, we computed the SASA distributions of the lignin tetramer in pure and mixed solvents. The tetramer structures in the considered solvents are illustrated in Fig. 4.17, and the results are summarized in Fig. 4.18. In agreement with our previous observation, the SASA of the lignin tetramer in pure organic solvents falls in a wider distribution between ~ 10 - 12 nm^2 , whereas a narrower distribution that falls near ~ 12 nm^2 is obtained in 75% ethanol and 25% acetonitrile. A somewhat narrow SASA distributions fall at ~ 11 nm^2 in 50% ethanol and 75% acetonitrile, while the SASA distributions in 25% ethanol and 50% acetonitrile behaves as their pure organic counterparts.

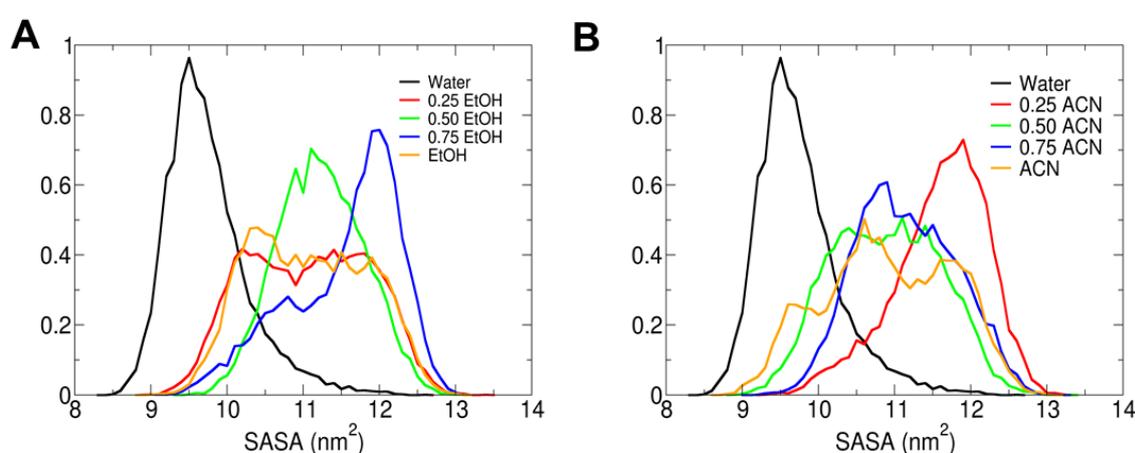


Figure 4.19 – Average distributions of solvent accessible surface area, SASA, of lignin tetramer models in (A) pure ethanol and (B) acetonitrile, and their respective aqueous mixtures.

These results suggest that concentrated mixtures ethanol near 75 w% and 25% acetonitrile solutions are better solvents than the pure organic solvents for lignin molecules with guaiacyl groups and β -O-4 linkages. Another point is to notice that the surface area of the lignin models, here studied, vary in a wide range (particularly when compared with the surface of the cellulose nanocrystal), as both experimental and theoretical studies suggest that β -O-4 structures are flexible molecules that can adopt a large number of conformations [23, 28, 29]. These results, although obtained for limited lignin size models, agree reasonably well with the results of Smith *et al.*[35], showing that lignin polymers containing up to 60 units in water-tetrahydrofuran (THF) mixtures adopt an unfolded conformation with increased SASA, whereas, in pure water, lignin polymers adopt a crumbled globular-like shape, with a considerably reduced SASA. Additionally, very recent studies of lignin solubility in water-ethanol mixtures have identified a solubilisation maximum at 60 wt% ethanol [79], which is expected to result in more extended conformations of lignin polymers, in agreement with our observations.

3.3.2 Organization of solvents around lignin models

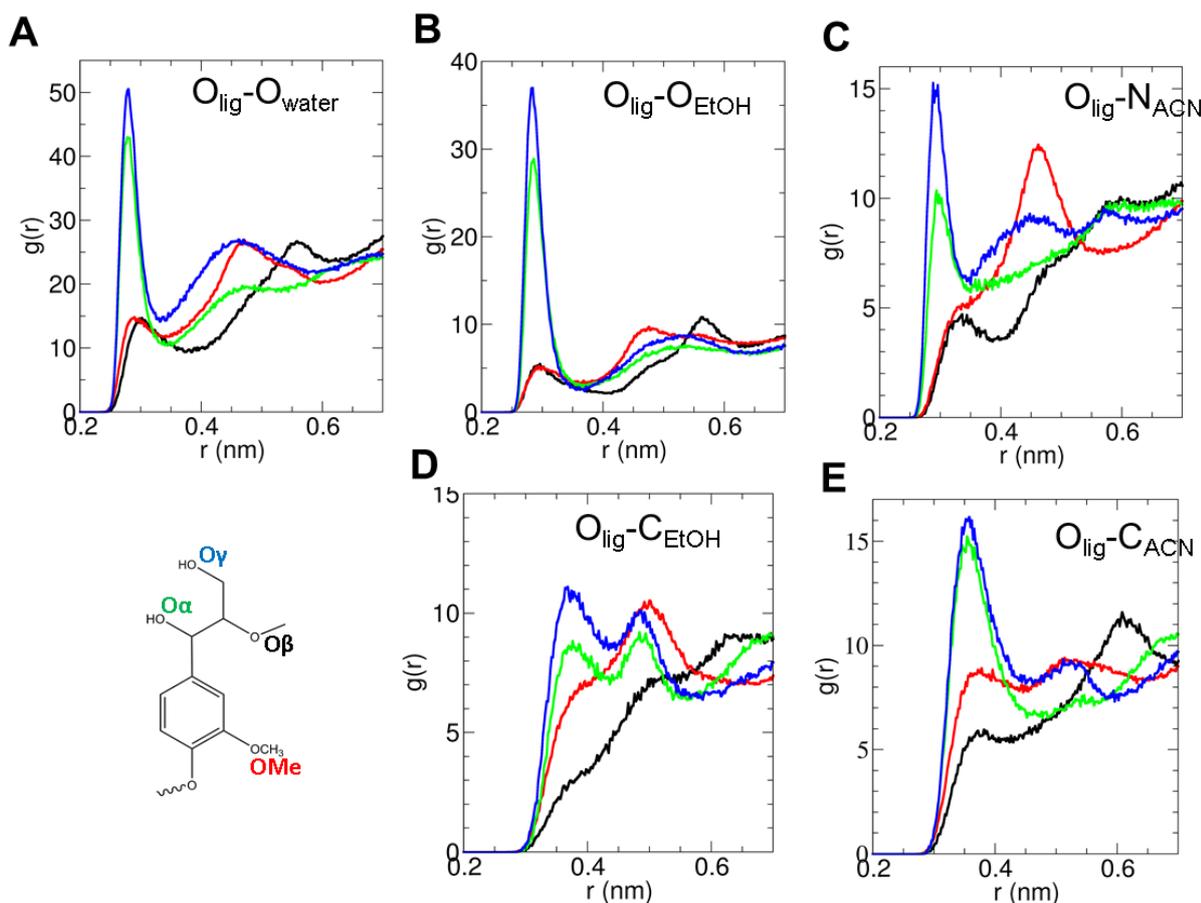


Figure 4.20 - Solvation of lignin in pure solvents. Radial distribution functions, $g(r)$, of the Olig-X_{solv} pairs. The reference site, X, in the solvents are water oxygen (A), EtOH oxygen (B), and ACN nitrogen (C), and $g(r)$ of Olig-C_{Me} with C_{Me}, the carbon of methyl groups, in ethanol (D) and in acetonitrile (E). Olig colour code is OMe (red), O β (black), O α (green), and O γ (blue), see Figure 4.1 for lignin O notation

The organization of solvent molecules around the lignin dimer has been studied from the computed RDF $g(r_{\text{Olig-X}_{\text{solv}}})$ between lignin oxygens (O α , O γ , OMe, and O β) and the solvents. Here X indicates O and C_{Me} atoms in ethanol, O atom in water, and N and C_{Me} atoms in acetonitrile. The $g(r_{\text{Olig-X}_{\text{solv}}})$ results for the pure solvents in Figure 4.19 reveal well-structured solvation layers around the lignin dimer. Higher coordination is found between solvents sites and O α and O γ of the propyl alcohol chains, whereas the O β and OMe are significantly less coordinated. Water and ethanol oxygen atoms display the highest number density when coordinating O α and O γ . This observation is in agreement with the lignin-water H-bond frequencies reported by Charlier and Mazeau [30], who showed that hydroxyl groups (O α and O γ) are 5 times more involved in H-bonds than OMe and 10 times more than the backbone O β . On the other hand, acetonitrile coordination towards O α and O γ is much weaker when compared with water and ethanol solvents. This is evidenced by nearly two times lower peak intensity at ~ 0.30 nm in $g(r_{\text{Olig-NACN}})$ profiles. On the other hand,

$g(r_{\text{Olig-CACN}})$ acetonitrile shows a significant number density of methyl carbons, which interact through the aliphatic hydrogens with the lignin alcohol groups. Thus, acetonitrile displays ahead (N, H-bond acceptor) and tail (methyl) interaction with lignin (as with cellulose), whereas ethanol shows predominantly a head-type (O, H-bond acceptor/donor) interaction.

The organization of the water-organic solvent mixtures at 25, 50, and 75 wt% is also described with $g(r)$ between $\text{O}_{\text{lig-X}_{\text{solv}}}$, and the results are reported in the Supporting Materials, Figure S4.5 and S4.6. The $\text{O}_{\text{lig-O}_{\text{solv}}}$ lignin-water and lignin-ethanol $g(r)$ s display similar coordination-site trends in comparison with those found for the pure solvents, varying in the magnitude of the $g(r)$ as a function of solvent content. The lignin-acetonitrile $g(r)$ in water-acetonitrile mixtures, however, show a significant decrease of the O_{γ} coordination as it becomes rapidly coordinated by water. Thus, the $\text{O}_{\text{lignin-N}_{\text{ACN}}}$ coordination is inverted and follows the order $\text{O}_{\alpha} > \text{O}_{\gamma}$, whereas OMe and O_{β} remain mostly coordinated from the second shell. Moreover, the $\text{O}_{\text{lignin-C}_{\text{Me}}}$ coordination in 25% acetonitrile and ethanol is also inverted and follows the order $\text{OMe} > \text{O}_{\gamma} > \text{O}_{\alpha} > \text{O}_{\beta}$, being this more evident in the acetonitrile mixture.

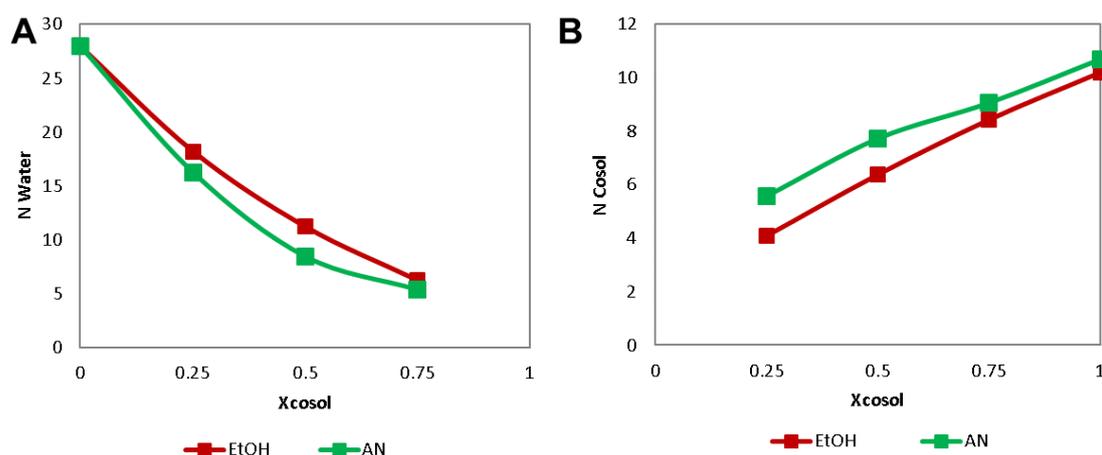


Figure 4.21. Cumulative number of molecules (N) of water (A) and organic solvent (B) within 0.7 nm from the lignin methoxy oxygen, as a function of cosolvent concentration. Green circles: acetonitrile; red squares: ethanol.

Furthermore, we computed the $cn(0.7)$, i.e., the number of water and organic solvent (cosol) molecules within $r = 0.7$ nm from the methoxy oxygen, OMe , as a function of organic solvent concentration, as summarized in Figure 4.20. The number of water molecules surrounding the lignin monomers (Figure 4.20A) decreases with cosolvent concentration, as a result of the water displacement by the cosolvent. Moreover, more water molecules within this solvent shell are present in ethanol than in acetonitrile at all mixed solvent concentrations. In Figure 4.20B, correspondently, the number of organic molecules surrounding the lignin monomer decreases with the organic fraction. Moreover, more acetonitrile molecules are found near the lignin surface than ethanol molecules at all concentrations (Fig. 4.20 C). We note that keeping in mind that the volume of an H_2O molecule is about 1/3 of the volume of an EtOH or ACN molecule, at 100% organic solvent concentration approximately three water molecules should be replaced by one ethanol molecule. Our

analysis in Fig. 4.5, indeed, indicates that nearly 2.7 and 2.6 water molecules are replaced by one EtOH and ACN molecule, respectively.

3.2.3 H-bonds analysis

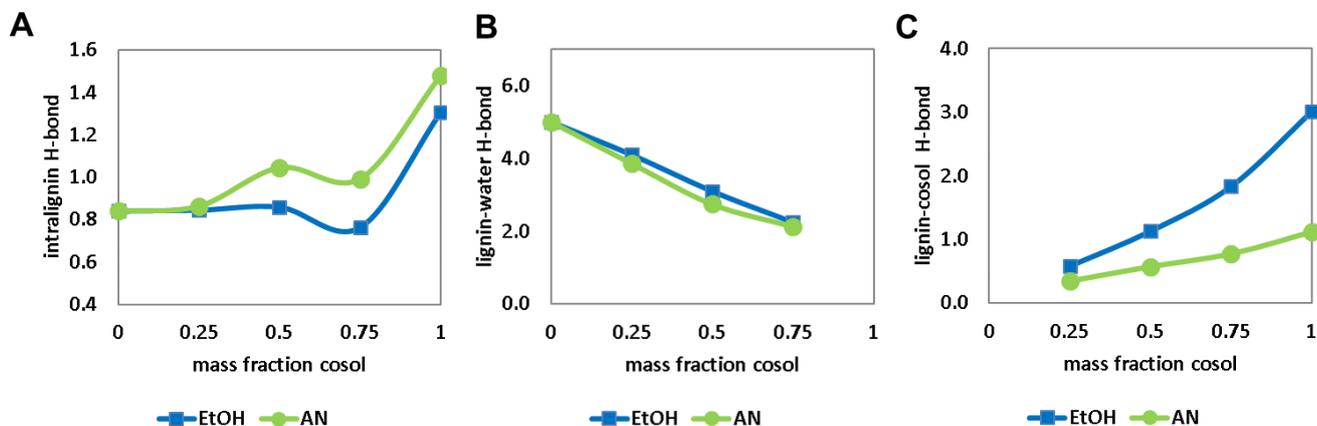


Figure 4.22-(A) Average number of lignin tetramer intra-molecular H-bonds per guaiacyl monomer at different ethanol and acetonitrile concentrations. Average number of H-bonds per guaiacyl monomer between lignin and water (B) or lignin and organic solvent (C) at different organic solvent concentrations. Green circles: acetonitrile; blue squares: ethanol

We monitored the average number of intra- and intermolecular H-bonds between the lignin tetramer and the solvents, and the results are summarized in Figure 4.21. As with cellulose and the xylan tetramer, the frequency of lignin intramolecular H-bonds (Figure 4.21A) increases from water to pure organic solvent, except at 75% ethanol where a minimum is observed. Also, at 25% organic solvent, the intramolecular H-bonds remain practically unchanged compared to water. As with xylan, the frequency of intramolecular H-bonds is lower in pure ethanol than in pure acetonitrile. Furthermore, this trend is maintained throughout all the mixed solvent range. The number of lignin-water H-bonds, in Figure 4.18B, show a decrease as the organic fraction increases and a higher frequency is observed in the presence of ethanol than acetonitrile. In agreement with the cellulose-cosolvent and xylan-cosolvent H-bonds, the frequency of lignin-cosolvent increases with the organic fraction and significantly less H-bonds are formed between lignin and acetonitrile at all concentrations, due to its aprotic nature. Nonetheless, the cumulative numbers in Figure 4.20 show that more acetonitrile molecules are surrounding lignin than ethanol molecules. Therefore, this suggests the lignin-acetonitrile interactions are mainly driven by van der Waals rather than electrostatic interactions and play a significant role in stabilizing lignin.

Table 4.6. H-bond lifetimes for the lignin-water (L-W) and lignin-cosolvent (L-Cosol) interaction types, obtained from the H-bond autocorrelation functions of each H-bond type.

	HB lifetimes (ps)	
Solvent	Water-EtOH	Water-ACN

mixture				
X_{cosol}	L-W	L-Cosol	L-W	L-Cosol
0	8.1	-	8.1	-
0.25	17.7	22.7	12.5	9.9
0.50	24.9	30.1	17.8	11.3
0.75	48.1	43.3	25.7	10.4
1	-	53.1	-	8.0

We analyzed the H-bond lifetimes of the lignin-water and lignin-organic solvent H-bond types from the autocorrelation functions, and the results are summarized in Table 4.4. In agreement with the results of cellulose (Table 4.2) and xylan (Table 4.3) in the previous sections, the lifetime of lignin-water H-bonds increase with the organic fraction. As we showed in the previous chapter, the water H-bond lifetimes are shorter when richer water interfaces exist, since this allows faster water exchange between the bulk and surface water. Correspondently, shorter polymer-water lifetimes are observed for the xylan-water (5.3 ps) than the lignin-water (8.1 ps) H-bonds, followed by the cellulose-water H-bonds (9.8 and 12.3 ps for chains AF and BDEG, respectively). Water-rich interfaces with xylose ($cn(0.7 \text{ nm}) = \sim 35$) allow water exchange between the bulk and surface water faster than lignin, whose number of water molecules at 0.7 nm is ~ 28 . More water molecules surrounding xylose and lignin are observed since the monomers of the polymers are considerably exposed, in comparison with the glucose monomer in cellulose that remain less solvent-exposed in the nanocrystal assembly. Therefore, cellulose-water H-bonds last longer, even at the water-rich surface (chain AF), whose number of water molecules at 0.7 nm is ~ 18 . Furthermore, we also observed that the polymer-water H-bond lifetimes in water-ethanol mixtures are longer lasting than in water-acetonitrile mixtures. As we showed in the previous chapter, this is explained by the higher frequency of H-bond formation between water and ethanol than with acetonitrile and water. Thus, we believe that the polymer-water H-bond lifetimes are longer-lasting in ethanol mixtures than in acetonitrile mixtures, despite having richer water interfaces in ethanol mixtures, because the constant interaction between water and ethanol that constrains the water movement. Finally, the polymer-ethanol H-bond lifetimes are always longer-lasting than the polymer-acetonitrile ones due to their higher frequency, given ethanol's H-bond donor and acceptor capacity, whereas acetonitrile remains only as an H-bond acceptor.

3.4 Polarization effects on lignocellulose polymers, evaluated at the QM (DFT) level

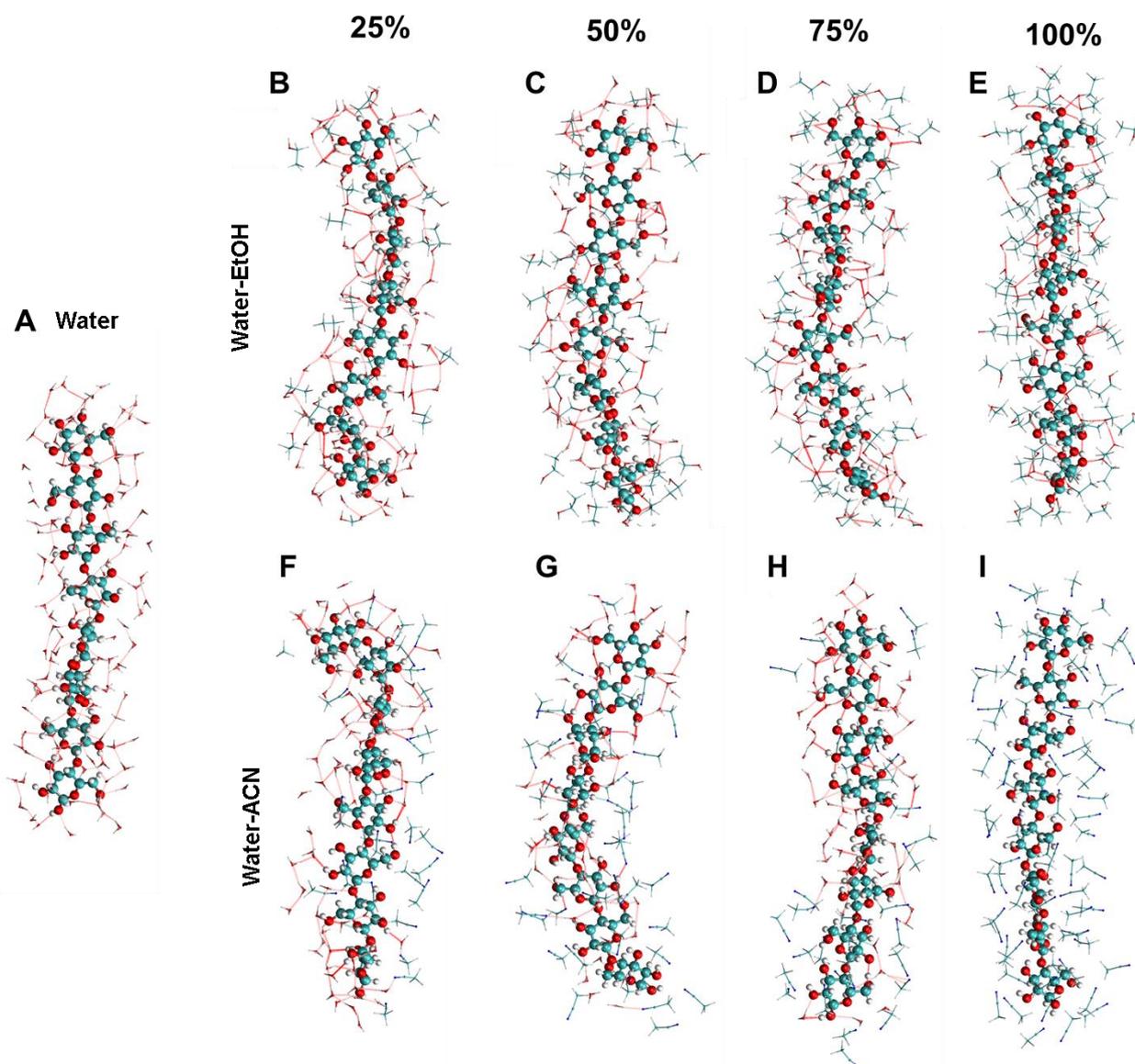


Figure 4.23. MD snapshots of single cellulose chain structures obtained from the cluster analysis in pure and mixed solvents, along with a 3.5 Å solvent shell. CPK colour code: oxygen (red), hydrogen (white), carbon (cyan), and nitrogen (blue). Hydrogen bonds are shown in dashed red lines.

We studied at the QM level the solvent effects on the structures of cellulose chains (Figure 4.22) and the lignin tetramers (Figure 4.17) obtained from MD simulations. The scope was to obtain an estimation of the trends of charge polarizations of the selected oxygen atoms (*vide supra*) in cellulose and lignin models in the presence of the considered pure and binary aqueous organic solvents. Whereas from MD simulations we could gain knowledge about the statistical accessibility and the dynamic evolution of H-binding of solvent molecules to specific oxygens in the truncated lignocellulosic polymers models, we completely neglect the mutual charge polarization using non-polarizable force fields (FF) methods. Note that the use of polarisable

FF approaches is not practical for the present studies, because of the significant computational demand. To provide a better insight into the polarization effect we therefore evaluated the atomic charges using DFT methods (see the Computational details section in this chapter) on the energy minimized structures, taken from the cluster analysis of the MD simulations, and analyzed these charges in relation to the MD results, presented in the previous sections. The atomic charges were computed using the Mulliken population scheme, which provides reliable results for not extended basis sets, as are the double- ζ types bases, used by us. We have considered the cellulose single chain dynamics in various solvents as a reasonable approximation because all the OH sites are accessible to the solvent molecules.

The oxygen partial atomic charges (q_O) obtained for the cellulose chain are illustrated in Figure 4.23, and the selected average q_O and q_H (for hydrogen atoms in OH groups) values are reported in Table 4.5, without considering the reducing and non-reducing ends of the cellulose chain in the absence of the solvent. The O partial charges in Figure 4.23 are displayed by following the number index in the topology file in the periodic order O5-O2-O3-O4-O6. Well-formed repetitive patterns are found along the cellulose chain in gas face. O2 shows the lowest charge value, with $q_{O2} = -0.542 e^-$. In contrast, the less negative value is found for O4, with a charge of $-0.380 e^-$. Moreover, the H-bond donor/acceptor character of each hydroxyl oxygen can be deduced from the partial charges. Among the oxygen sites of interest, the O2-H is the most polarized, because O2 is the most negative O-site ($q_O = -0.542 e^-$) and H is slightly more positive than the other hydrogen sites of interest ($q_H(\text{H}_{O2}) = 0.288$), as follows from the QM results in Table 4.5. Therefore, the O2-H group is more prone to lose its H, which reveals O2 as a better H-bond donor than O3 and O6 sites. This correlates well with the observed behaviour reported for the $O_{\text{cell}}-N_{\text{ACN}}$ RDF (Figure 4.6), which shows the highest coordination of acetonitrile N for O2 in cellulose. On the contrary, cellulose O6 is the less negative alcohol oxygen, and its H has a slightly less positive charge than the other hydrogen alcohols. Thus, it would perform better as H acceptor, which correlates well with the observed preferential interactions of water and ethanol oxygens for this site.

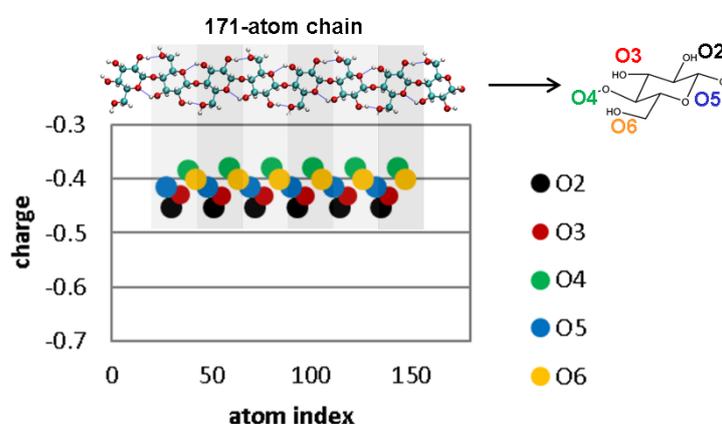


Figure 4.24. Mulliken charges in e^- of a cellulose chain with 8 glucose monomers. Charges of the reducing and not-reducing ends are not shown.

We also examined the charges of the cellulose oxygen in the presence of explicit solvents. First, we compared the average partial charges obtained in water with the charges used in the force field during the MD simulations (Table 4.5 and Figure 4.24A). The partial charges used for the alcohol H ($0.42 e^-$) and O ($-0.65 e^-$) are the same for O2, O3, and O6 in the MM force-field, whereas different polarization effects can be observed for each of these atoms at QM level. Moreover, O partial charges are more negative in the MM force field than the ones reported at the QM level. The water solvent effect is more prominent for O3 and O6, which become rapidly coordinated and more negative, and participate in hydrogen bonding with water. The glycosidic oxygen O4 is significantly affected by the polarization effects, despite being poorly coordinated by water according to our MD $O_{\text{cell}}-O_{\text{water}}$ $g(r)$ analysis, because of its relatively lower accessibility to the solvent. The result that despite its lower coordination O4 polarization changes with the solvents, suggests that the solvent to cellulose OH group coordination causes a non-negligible charge redistribution within the glucose monomer(s).

Table 4.7. Average partial Mulliken charges in e^- of O_n ($n=2, 3, 4, 5$ and 6) and H_{Om} ($m=2,3,$ and 6) atoms in a cellulose chain with 8 glucose monomers, in gas phase and in water. Mulliken charges of the reducing and not-reducing ends are not included in the averages. The standard deviations of the average QM charges are reported in parentheses. MM [58, 59]charges used during simulation have been included for comparison.

O_{cell}	QM		MM [58, 59]
	Gas	Water	
O2	-0.542	-0.548(0.017)	-0.65
O3	-0.430	-0.536(0.030)	-0.65
O4	-0.380	-0.432(0.033)	-0.36
O5	-0.414	-0.438(0.045)	-0.40
O6	-0.399	-0.513(0.015)	-0.65
H _{O2}	0.288	0.308(0.021)	0.42
H _{O3}	0.281	0.346(0.021)	0.42
H _{O6}	0.268	0.305(0.018)	0.42

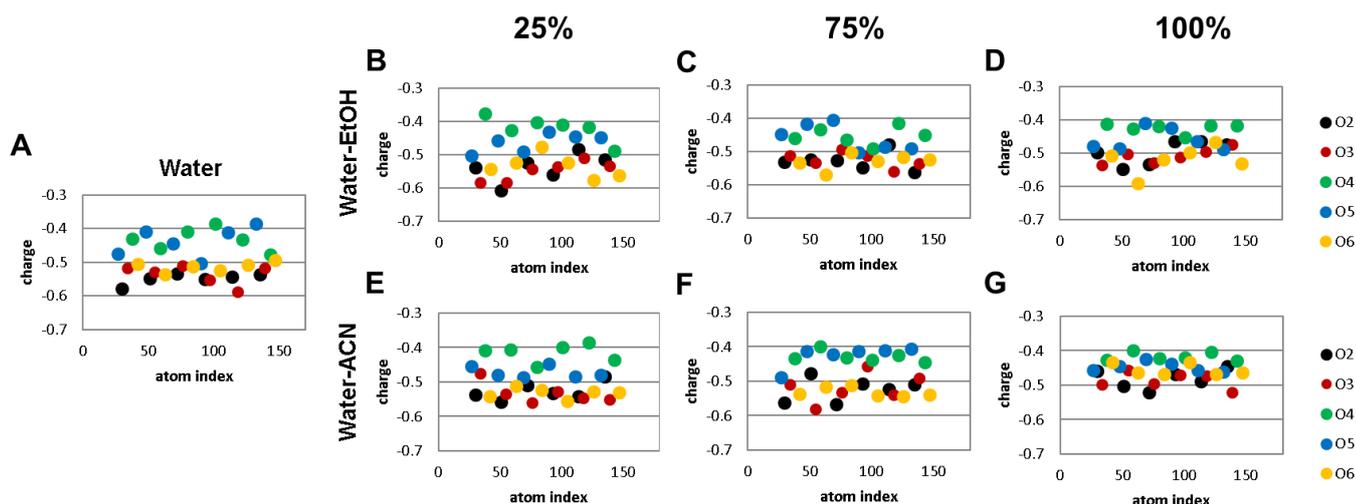


Figure 4.25. Mulliken charges in e- of a 8DP cellulose chain in pure and mixed solvents. Charges of the reducing and not-reducing ends are not shown.

We further compared the polarization effects in water with those in pure and mixed water-organic solvents. The polarization effects in diluted water-cosolvent mixtures (Figure 4.24.B, E) show a similar behaviour to those observed in water, with a lower effect on O5 and this effect persists with the addition of the organic phase. In 75% acetonitrile (Figure 4.24 D) the polarization effect of O5 and O4 is significantly decreased, whereas O2, O3, and O6 show significant interactions with the solvent. Moreover, the polarization effects are significantly reduced in the presence of the pure organic solvents, particularly in acetonitrile. As shown for the RDFs previously, the O5 coordination by the methyl group of acetonitrile seems to provide additional order of the cellulose chain, since the number of O3-H...O5 hydrogen bonds is better preserved for acetonitrile mixtures than in the other solvents. Although our static charge analysis is made on one minimum energy structure, obtained from the cluster analysis of the MD simulations, it is in line with the MD statistical analysis for the specific coordination of solvent molecules to distinct oxygen sites of glucose monomers. This underlines the microscopic nature of the preferential cellulose-solvent interactions.

We now perform the same partial-charge analysis on the lignin tetramer structures, again extracted from the MD cluster analysis, in mixed solvents (Figure 4.17). We observe that the stronger polarization occurs in the presence of water. The O α and O γ atoms are more affected than in ethanol and acetonitrile. A strong polarization of O β and OMe is observed in water and ethanol, whereas this is significantly lower in acetonitrile. Thus, our results show that different polarization patterns result due to the interactions between the solutes and the solvents. This has a direct effect on the magnitude of non-covalent interactions and highlights the need for more accurate than non-polarizable force-field methods when understanding molecular-scale phenomena.

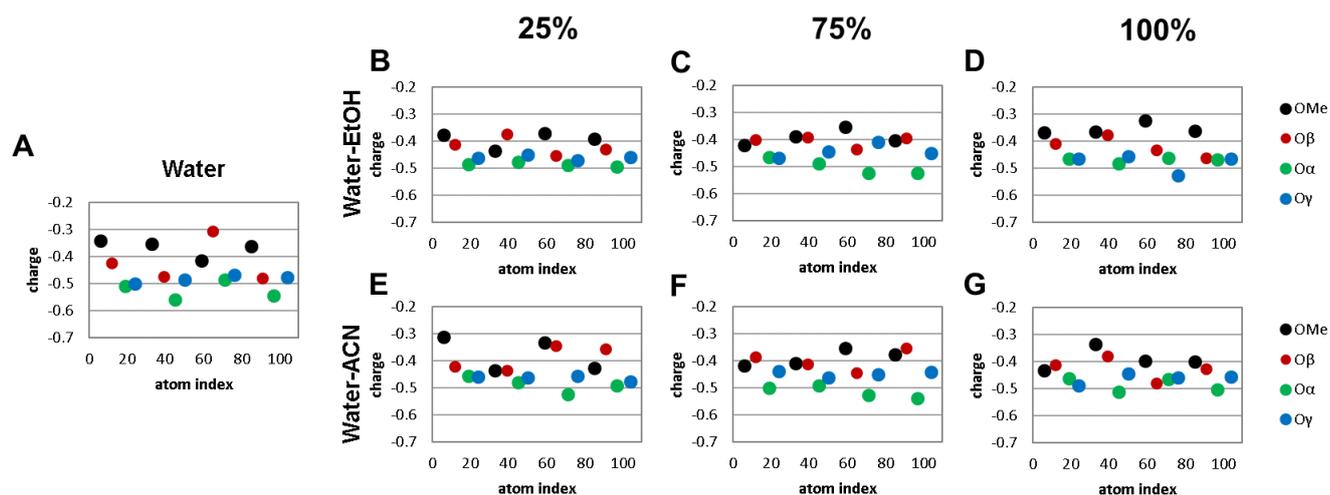


Figure 4.26. Mulliken charges in e^- of OMe, O_α , O_β , and O_γ oxygens of the lignin tetramer in pure and mixed solvents.

Table 4.8. Average partial Mulliken charges in e^- of O_n ($n = \text{Me}, \alpha, \beta, \text{ and } \gamma$) and H_{Om} ($m = \text{Me}, \alpha, \text{ and } \gamma$) atoms in lignin tetramers in water. The standard deviations of the average QM charges are reported in parentheses. MM [60] charges used during simulation have been included for comparison.

Atom	QMwater	MM[60]
OMe	-0.368(0.033)	-0.28
O_α	-0.526(0.034)	-0.66 ^a
O_β	-0.420(0.098)	-0.28
O_γ	-0.483(0.013)	-0.66 ^a
H_{OMe}	0.116(0.022)	0.09
$H_{O\alpha}$	0.310(0.023)	0.43
$H_{O\gamma}$	0.309(0.023)	0.43

^aThe author in ref. [60] kindly provided us with the lignin force field. In all β -O-4 linked monomers, a charge of -0.66 is used for O_α and O_γ atoms; however, a partial charge of -0.54 is in our opinion wrongly reported in ref. [60]

Nonetheless, and despite the differences found between the charges obtained by our DFT calculations and the ones reported by the MM force fields, our statistical MD results reproduce correctly the cellulose-water RDFs profiles. Moreover, they allowed us to gain insight on the preferential interactions of the organic solvents with specific sites of the lignocellulosic truncated model polymers. Furthermore, the correlation trend between the static atomic charges at QM level, with the established from the MD solvent preferential interactions to cellulose and lignin O-sites evidences better the local, microscopic nature of those H-bond interactions types.

3.5 MD and BOMD/MD evaluation of the stability of lignin dimer structures

In this last section of the chapter devoted to the atomistic study of the physicochemical solvent-lignocelluloses interactions, we present the interaction energies between the considered solvents and the lignocellulosic

truncated polymer models. The lignin dimer energies in the pure organic and mixed aqueous solvents, relative to the water solvent, are also evaluated from the BOMD/MD simulations in order to gain better insight on the effect of ethanol and acetonitrile solvent components on the energy stability of lignin dimer.

First, we analyzed the interaction energies between the individual lignocellulosic components and the solvents at the MM level. The interaction energies are computed from the sum of the non-covalent interactions (LJ and Coulomb) of the individual polymer-water and polymer-cosolvent energies, and the complete results are summarized in Figures S4.7 to S4.9. Overall, we found that the polymer-water interactions are always electrostatically driven (Fig. 4.25A, C). In pure ethanol, all the interactions with the lignocellulosic components display a stronger Coulomb character. In the presence of water, however, ethanol interactions have stronger LJ contributions with lignin, whereas this contribution is almost equivalent to hemicellulose and cellulose. On the other hand, acetonitrile interactions display always stronger LJ energies. The total polymer-solvent interactions energies summarized in Figure 4.25 show that cellulose and hemicellulose interact strongest with pure water, and this interaction gradually decreases with the addition of the organic phase. Moreover, the interactions with ethanol are stronger than with acetonitrile. On the contrary, the lignin-solvent interactions are stronger than the lignin-pure water interactions in solvent mixtures of ethanol and water, and in 25% acetonitrile-water. Notably, the polymer-solvent interactions inversely correlate with the intramolecular H-bond frequencies in Figures 4.5, 4.13, and 4.21, indicating that non-covalent intramolecular interactions are weaker in pure water than in the organic solvents.

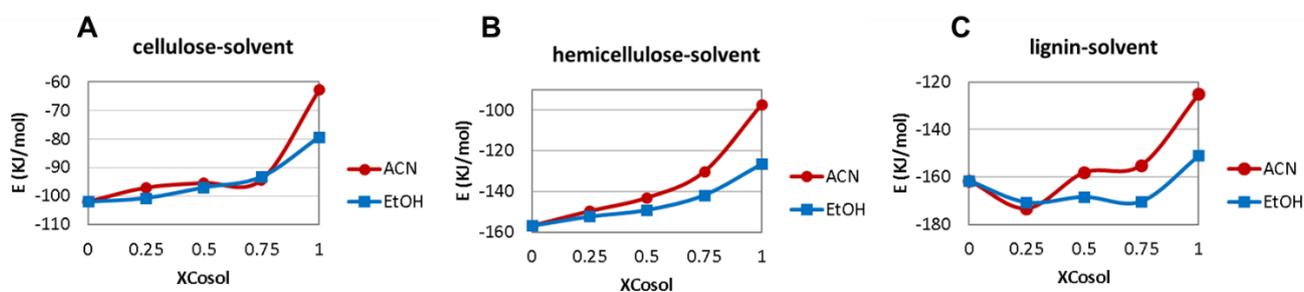


Figure 4.27. Solute-solvent (water+cosolvent) interaction energies for (A) cellulose, (B) xylan, and (C) lignin tetramer, normalized by the number of monomers in each model.

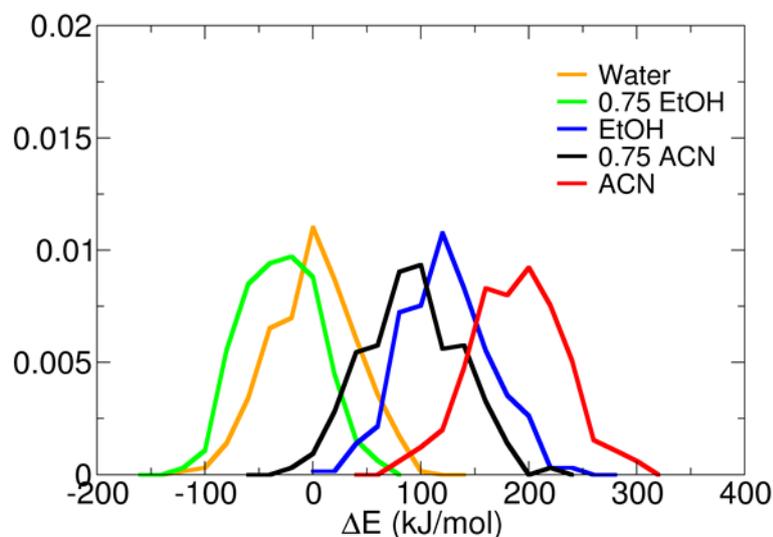


Figure 4.28. DFT relative energies, ΔE , of lignin dimers per monomer computed in the MM explicit solvents. $\Delta E = (\text{Elig-solvent} - \text{Elig-water}) / N_{\text{monomers}}$. Reference energy of lignin in water is $\text{Elig-water} = 1454.616162 \text{ Ha}$. $1 \text{ Ha} = 2625.5 \text{ kJ/mol}$.

To gain better understanding of the stability and non-linear behaviour of the lignin structures in the solvent mixtures we performed QM/MM (BOMD/MD) dynamics for 12 ps, treating lignin at DFT level and the solvents at MM level. Note that the comparison between the full DFT/MM energy was not possible because of the different number of MM solvent molecules due to the different solvent compositions. The DFT energies of the lignin in the QM/MM systems were considered for the last 330 frames, corresponding to the last 1 ps dynamics of the total 12 ps BOMD/MD trajectory length. As reference energy, we chose the lignin DFT energy in MM water (see Computational Details section). The DFT lignin energies in the pure organic and water-organic solvents (solvent molecules treated at MM level), relative to the DFT lignin energy in water were subsequently normalized by the number of lignin monomers. The equation, used by us, to compute these relative DFT energies is reported in the Figure caption of Fig. 4.27. The relative DFT energies in Figure 4.27 show that the stability of the lignin dimers in the MM solvents follows the order $0.75 \text{ EtOH} > \text{Water} > \text{EtOH} > 0.75 \text{ ACN} > \text{ACN}$. The lignin structure in 75% EtOH is stabilized by $\sim 30 \text{ kJ/mol}$ per monomer in comparison to the structure of lignin in water. On the contrary, the lignin structure of lignin in 75% acetonitrile, EtOH and ACN is destabilized by $\sim 90, 120$, and 190 kJ/mol per monomer, relative to water. It is well known that water is a bad solvent for lignin; therefore, the trend of our observations can be contradictory at first glance. Moreover, the lignin intra-molecular hydrogen bonding profiles in Figure 4.21 show a frequency of H-bonds in the order $\text{ACN} > \text{EtOH} > 0.75 \text{ ACN} > \text{Water} > 0.75 \text{ EtOH}$, indicating that the non-covalent intramolecular interactions are weaker in 0.75 EtOH than in the pure ACN. However, we observe that the lignin-solvent interaction profiles, computed from the MD simulations, in Figure 4.26C reveal that the stronger lignin-solvent interaction follows the same trends obtained from the BOMD/MD simulations. Thus,

the strongest lignin-solvent interaction is observed in 75% ethanol, whereas the weakest lignin-solvent interactions are observed in pure organic solvents, particularly in acetonitrile. Therefore, our results indicate that solvent interactions can stabilize the structure, and the energetic gain of forming H-bonds with the solvent compensates the loss of intramolecular H-bonds. In the case of 75% ethanol (and all water-ethanol and 25% acetonitrile) the lignin-solvent interactions are maximized due to favourable interactions that result in a more stable and more extended structure. The lignin structure in water, however, gains stability by reducing the surface-accessible area and reducing the hydrophobic interactions with water.

4. Conclusions

Classical molecular dynamics simulations (MD), quantum chemistry calculations (QM-DFT), and combined QM/MM (BOMD/MD) approaches were employed to describe at the atomistic level the effect of the solvents on the structure and dynamics of lignocellulose compounds - cellulose, hemicellulose, and lignin, truncated polymer models. The considered solvents are first pure water, ethanol and acetonitrile, and second, the binary mixtures of water with the two organic solvents.

At the classic level, the structure and solvation of the cellulose nanocrystal, lignin dimer and tetramer, and xylose were investigated through solvent accessible surface area (SASA), conformational analysis, H-bonding and H-bond lifetimes, and radial distribution functions. The conclusions drawn from the analysis of these properties are the following:

1. The SASA distributions analysis shows the largest cellulose nanocrystal solvent-accessible surfaces in the presence of mixed solvents, whereas it has a minimum in the presence of pure acetonitrile, followed by pure ethanol and water. Likewise, lignin displays the highest surface area in the presence of mixed solvents, while water remains a bad solvent for lignin. On the contrary, the hygroscopic nature of xylose results in a trend to find the lowest surface area in pure acetonitrile and ethanol, whereas a tendency of a more extended surface is observed in pure water and the diluted mixtures.

2. The analysis of the hydroxymethyl group conformations in cellulose shows a higher frequency of tg conformers in pure organic solvents indicating lower capacity of hydroxymethyl to form intramolecular H-bonds; following a transition from tg to more solvent-exposed gg conformation in the mixed solvents. In the pure water solvent, the solvent-exposed gg-hydroxymethyl conformer is predominant. The torsion angle analysis of the xylose glycosidic bond has shown that xylose adopts a left-handed threefold conformation in water and diluted mixtures, whereas a twofold conformation is favoured in pure and concentrated organic solvents. The analysis of the β -O-4 torsion angle in lignin shows a dihedral conformation distribution near $\theta \sim -90^\circ$ in pure water and diluted ethanol which corresponds to a stacked conformation. In contrast, a more extended T-stacked conformation with a distribution that falls near $\theta \sim -150^\circ$ is found in pure and concentrated organic solvents.

3. The polymer-solvent correlation analysis using the radial distribution functions show that lignin-water and lignin-ethanol display similar coordination-site trends in pure and mixed solvents, favouring the electrostatic interactions and coordinating preferentially the $O\gamma > O\alpha$ sites. In acetonitrile-water mixtures, however, the $O\gamma$ site becomes rapidly coordinated by water; therefore, the lignin-acetonitrile coordination is inverted and follows the order $O\alpha > O\gamma$, whereas OMe becomes coordinated at diluted acetonitrile content by the methyl group of acetonitrile. Likewise, the xylan coordination analysis shows that water and ethanol interact mostly electrostatically with the $O2$ and $O3$ sites, whereas apolar interactions between the methyl group of acetonitrile and $O2$ are preferred. Moreover, the $O_{xylose}-C_{EtOH}$ and the $O_{xylose}-C_{ACN}$ display a coordination-site inversion and follow the order $O3 > O2$ in all water-organic solvent mixtures. On the other hand, the cellulose-water coordination follows the order $O6 > O2$ and displays higher number densities in the hydrophilic than the hydrophobic surfaces. Similarly, ethanol follows the order $O6 > O2$ whereas acetonitrile shows higher coordination for $O2$ than $O6$. Furthermore, while the water number density is expected to decrease with the addition of organic solvent, the water exclusion in acetonitrile mixtures from the first cellulose solvation shell, especially in the hydrophobic surface, is stronger than for ethanol mixtures. An increase of the co-solvent phase could yield a more hydrated and more extended water shell around cellulose hydrophilic surfaces, which, enhanced by local microheterogeneity of water-acetonitrile mixtures[78], results in phase-separation when in contact with the different cellulose surfaces. Ethanol mixtures, on the contrary, remain less separated as ethanol-water mixtures are more homogeneous and also share a similar affinity for a specific oxygen site in cellulose.

Quantum chemistry calculations were used to study the polarization effect of solvents in the partial charges of lignin tetramers and cellulose chains. The Mulliken charges of cellulose oxygens in the presence of water are considerably affected by the solvent, whereas the effect is significantly reduced in the presence of acetonitrile. The partial charge of the glycosidic oxygen is affected despite the significant lower coordination reported in the RDFs. Similarly, the partial charges of OMe, $O\beta$, $O\gamma$, and $O\alpha$ oxygens in lignin tetramers show that water strongly interacts with $O\gamma$, and $O\alpha$ oxygens, while the others remain less disrupted in the presence of acetonitrile than water or ethanol.

DFT energies of lignin dimers obtained from BOMD/MD dynamics showed that low energy conformations with a low number of intramolecular H-bonds can exist thanks to a stabilization effect of the solvent. The results suggest that energy gain due to H-bond interactions with the solvent greatly compensates the loss of intramolecular H-bonding, allowing the existence of low energy conformers in the presence of water. Moreover, the lignin-solvent interactions pass through an optimization when adding ethanol to water solvent, as found for the most energetically stable and more solvent-interacting lignin dimer in 75wt% ethanol-water mixtures. Despite the relatively small lignin models, our results are in a qualitative agreement with recent studies that show an increase of lignin solubility in water-ethanol mixtures[79].

References

1. McKendry, P., *Energy production from biomass (part 1): overview of biomass*. Bioresource Technology, 2002. **83**(1): p. 37-46.
2. Ragauskas, A.J., et al., *The Path Forward for Biofuels and Biomaterials*. Science, 2006. **311**(5760): p. 484-489.
3. Parham, R.A. and R.L. Gray, *Formation and Structure of Wood*, in *The Chemistry of Solid Wood*. 1984, American Chemical Society. p. 3-56.
4. Polavarapu, P.L. and C.S. Ewig, *Ab Initio computed molecular structures and energies of the conformers of glucose*. Journal of Computational Chemistry, 1992. **13**(10): p. 1255-1261.
5. Kroon-Batenburg, L.M.J. and J. Kroon, *Solvent effect on the conformation of the hydroxymethyl group established by molecular dynamics simulations of methyl- β -D-glucoside in water*. Biopolymers, 1990. **29**(8-9): p. 1243-1248.
6. van Eijck, B.P., R.W.W. Hooft, and J. Kroon, *Molecular dynamics study of conformational and anomeric equilibria in aqueous D-glucose*. The Journal of Physical Chemistry, 1993. **97**(46): p. 12093-12099.
7. Heiner, A.P. and O. Teleman, *Interface between Monoclinic Crystalline Cellulose and Water: Breakdown of the Odd/Even Duplicity*. Langmuir, 1997. **13**(3): p. 511-518.
8. Heiner, A.P., L. Kuutti, and O. Teleman, *Comparison of the interface between water and four surfaces of native crystalline cellulose by molecular dynamics simulations*. Carbohydrate Research, 1998. **306**(1-2): p. 205-220.
9. Matthews, J.F., M.E. Himmel, and J.W. Brady, *Simulations of the Structure of Cellulose*, in *Computational Modeling in Lignocellulosic Biofuel Production*. 2010, American Chemical Society. p. 17-53.
10. Matthews, J.F., et al., *Comparison of Cellulose I β Simulations with Three Carbohydrate Force Fields*. Journal of Chemical Theory and Computation, 2012. **8**(2): p. 735-748.
11. Matthews, J.F., et al., *High-Temperature Behavior of Cellulose I*. The Journal of Physical Chemistry B, 2011. **115**(10): p. 2155-2166.
12. Hadden, J.A., A.D. French, and R.J. Woods, *Unraveling Cellulose Microfibrils: A Twisted Tale*. Biopolymers, 2013. **99**(10): p. 746-756.
13. Petridis, L., et al., *Hydration Control of the Mechanical and Dynamical Properties of Cellulose*. Biomacromolecules, 2014. **15**(11): p. 4152-4159.
14. Matthews, J.F., et al., *Computer simulation studies of microcrystalline cellulose I β* . Carbohydrate Research, 2006. **341**(1): p. 138-152.
15. Yui, T., et al., *Swelling behavior of the cellulose I β crystal models by molecular dynamics*. Carbohydrate Research, 2006. **341**(15): p. 2521-2530.
16. Mazeau, K. and L. Heux, *Molecular Dynamics Simulations of Bulk Native Crystalline and Amorphous Structures of Cellulose*. The Journal of Physical Chemistry B, 2003. **107**(10): p. 2394-2403.

17. Berglund, J., et al., *A molecular dynamics study of the effect of glycosidic linkage type in the hemicellulose backbone on the molecular chain flexibility*. The Plant Journal, 2016. **88**(1): p. 56-70.
18. Mazeau, K. and L. Charlier, *The molecular basis of the adsorption of xylans on cellulose surface*. Cellulose, 2012. **19**(2): p. 337-349.
19. Almond, A. and J.K. Sheehan, *Predicting the molecular shape of polysaccharides from dynamic interactions with water*. Glycobiology, 2003. **13**(4): p. 255-264.
20. Falcoz-Vigne, L., et al., *Quantification of a tightly adsorbed monolayer of xylan on cellulose surface*. Cellulose, 2017. **24**(9): p. 3725-3739.
21. Sauter, J. and A. Grafmüller, *Solution Properties of Hemicellulose Polysaccharides with Four Common Carbohydrate Force Fields*. Journal of Chemical Theory and Computation, 2015. **11**(4): p. 1765-1774.
22. Stark, N.M., D.J. Yelle, and U.P. Agarwal, *4 - Techniques for Characterizing Lignin*, in *Lignin in Polymer Composites*, O. Faruk and M. Sain, Editors. 2016, William Andrew Publishing. p. 49-66.
23. Besombes, S. and K. Mazeau, *Molecular dynamics simulations of a guaiacyl β -O-4 lignin model compound: Examination of intramolecular hydrogen bonding and conformational flexibility*. Biopolymers, 2004. **73**(3): p. 301-315.
24. Petridis, L. and J.C. Smith, *Conformations of Low-Molecular-Weight Lignin Polymers in Water*. ChemSusChem, 2016. **9**(3): p. 289-295.
25. Petridis, L., R. Schulz, and J.C. Smith, *Simulation Analysis of the Temperature Dependence of Lignin Structure and Dynamics*. Journal of the American Chemical Society, 2011. **133**(50): p. 20277-20287.
26. Hong, L., L. Petridis, and J.C. Smith, *Biomolecular Structure and Dynamics with Neutrons: The View from Simulation*. Israel Journal of Chemistry, 2014. **54**(8-9): p. 1264-1273.
27. Petridis, L., et al., *Self-similar multiscale structure of lignin revealed by neutron scattering and molecular dynamics simulation*. Physical Review E, 2011. **83**(6): p. 061911.
28. Stomberg, R. and K. Lundquist, *Stereochemistry of lignin structures of the β -O-4 type*, in *Nordic Pulp & Paper Research Journal*. 1994. p. 37.
29. Besombes, S., et al., *Molecular Modeling of Lignin β -O-4 Model Compounds. Comparative Study of the Computed and Experimental Conformational Properties for a Guaiacyl β -O-4 Dimer*, in *Holzforschung*. 2003. p. 266.
30. Charlier, L. and K. Mazeau, *Molecular Modeling of the Structural and Dynamical Properties of Secondary Plant Cell Walls: Influence of Lignin Chemistry*. The Journal of Physical Chemistry B, 2012. **116**(14): p. 4163-4174.
31. Durbeej, B. and L.A. Eriksson, *A Density Functional Theory Study of Coniferyl Alcohol Intermonomeric Cross Linkages in Lignin - Three-Dimensional Structures, Stabilities and the Thermodynamic Control Hypothesis*, in *Holzforschung*. 2003. p. 150.
32. Leeflang, B.R., et al., *A 1H-NMR and MD study of intramolecular hydrogen bonds in methyl β -cellobioside*. Carbohydrate Research, 1992. **230**(1): p. 41-61.
33. Sawada, D., et al., *Structure and dynamics of a complex of cellulose with EDA: insights into the action of amines on cellulose*. Cellulose, 2013. **20**(4): p. 1563-1571.

34. Aguilera-Segura, S.M., et al., *Synergistic sorption of mixed solvents in wood cell walls: experimental and theoretical approach*. Macromolecular symposia, 2019. **386**(1): p. 1900022.
35. Smith, M.D., et al., *Cosolvent pretreatment in cellulosic biofuel production: effect of tetrahydrofuran-water on lignin structure and dynamics*. Green Chemistry, 2016. **18**(5): p. 1268-1277.
36. Smith, M.D., et al., *Enhanced sampling simulation analysis of the structure of lignin in the THF-water miscibility gap*. Physical Chemistry Chemical Physics, 2016. **18**(9): p. 6394-6398.
37. Vasudevan, V. and S.H. Mushrif, *Insights into the solvation of glucose in water, dimethyl sulfoxide (DMSO), tetrahydrofuran (THF) and N,N-dimethylformamide (DMF) and its possible implications on the conversion of glucose to platform chemicals*. RSC Advances, 2015. **5**(27): p. 20756-20763.
38. Mostofian, B., et al., *Local Phase Separation of Co-solvents Enhances Pretreatment of Biomass for Bioenergy Applications*. Journal of the American Chemical Society, 2016. **138**(34): p. 10869-10878.
39. Smith, M.D., et al., *Organosolv-Water Cosolvent Phase Separation on Cellulose and its Influence on the Physical Deconstruction of Cellulose: A Molecular Dynamics Analysis*. Scientific Reports, 2017. **7**(1): p. 14494.
40. Umemura, M. and Y. Yuguchi, *Solvation of xyloglucan in water/alcohol systems by molecular dynamics simulation*. Cellulose, 2009. **16**(3): p. 361-371.
41. Durbeej, B., Y.-N. Wang, and L.A. Eriksson, *Lignin Biosynthesis and Degradation — a Major Challenge for Computational Chemistry*, in *High Performance Computing for Computational Science — VECPAR 2002: 5th International Conference Porto, Portugal, June 26–28, 2002 Selected Papers and Invited Talks*, J.M.L.M. Palma, et al., Editors. 2003, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 137-165.
42. Sangha, A.K., et al., *Molecular simulation as a tool for studying lignin*. Environmental Progress & Sustainable Energy, 2012. **31**(1): p. 47-54.
43. Yuguchi, Y., et al., *Gelation of xyloglucan in water/alcohol systems*. Cellulose, 2004. **11**(2): p. 203-208.
44. Yamanaka, S., et al., *Gelation of tamarind seed polysaccharide xyloglucan in the presence of ethanol*. Food Hydrocolloids, 2000. **14**(2): p. 125-128.
45. Nishiyama, Y., P. Langan, and H. Chanzy, *Crystal Structure and Hydrogen-Bonding System in Cellulose I β from Synchrotron X-ray and Neutron Fiber Diffraction*. Journal of the American Chemical Society, 2002. **124**(31): p. 9074-9082.
46. Gomes, T.C.F. and M.S. Skaf, *Cellulose-Builder: A toolkit for building crystalline structures of cellulose*. Journal of Computational Chemistry, 2012. **33**(14): p. 1338-1346.
47. Pettersen, R.C., *The Chemical Composition of Wood*, in *The Chemistry of Solid Wood*. 1984, American Chemical Society. p. 57-126.
48. Ebringerová, A. and T. Heinze, *Xylan and xylan derivatives – biopolymers with valuable properties, I. Naturally occurring xylans structures, isolation procedures and properties*. Macromolecular Rapid Communications, 2000. **21**(9): p. 542-556.
49. Woods group. Complex Carbohydrate Research Center, U.o.G., Athens, GA. (<http://glycam.org>). *GLYCAM Web*. 2005-2019.

50. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers*. SoftwareX, 2015. **1–2**: p. 19-25.
51. Bekker, H., et al. *Gromacs: A parallel computer for molecular dynamics simulations*. in *Physics computing*. 1993.
52. Berendsen, H.J.C., D. van der Spoel, and R. van Drunen, *GROMACS: A message-passing parallel molecular dynamics implementation*. Computer Physics Communications, 1995. **91**(1): p. 43-56.
53. Van Der Spoel, D., et al., *GROMACS: Fast, flexible, and free*. Journal of Computational Chemistry, 2005. **26**(16): p. 1701-1718.
54. Páll, S., et al., *Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS*, in *Solving Software Challenges for Exascale: International Conference on Exascale Applications and Software, EASC 2014, Stockholm, Sweden, April 2-3, 2014, Revised Selected Papers*, S. Markidis and E. Laure, Editors. 2015, Springer International Publishing: Cham. p. 3-27.
55. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water*. The Journal of Chemical Physics, 1983. **79**(2): p. 926-935.
56. Dick, T.J. and J.D. Madura, *Chapter 5 A Review of the TIP4P, TIP4P-Ew, TIP5P, and TIP5P-E Water Models*, in *Annual Reports in Computational Chemistry*. 2005, Elsevier. p. 59-74.
57. Mark, P. and L. Nilsson, *Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K*. The Journal of Physical Chemistry A, 2001. **105**(43): p. 9954-9960.
58. Guvench, O., et al., *Additive empirical force field for hexopyranose monosaccharides*. J Comput Chem, 2008. **29**.
59. Guvench, O., et al., *CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses*. J Chem Theory Comput, 2009. **5**.
60. Petridis, L. and J.C. Smith, *A molecular mechanics force field for lignin*. Journal of Computational Chemistry, 2009. **30**(3): p. 457-467.
61. Petridis, L. and J.C. Smith, *Cellulosic ethanol: progress towards a simulation model of lignocellulosic biomass*. Journal of Physics: Conference Series, 2008. **125**(1): p. 012055.
62. van der Spoel, D., P.J. van Maaren, and C. Caleman, *GROMACS molecule & liquid database*. Bioinformatics, 2012. **28**(5): p. 752-753.
63. Eisenhaber, F., et al., *The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies*. Journal of Computational Chemistry, 1995. **16**(3): p. 273-284.
64. Luzar, A. and D. Chandler, *Hydrogen-bond kinetics in liquid water*. Nature, 1996. **379**: p. 55.
65. Luzar, A., *Resolving the hydrogen bond dynamics conundrum*. The Journal of Chemical Physics, 2000. **113**(23): p. 10663-10675.
66. van der Spoel, D., et al., *Thermodynamics of Hydrogen Bonding in Hydrophilic and Hydrophobic Media*. The Journal of Physical Chemistry B, 2006. **110**(9): p. 4393-4398.
67. Geudtner, G., et al., *deMon2k*. Wiley Interdisciplinary Reviews: Computational Molecular Science, 2012. **2**(4): p. 548-555.

68. de la Lande, A., et al., *Molecular Simulations with in-deMon2k QM/MM, a Tutorial-Review*. *Molecules*, 2019. **24**(9): p. 1653.
69. Perdew, J.P., K. Burke, and M. Ernzerhof, *Generalized Gradient Approximation Made Simple*. *Physical Review Letters*, 1996. **77**(18): p. 3865-3868.
70. Perdew, J.P., K. Burke, and M. Ernzerhof, *Generalized Gradient Approximation Made Simple [Phys. Rev. Lett. 77, 3865 (1996)]*. *Physical Review Letters*, 1997. **78**(7): p. 1396-1396.
71. Zhang, Y. and W. Yang, *Comment on "Generalized Gradient Approximation Made Simple"*. *Physical Review Letters*, 1998. **80**(4): p. 890-890.
72. Lee, C., W. Yang, and R.G. Parr, *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density*. *Physical Review B*, 1988. **37**(2): p. 785-789.
73. Godbout, N., et al., *Optimization of Gaussian-type basis sets for local spin density functional calculations. Part I. Boron through neon, optimization technique and validation*. *Canadian Journal of Chemistry*, 1992. **70**(2): p. 560-571.
74. Köster, A.M., J.U. Reveles, and J.M. del Campo, *Calculation of exchange-correlation potentials with auxiliary function densities*. *The Journal of Chemical Physics*, 2004. **121**(8): p. 3417-3424.
75. Jorgensen, W.L., 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*. *Journal of the American Chemical Society*, 1996. **118**(45): p. 11225-11236.
76. Maurer, R.J., A.F. Sax, and V. Ribitsch, *Molecular simulation of surface reorganization and wetting in crystalline cellulose I and II*. *Cellulose*, 2013. **20**(1): p. 25-42.
77. Miyamoto, H., et al., *Molecular dynamics simulation of dissociation behavior of various crystalline celluloses treated with hot-compressed water*. *Cellulose*, 2014. **21**(5): p. 3203-3215.
78. Aguilera-Segura, S.M., F. Di Renzo, and T. Mineva, *Structures, intermolecular interactions, and chemical hardness of binary water-organic solvents: a molecular dynamics study*. *J Mol Model*, 2018. **24**: p. 292.
79. Goldmann, W.M., et al., *Solubility and fractionation of Indulin AT kraft lignin in ethanol-water media*. *Separation and Purification Technology*, 2019. **209**: p. 826-832.

Chapter 5. Solvent effect on the adhesion between lignocellulose compounds

Contents

Abstract	175
1. Introduction.....	175
2. Computational details	177
2.1 Models	177
2.2 Molecular Dynamics simulations	178
3. Results and discussion.....	181
3.1 Solvent effect on the adhesion between lignin and cellulose	181
3.2 Solvent effect on the adhesion between xylan and lignin.....	189
4. Conclusions.....	195
References.....	196

Abstract

Understanding the factors that drive the interaction of lignocellulose components with different solvents is the key to successful optimisation of fractionation processes. The effect of binary water-organic solvents on the adhesion of lignocellulose components was studied with molecular dynamics simulations, considering ethanol and acetonitrile as organic co-solvents in binary water-organic solvent mixtures. The lignin-xylan and lignin-cellulose binding is significantly disrupted in the mixed solvents, which could be explained with preferential interactions between different solvent components and the individual polymers. In pure water, lignin-cellulose interactions are enhanced, and long-lasting H-bonds between lignin and cellulose are observed. On the contrary, xylose-lignin interactions are rapidly interrupted upon hydration of the xylose chain. The analysis of the radial distribution functions and interaction energies shows that water-organic solvent mixtures act together to disrupt more effectively the cellulose-lignin interactions. However, our analyses show that the presence of the organic phase can enhance and slow down the detachment between xylose and lignin. This suggests that a detailed description of the intermolecular interactions between lignocellulose models and solvent molecules can provide better insight into the effect of water-organic mixtures on the xylan-lignin and lignin-cellulose adhesion.

1. Introduction

Understanding how pre-treatment will affect biomass chemical and physical features can offer meaningful improvements in biomass conversion [1]. The use of non-aqueous solvents has provided remarkable results in fractionation processes of several levels of severity, ranging from swelling, dilute acid, hydrothermal, steam explosion, alkaline treatment, to organosolv pulping [2-13]. Non-linear effects of aqueous solutions of organics have been observed in solvent fractionation techniques, where aqueous-organic solvent solutions led to higher lignin solubilisation and extraction of polyphenols than pure solvents [14, 15]. Recent experimental work based on optical microscopy experiments has evidenced a detachment between the primary wall and middle lamella of wood cells in pure and ethanol aqueous solutions [16]. However, the current understanding of the solvent effect on the interactions within the cell wall is limited. Hence, molecular-scale studies achieved by experimental and theoretical approaches are needed for the understanding and optimisation of biorefinery processes [17-21].

Molecular dynamics (MD) simulations, often in combination with experimental investigations, have provided valuable information about the adhesion between lignocellulose components -i.e., cellulose, hemicellulose, and lignin- [22-42]. The complexity and variability of both lignin and hemicelluloses present a challenge, but significant advances have been developed to lower the barrier for modelling these biomolecules [36, 37, 43]. Considerable theoretical work has been performed in aqueous phase to gain understanding of the interactions between components of the cell wall and the mechanisms for wood deconstruction during fractionation [22-

42, 44]. This is motivated by the intrinsic presence of water in the cell wall structure and the extensive use of steam explosion pre-treatment [45] for biorefineries. The theoretical and experimental results of Langan et al. [39, 44] sensitive at different scales, for example, have elucidated processes responsible for morphological changes in biomass during pretreatment, through cellulose dehydration and lignin-hemicellulose phase separation. It is highlighted that strong van de Waals forces between lignin molecules and cellulose microfibrils are responsible for higher adhesion energy in the complex [25, 28] showing that hemicellulose has stronger mechanical properties than lignin whereas lignin exhibits greater tendency to adhere to cellulose microfibrils.

Because of the amount of solvents available, and the computational cost to consider different non-aqueous solvents and compositions, only few theoretical studies approach the interactions of the biomass components with pure or binary mixed solvents [46-57]. However, the non-aqueous solvent effect on the adhesion between lignocellulose components is still unexplored. The non-aqueous solvent effect has been addressed in studies of conformation and dynamics of individual lignocellulose compounds, and the results have provided insight on the mechanisms of biomass deconstruction of the lignocellulose compounds in water-cosolvent mixtures at different scales [46-57]. Nonetheless, the variety and complexity of the lignocellulose biomass makes the modelling of all component interactions complex and somewhat impractical.

However, the systematic modelling of pairwise interactions between the main lignocellulose compounds can provide insight into the fundamental interactions on the cell wall. Recently, Yang et al. [25] studied at the quantum level the solvent effect on the pairwise adhesion between several tetramer models of cell wall components using polarised continuum models to represent water, methanol, and chlorophrom. Their results show that implicit solvation methods does not replicate the interaction energies calculated with explicit solvation of water molecules, as expected, because polarised continuum models do not consider the strong short-range interactions such as hydrogen bonding or vdW interactions. Moreover, their results show that the variations in model interaction with the solvent dielectric constant in the polarised continuum model are significant and non-systematic [25]. Thus, the explicit inclusion of the solvent is fundamental to consider the strong short-range interactions. The quantum inclusion of the solvent is computationally expensive and can result in local potential energy minima [25]; however, MD simulation of explicitly solvated interactions can provide a compromise between cost and accuracy.

Thus, we aim to gain understanding at the molecular level of the interactions between wood components in the presence of pure and water-organic solvent mixtures. In the present Chapter, we assess the evolution of xylose-lignin and cellulose-lignin complexes in the presence of water-ethanol and water-acetonitrile binary solvents at 25wt. % (only xylose-lignin), 50wt. % and 75wt. % concentrations. The results are compared to those obtained for the respective pure solvents. To achieve the scope, we carried out a detailed analysis of the MD dynamics of the model polymer complexes using the non-covalent interaction energies, number of H-bonds, and distribution functions (RDFs) as indicators.

2. Computational details

2.1 Models

To study the cellulose-lignin interactions, we used the cellulose nanocrystal (reported in Chapter 4) in complex with four lignin dimers, illustrated in Figure 5.1A. The cellulose I β nano-crystallite [58-60] was built with seven cellulose chains, eight-monomers long (56 anhydroglucose units) and two β -O-4 linked guaiacyl monomers formed the lignin dimers. Each lignin dimer was placed close to the cellulose surfaces, at a distance below 3 Å of the atomic centre of the closest atom in order to favour the initial cellulose-lignin interactions.

To study the xylan-lignin interactions, we used the complex of the xylan (β -(1 \rightarrow 4) linked xylose) [43] and lignin (β -O-4 linked guaiacyl) tetramer models studied in chapter 4. The xylan-lignin complex was built by aligning the tetramer molecules at a distance below 3 Å. Then, we performed a 10-ns MD in gas phase to let the xylan-lignin system to complex in the absence of solvent. This equilibration resulted in a folded lignin conformer wrapped by the xylan tetramer, as shown in Figure 5.1 B. In the absence of the solvent, xylan undergoes a transition from β -(1 \rightarrow 4) to α -(1 \rightarrow 4) bonds to favour the intermolecular interactions. However, the β -(1 \rightarrow 4) conformation is recovered during the solvated simulations.

The xylan-lignin and cellulose-lignin complexes were centred in cubic boxes, leaving at least one nm to the longest side of the model, avoiding interactions between the cellulose or lignin models and their images in the neighbouring boxes. We used this rim spacing to determine the size of the box. Each system was further solvated with the number of solvent molecules needed to fill the box size, summarised in Table 5.1. Each organic solvent, i.e. ethanol and acetonitrile, was examined at three different concentrations: pure, 75 wt.%, and 50 wt.% of organic solvent (25% was also evaluated for the xylan-lignin complex), and compared with the simulations in pure water.

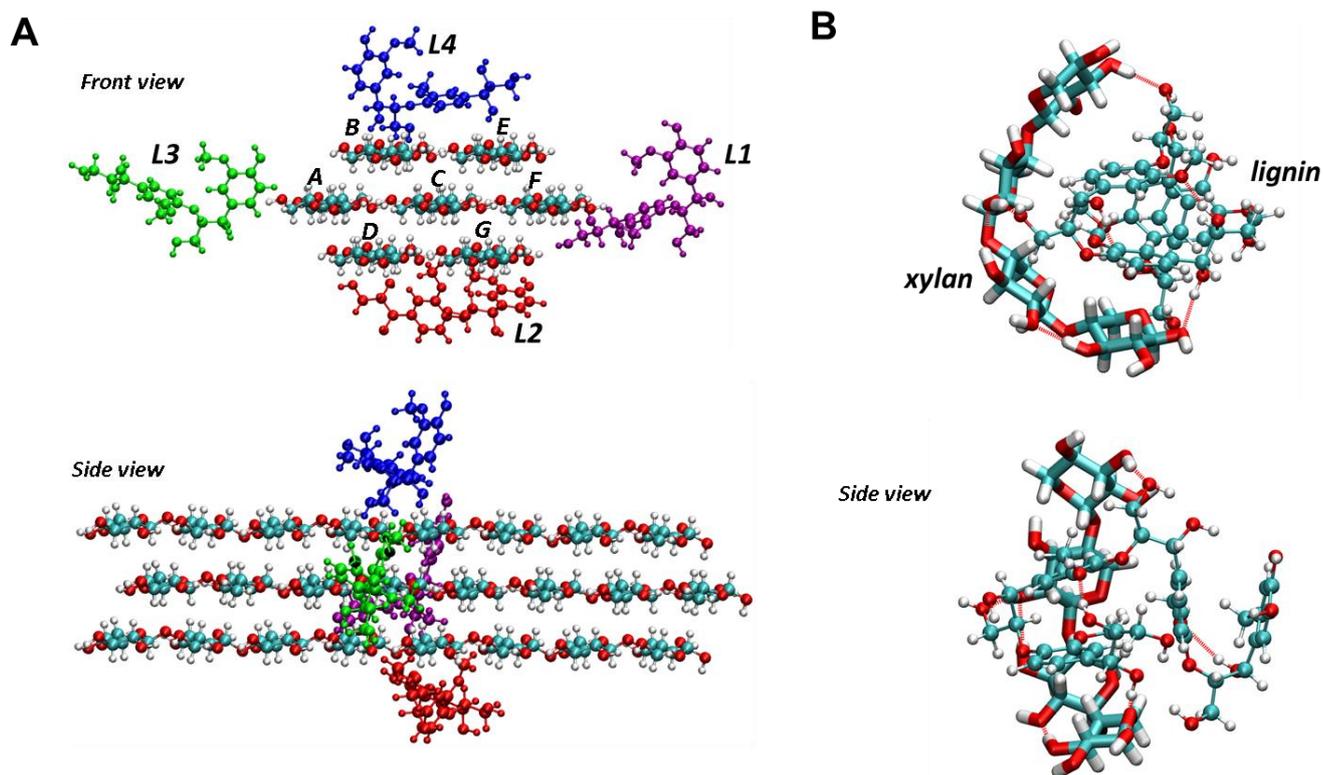


Figure 5. 1. (A) lignin-cellulose complex between the cellulose nanocrystal model with seven chains (ABCDEFG) and four lignin dimers. Cellulose chains AF are referred to as the hydrophilic surface while chains BDEG are referred to as the hydrophobic surface. Lignin dimer colour code of each of the four dimers in the complex: L1 (purple), L2 (red), L3 (green), and L4 (blue). Top: front view; Down: side view. (B) xylan-lignin complex between β -(1 \rightarrow 4) xylan and β -O-4 guaiacyl tetramers. Top: view of bent xylan and stacked lignin; down: detail of lignin ring-ring stacking in side view. Colour code CPK and licorice representations: oxygen (red), hydrogen (white), and carbon (cyan). Hydrogen bonds are shown in red dashed lines.

2.2 Molecular Dynamics simulations

All-atom Molecular Dynamics (MD) simulations of each system described in Table 5.1 were carried out using the GROMACS package version 2016.3 [61-65], along with the 4-sites Transferable Intermolecular Potential (TIP4) for liquid water [66-68], the CHARMM36 additive force field [69, 70], and the CHARMM-compatible force field for lignin [71]. Solvent structure for the organic solvents was available at the GROMACS molecule and liquid database [72]. For each simulation box, energy minimisation was performed using the steepest descent algorithm until convergence to a tolerance of $100 \text{ kJ mol}^{-1} \text{ nm}^{-1}$. After minimisation, restrained simulations were performed for 200 ps at 298.15 K to allow solvent equilibration around the polymers. Afterwards, 20-ns MD simulations were performed with a frame-saving rate (for analysis) of 1 ps, in order to study the interaction of the cellulose-lignin complex in the solvent mixtures. 10-ns MD simulations were performed for xylan-lignin complex due to the reduced size of the models in comparison with the cellulose nanocrystal. Temperature and pressure coupling were handled using the leap-frog stochastic dynamics integrator and the Parrinello-Rahman method, respectively. Initial velocities were generated from a Maxwell distribution at 298.15 K and the isothermal-isobaric (NPT) ensemble was considered for data collection.

Neighbour searching and short-range non-bonded interactions were handled with the Verlet cut-off scheme. Electrostatics were treated with the Fast smooth Particle-Mesh Ewald (SPME) method, with a Coulomb cut-off of 1.2 nm, a fourth order interpolation and Fourier spacing of 0.12 nm. Vander Waals (vdW) interactions were treated using the Lennard- Jones potential with a cut-off distance of 1.2 nm. Simulations were carried out in an Intel Xeon CPU with 2.10Gz with 32 logical cores.

Table 5.1. Configuration of simulated systems and equilibrium size of simulation boxes for the xylan-lignin and cellulose-lignin complexes studied in this work Solvents studied include water, ethanol (EtOH), acetonitrile (ACN), and their binary aqueous mixtures.

Model wood component	Solvent system	Cosolvent fraction (wt. %)	Number of cosolvent molecules	Number of water molecules	Cubic box side length (nm)	Volume (nm ³)
Xylan + lignin tetramers	water	0	0	5578	5.55	170.82
	EtOH-water	50	513	3941	5.52	168.54
		75	975	2496	5.52	168.66
		50	1381	1178	5.52	168.72
	ACN-water	25	565	3868	5.52	168.48
		50	1058	2412	5.52	168.53
		75	1492	1134	5.52	168.57
	EtOH	100	1621	0	5.42	159.42
	ACN	100	1843	0	5.48	164.67
Cellulose nanocrystal + 4 lignin dimers	water	0	0	10370	6.886	326.56
	EtOH-water	50	1725	4415	6.749	307.46
		75	2443	2084	6.749	307.48
	ACN-water	50	1760	4013	6.615	
		75	2639	2006	6.746	307.056
	EtOH	100	2985	0	6.711	302.31
	ACN	100	3442	0	6.813	316.266

The structure, dynamics, and interactions of the wood component models with the pure and water-organic solvent mixtures were characterised using the incorporated tools within GROMACS. To quantify the solvent effect, we computed intermolecular energies, radial distribution functions $g(r)$, and the average number of

hydrogen bonds. Each of these descriptors was calculated for all the water-organic solvent concentrations considered in this study. Site-to-site $g(r)$ were computed for the water-polymer and organic solvent-polymer pairs (see Figure 5.2 for atom numbering) considering the polymer sites as follows: in cellulose, we considered the hydroxyl (O2 and O3) and hydroxymethyl (O6), the glycosidic bond (O4), and the monosaccharide ring (O5)). In lignin, we studied the solvation around the β -O-4 bond (O α , O β , and O γ) and the methoxy groups (OMe). The solvent sites considered here are the water O, ethanol O, and acetonitrile N, along with the methyl group carbon of ethanol and acetonitrile. Here, $g(r)$ is normalised by the number of reference points and the volume of the shell. Thus, $g(r)$ is expressed as number density (atoms/nm³) per monomer, and it tends to the bulk density of the particle as r increases. H-bonds were calculated using a geometrical criterion with a maximum donor-acceptor distance of 0.35 nm and a hydrogen-donor-acceptor angle of 30°.

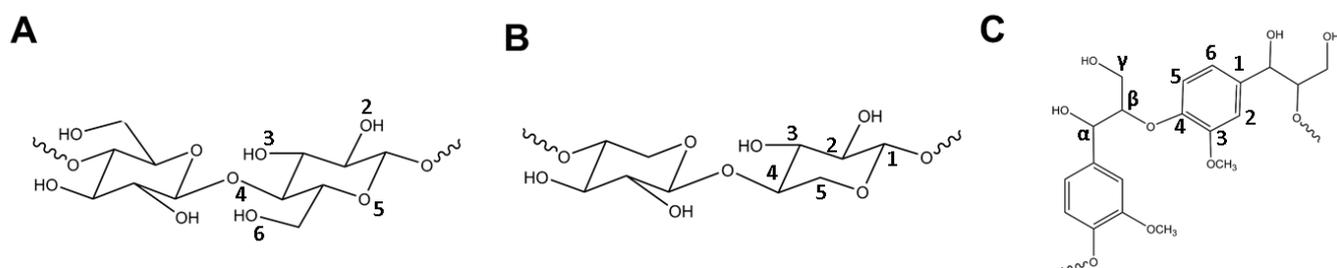


Figure 5.2. (A) Cellulose cellobiose fundamental unit. (B) Fundamental unit of xylan hemicellulose, a xylosyl dimer, bonded with a β -(1 \rightarrow 4) linkage. (C) A guaiacyl dimer (G-G) connected with a β -O²-4' linkage.

3. Results and discussion

3.1 Solvent effect on the adhesion between lignin and cellulose

3.1.1 Assessment of the interaction strength between lignin and cellulose in solvents and the gas-phase.

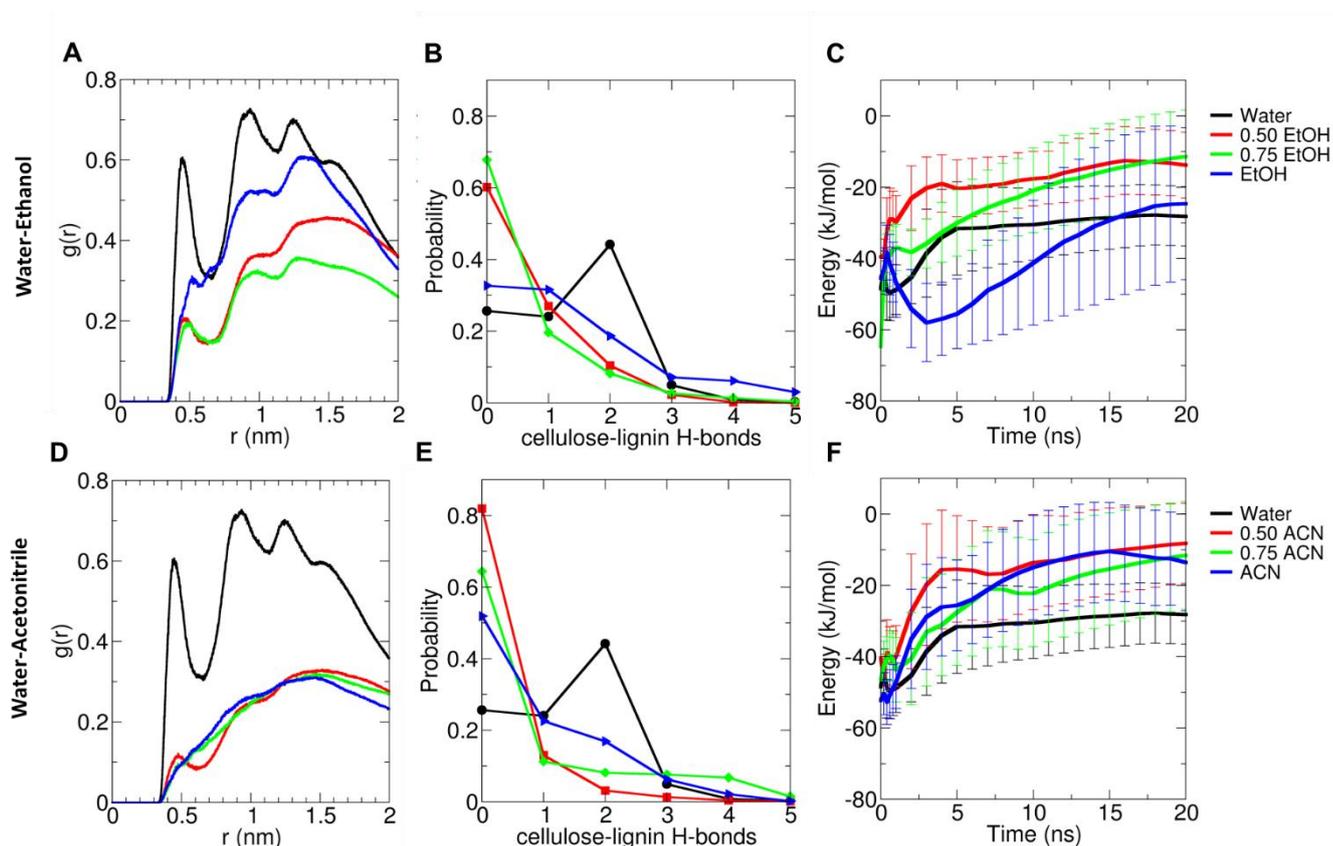


Figure 5.3. Cellulose-lignin interactions in the presence of water-ethanol (A-C) and water-acetonitrile (D-F) solvent mixtures at 50 wt. % and 75 wt. % and pure concentrations. (A, D) Cellulose-lignin radial distribution function, $g(r)$, between the surface cellulose monomers and the lignin phenolic rings, in pure and aqueous organic solvent mixtures. Only the carbons of glucose and the aromatic ring, respectively, have been used in this calculation. (B, E) Probability distribution of hydrogen bonds between lignin dimers and surface cellulose, in pure and aqueous organic solvent mixtures. H-bond intervals are discrete, but lines have been added to describe better the trends. (C, F) Average lignin-cellulose intermolecular energies along the 20-ns MD normalised by the number of lignin dimers in the simulation along the 20-ns MD simulations.

We evaluated the dynamics of the cellulose nanocrystal initially in complex with four lignin dimers shown in Figure 5.1A, in pure and binary mixtures of water-ethanol and water-acetonitrile. The interaction between cellulose and lignin dimers was estimated from the computed cellulose-lignin radial distribution functions (i.e. $g(r)$, in Figure 5.3A, D), the cellulose-lignin H-bond distributions (Figure 5.3B, E), and the averaged interaction energies during the dynamics (Figures 5.3C, F and Table 5.2). The cellulose-lignin $g(r)$ between the carbons of the aromatic and saccharide rings show that it is more likely to find lignin coordinating

cellulose in pure water. The presence of ethanol and acetonitrile decrease the cellulose-lignin coordination significantly, and the effect becomes more notorious in water-organic mixtures, particularly in the water-acetonitrile solvents.

The cellulose-lignin adhesion is controlled by both van der Waals and H-bond interactions [33]; therefore, we analysed the H-bond frequency since the radial distribution functions do not explicitly describe hydrogen bonding. As follows from the H-bond analysis, the cellulose-lignin H-bond distribution (Figure 5.3B) shows the highest probability of finding lignin hydrogen-bonded to cellulose ($H\text{-bond} \geq 1$) in water during more than ~75% of the simulation time, followed by ethanol, in which lignin remains bounded to cellulose more than 65% of the time. In pure acetonitrile, the bounded state is reduced to ~50%; however, lignin remains mostly unbounded in the mixed solvents, and the probability of forming H-bond with cellulose is less than ~40%. Accordingly, the average cellulose-lignin interaction energies in Figure 5.3C show the strongest average cellulose-lignin interaction in the presence of water and pure ethanol. We observe that while the adhesion between lignin and cellulose decreases rapidly in the presence of acetonitrile and mixed water-organic solvent, it tends to increase as a first response in the presence of pure water and ethanol. The cellulose-lignin adhesion in water stabilises after five ns, whereas the interaction in the presence of ethanol decreases slowly. Thus, the lignin-cellulose interactions are in average stronger in pure water and ethanol than in pure acetonitrile and mixed water-acetonitrile or water-ethanol solutions (Figure 3C, 3F and Table 5.2, column 3). The weakest adhesion between cellulose and acetonitrile is obtained in 50% acetonitrile-water mixture. Moreover, the interaction energies curves pass through a minimum interaction at 50% acetonitrile-water mixtures and 75% ethanol-water (also 50% ethanol during the first ns).

Table 5.2. Average Lennard-Jones (LJ) and Coulomb (Coul) lignin-cellulose energies with total standard deviation errors. Values are normalised by the number of lignin dimers and presented in kJ/mol.

Solvent	Average Energies and standard deviation errors		
	LJ	Coul	Total
Water	-17.0 ± 6.5	-11.2 ± 5.5	-28.2 ± 8.3
0.50 EtOH	-9.0 ± 5.8	-4.8 ± 4.8	-13.8 ± 9.3
0.75 EtOH	-6.7 ± 7.5	-4.7 ± 6.6	-11.4 ± 13.0
EtOH	-13.7 ± 11.6	-10.9 ± 10.7	-24.6 ± 21.3
0.5 ACN	-5.2 ± 7.1	-3.0 ± 5.0	-8.2 ± 11.2
0.75 ACN	-5.5 ± 7.1	-6.1 ± 9.1	-11.6 ± 15.1
ACN	-6.1 ± 6.9	-7.4 ± 8.2	-13.5 ± 13.6
No Solvent	-65.1 ± 4.0	-86.1 ± 8.5	-151.2 ± 9.4

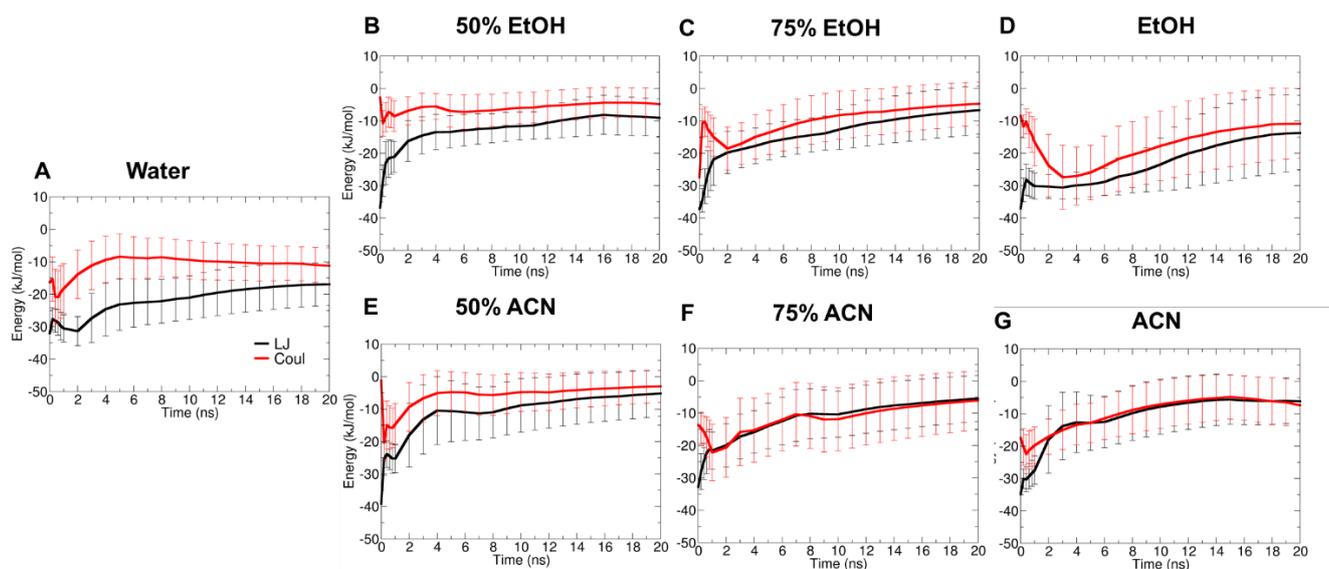


Figure 5.1. Lennard-Jones (LJ, black) and Coulomb (Coul, red) average lignin-cellulose energies along the 20-ns MD normalised by the number of lignin dimers.

To understand the factors responsible for the lignin-cellulose adhesion in the presence of the solvent we computed the Lennard-Jones (LJ) and Coulomb (Coul) contributions to the cellulose-lignin interaction energies, and the results are summarized in Table 5.2 and Figure 5.4. In the same Figure and Table, the standard deviation is reported, thus assessing the error bars of the calculated energies, averaged over the interaction energies of the 7-chain cellulose with the four lignin dimers. As follows, the error bars for pure organic and aqueous organic binary mixtures are as large as the magnitude of the energy values. These large magnitudes of the errors are because of the enhanced mobility of three of the four lignin dimers, not adhering to cellulose (*vide infra*). Although the error bars are significant, the energy trends, representative for the whole model and the conclusions remain unaltered due to the predominant contribution of only one lignin dimer, which adheres to cellulose.

The LJ contribution to the interaction energy is stronger than the Coulomb electrostatic interaction in water, ethanol, water-ethanol mixtures and 50 wt. % water-acetonitrile mixture. However, the Coulomb electrostatic interaction is slightly stronger in pure and 75 wt. % acetonitrile. The stronger affinity of lignin to cellulose can be easily explained by their stronger association (LJ) and the reduction of the hydrophobic lignin-water interactions, in agreement with the observation that lignin binding to cellulose is favoured in the presence of water [42]. The evolution of the LJ and Coulomb energies along the trajectory (Figure 5.4) shows that whereas the presence of pure water tends to enhance the associative LJ interactions, the cosolvents act at the first contact by reducing them. However, we notice that the Coulomb interaction in the cellulose-lignin ensemble prevails in water and ethanol, and this effect is less evident in pure acetonitrile and the water-organic mixtures. Thus, our results suggest that while the van der Waals (i.e., LJ) lignin-cellulose interactions are more disrupted in the presence of cosolvents, protic solvents such as water and ethanol are less effective

disruptors than the aprotic acetonitrile. This can be explained with the enhancement of the Coulomb electrostatic interactions in the lignin-cellulose ensemble.

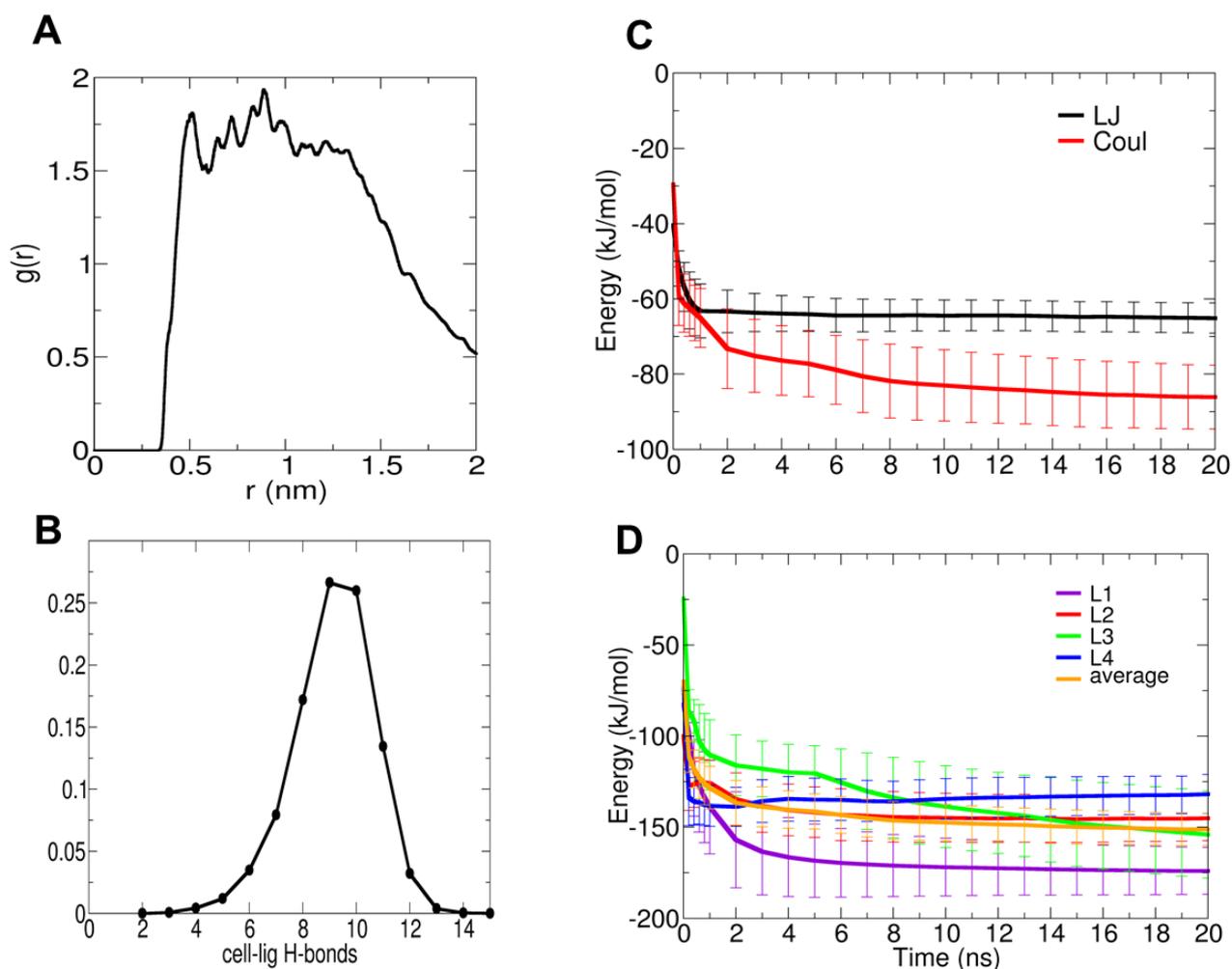


Figure 5.2. Cellulose-lignin interaction in the absence of solvent. (A) Cellulose-lignin radial distribution function, $g(r)$, between the surface cellulose monomers and the lignin phenolic rings, in pure and aqueous organic solvent mixtures. Only the carbons of glucose and the aromatic ring, respectively, have been used in this calculation. (B) Distribution of hydrogen bonds between lignin dimers and surface cellulose. H-bond intervals are discrete, but lines have been added to describe better the trends. (C) Average lignin-cellulose Lennard-Jones (LJ) and Coulomb (Coul) intermolecular energies per lignin dimer along the 20-ns MD simulations. (D) Lignin-cellulose intermolecular energies for each lignin dimer and their average along the 20-ns MD simulations.

Furthermore, we studied the cellulose-lignin (C-L) interactions in the absence of the solvents (Figure 5.5) to understand better the solvent effect. The descriptors of this C-L interaction in gas phase are the cellulose-lignin $g(r)$ (Figure 5.5A), the H-bond distribution (Figure 5.5B), the average interaction energies (Figure 5.5C and Table 5.2), and the cellulose-lignin interactions for each lignin dimer (Figure 5.5D) in the absence of the solvent. As expected, in the gas phase, the coordination between cellulose saccharides and phenolic lignin rings (Figure 5.5A) is 5-times higher when compared to the cellulose-lignin coordination in water (Figure 5.3A). Moreover, the first maximum in the RDF distribution falls at 5.0 Å, while in the presence of the solvent the first RDF maximum is below, at ~4.6 in all solvents studied (except ethanol). Thus, the presence of the

solvent results in a reduced coordination of lignin dimers to cellulose, which means less dimers are in contact with cellulose. The presence of the solvents, however, slightly improves the C-L complexation of the remaining lignin dimers as suggested from the $g(r)$ first maximum falling at a somewhat shorter distance. The H-bond distribution falls at 9-10 H-bonds; therefore, the hydrogen bond interactions are maximised in the absence of the solvent and strongly disrupted upon its addition (Figure 5.3B). Furthermore, the average interaction energies are also 5-times higher compared to water (Table 5.2) and are dominated by Coulomb interactions because of the high frequency of H-bonds, as illustrated in Figure 5.5C. Moreover, the cellulose-lignin LJ interactions without solvent are on average up to 4 times stronger than in water, but the Coulomb interactions are up to 8 times stronger. Thus, the addition of solvent is to primarily decrease the electrostatic interactions by disrupting the hydrogen bonding between lignin and cellulose, which separates the two polymers and subsequently decreases their LJ interactions.

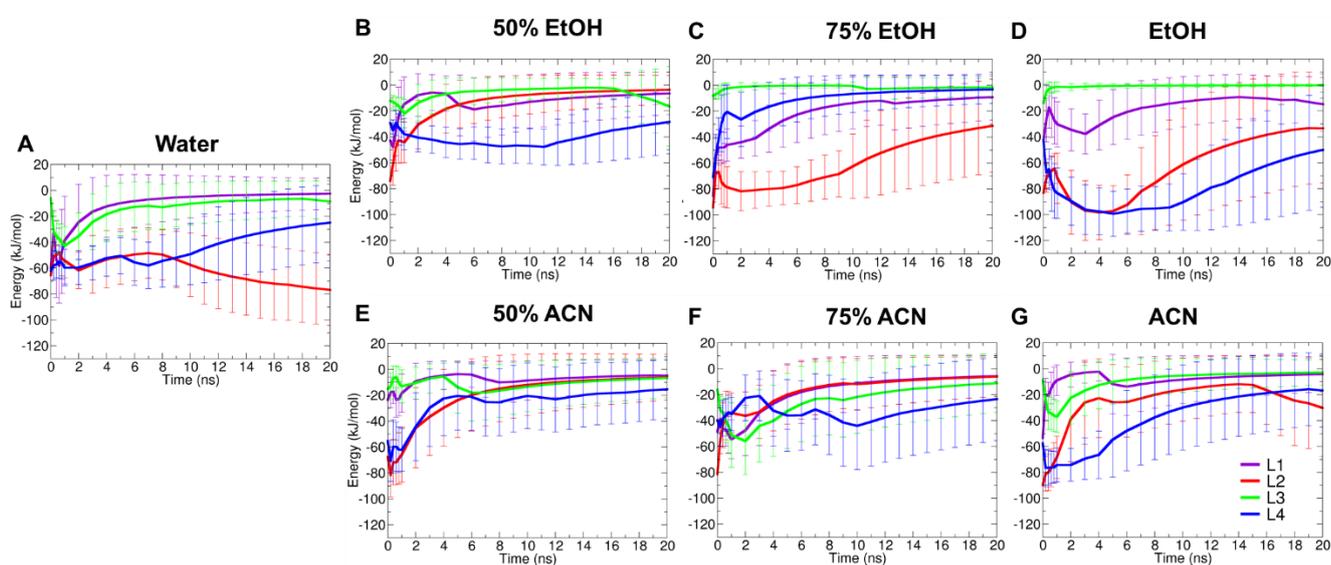


Figure 5.3. Evolution of average lignin-cellulose intermolecular energies between each lignin dimer along the 20-ns MD simulation. The colour code for each lignin dimer is L1 (purple), L2 (red), L3 (green), and L4 (blue), as in Figure 5.1A.

Furthermore, the cellulose-lignin interactions for each lignin dimer (see Figure 5.1B for notation) in the absence of solvent (Figure 5.5D) shows that lignin dimers L1 (purple) and L3 (green) initially in contact with the hydrophilic chains tend to interact stronger than L2 and L4. The evolution of the interactions between each lignin dimer and cellulose in the presence of solvents (Figure 5.6 and Table 5.3), on the contrary, shows that lignin dimers L2 (red) and L4 (blue) display in average stronger interaction energies with cellulose than lignin dimers L1 and L3. The interaction energies between lignin and the hydrophobic and hydrophilic cellulose surfaces summarised in Table 5.3, indeed, show that L1 interacts stronger with the hydrophilic cellulose surface whereas L2 and L4 interact preferentially with the hydrophobic surface in the absence of the solvent. L3, on the other hand, shows a tendency to interact stronger as it becomes closer to the hydrophobic chains. Furthermore, these observations are in agreement with previous authors that have shown that lignin molecules

preferentially aggregate into the hydrophobic faces of crystalline cellulose fibres [38, 42]. In the absence of solvent, however, the interactions with the hydrophilic surfaces are favoured. This is most likely because the lack of cellulose-solvent or lignin-solvent interactions modifies both the cellulose and lignin conformations and therefore their characteristics as hydrophilic or hydrophobic in gas phase become obsolete.

Table 5.3. Average lignin-cellulose intermolecular energies in kJ/mol between each lignin dimer and the hydrophilic (chains AF) or hydrophobic (chains BDEG) surface chains of cellulose, as a function of cosolvent fraction (Xcosol). Lignin number and colour code is L1 (purple), L2 (red), L3 (green), and L4 (blue), as in Figure 5.1A.

Solvent		No Solvent	Water	Ethanol			Acetonitrile		
Xcosol		-	0	0.50	0.75	1	0.50	0.75	1
Hydrophilic	L1	-132.0 ± 1.9	-1.6 ± 1.0	-3.9 ± 2.0	-6.5 ± 4.3	-13.0 ± 4.9	-2.5 ± 1.2	-4.2 ± 2.8	-0.5 ± 0.3
	L2	-7.6 ± 0.1	-1.1 ± 0.2	-1.0 ± 0.8	-0.4 ± 0.2	-0.9 ± 0.2	-0.2 ± 0.2	-0.3 ± 0.3	-12.8 ± 6.3
	L3	-61.5 ± 3.8	-3.9 ± 1.5	-4.0 ± 1.9	-1.6 ± 1.2	-0.1 ± 0.1	-1.2 ± 0.6	-6.9 ± 3.9	-2.4 ± 1.5
	L4	-4.4 ± 0.2	-1.6 ± 0.8	-1.8 ± 0.9	0.0 ± 0.0	-1.0 ± 0.1	-2.2 ± 1.1	-15.7 ± 5.9	-8.1 ± 4.4
Hydrophobic	L1	-42.0 ± 0.2	-1.0 ± 0.7	-2.6 ± 1.8	-2.6 ± 1.2	-1.8 ± 0.8	-2.1 ± 1.0	-1.5 ± 1.1	-3.7 ± 2.8
	L2	-137.5 ± 1.3	-75.7 ± 9.1	-2.7 ± 2.1	-31.0 ± 12.0	-32.5 ± 13.2	-5.9 ± 4.2	-5.8 ± 3.5	-17.4 ± 6.2
	L3	-92.5 ± 8.4	-4.6 ± 1.8	-12.6 ± 8.7	0.0 ± 0.0	-0.1 ± 0.1	-5.5 ± 3.8	-4.2 ± 1.6	-0.5 ± 0.3
	L4	-127.4 ± 2.0	-23.2 ± 9.5	-26.6 ± 7.9	-3.3 ± 2.4	-48.9 ± 14.3	-13.1 ± 4.0	-7.7 ± 3.0	-8.7 ± 4.6

In the presence of solvent, moreover, the interactions of L1 and L3 with both cellulose surfaces are drastically reduced, accordingly with the reduction of the hydrogen bonding in Figure 3B and 5B. Moreover, lignin dimers L2 and L4 display the highest recalcitrance in all the solvents, and they are found to interact preferentially with the hydrophobic surfaces, except in pure and 75 wt. % acetonitrile. L2 and L4 in pure and 75 wt. % acetonitrile, respectively, are found to have a significant interaction with the hydrophilic surface, which can explain the slightly stronger electrostatic interaction observed in these solvents (Figure 5.4 and Table 5.2). Thus, our results suggest that the preferential interaction of acetonitrile with hydrophobic surfaces of cellulose (discussed in chapter 4) disrupts the lignin-cellulose adhesion, but due to acetonitrile protic character, this solvent also enhances the interaction of lignin with the surfaces with solvent-exposed polar groups due to its aprotic character. Pure water and ethanol, on the other hand, disrupt more effectively the cellulose-lignin interactions with hydrophilic surfaces, but are less effective at disrupting the associative interactions with the saccharide rings in hydrophobic cellulose surfaces. Thus, our results demonstrate that the action of water-organic solvent mixtures is more efficient at disrupting both electrostatic and associative interactions with hydrophilic and hydrophobic cellulose surfaces.

3.1.2 Structural characteristics of cellulose-lignin complexation in solvents.

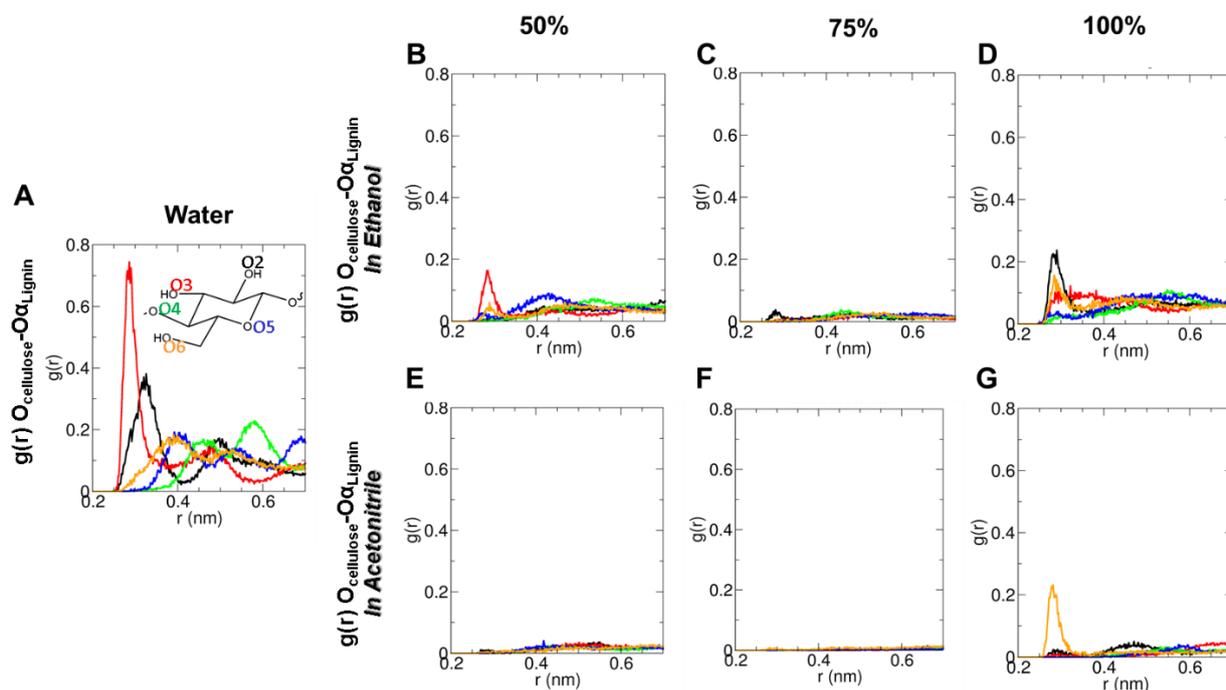


Figure 5.4. Cellulose-lignin radial distribution functions, $g(r)$, of the Ocellulose-O α Lignin pairs. Ocellulose colour code is O2 (black), O3 (red), O4 (green), O5 (blue), and O6 (orange), see Figure 5.2A for cellulose oxygen notation.

To disentangle the origin of the lignin-cellulose coordination in the pure and mixed solvents, we computed the site-to-site radial distribution functions between the cellulose oxygen sites (O2, O3, O4, O5, and O6, see Figure 5.2A for notation) and the lignin oxygens (O α , O γ , and OMe, see Figure 5.2B for notation). The O_{cell}-O α _{lig} $g(r)$ s in Figure 5.7, and the O_{cell}-O γ _{lig} and O_{cell}-OMe_{lig} $g(r)$ s in Supplementary Information (Figure S5.1) show the highest probability of coordination for the O3_{cell}-O α _{lig} and O2_{cell}-O γ _{lig} in pure water. In ethanol and ethanol mixtures (Figures 5.7B-D), the O3_{cell}-O α _{lig} coordination is significantly disrupted, but in pure ethanol, a residual coordination towards O2 is observed. Similarly, water-acetonitrile mixtures (Figures 5.7E-F) prove to be more effective at disrupting the lignin-cellulose coordination, and, as in ethanol, a residual coordination towards the O6 remains in pure acetonitrile (Figure 5.8G). Thus, water-organic solvents mixtures act together to disrupt more effectively the cellulose-lignin interactions not only by preferential interactions with different cellulose surfaces but also by selective coordination of specific cellulose and lignin sites. Organic solvents disrupt both stacking and hydrogen bonding, but their combination with water seems to enhance the cleavage of the remaining cellulose-lignin H-bonds. This suggests that preferential interactions between cellulose and lignin with water-organic solvent molecules influences the lignin-cellulose adhesion.

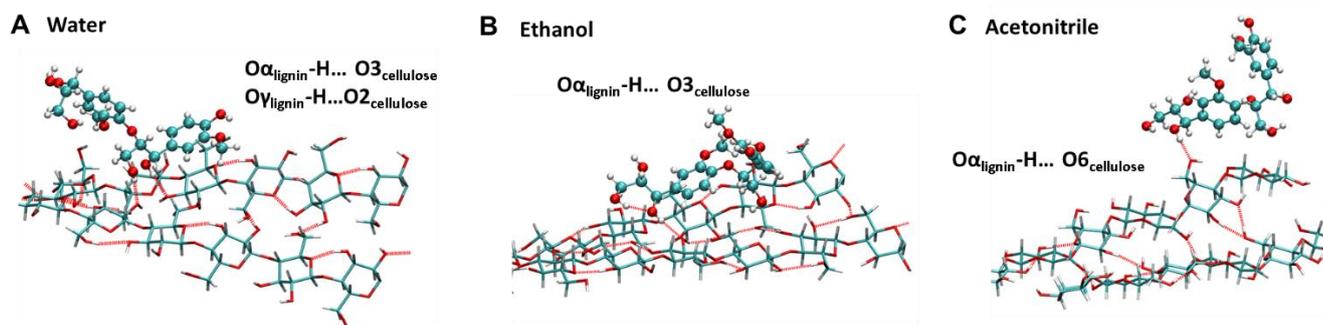


Figure 5.5. MD snapshots of most stable cellulose-lignin complexes to illustrate the binding of lignin to cellulose in water (A), ethanol (B), and acetonitrile (C) solvents (solvents not shown). Stacking between cellulose glucose and lignin phenol rings, and hydrogen bonding between cellulose-lignin alcohol groups is observed. H-bonds are shown with red dashed lines.

Furthermore, we have extracted the most stable cellulose-lignin complexes from the trajectories, as displayed in Figure 5.8, and the detailed example of the lignin-cellulose binding of the most strongly interacting dimer, accordingly to Figure 5.6, Table 5.3 and Figure 5.7, is shown in pure water, ethanol, and acetonitrile solvents. As illustrated in Figure 5.8A, aromatic rings of lignin adopt a preferential parallel orientation relative to the cellulose surface in water, as already reported previously [33, 34], and electrostatic interactions between the alcohol groups (i.e. $O\alpha H$ and $O\gamma H$) of lignin and the hydroxyls of cellulose are favoured. Lignin acts as an H-bond donor [33], and the $O\alpha_{\text{lig}}-H\dots O3_{\text{cell}}$ and $O\gamma_{\text{lig}}-H\dots O2_{\text{cell}}$ H-bonds are formed. In ethanol (Figure 5.8B), the H-bond between $O\alpha$ of the lignin vinyl group and $O3$ of cellulose prevails. Regardless of the H-bond formed between lignin vinyl $O\alpha$ and the glucose $O6$ in cellulose, significantly less stacking between the lignin aromatic ring and the cellulose glucose is observed in the presence of acetonitrile. In mixed solvents, moreover, the cellulose-lignin complex remains mostly unbounded.

3.2 Solvent effect on the adhesion between xylan and lignin

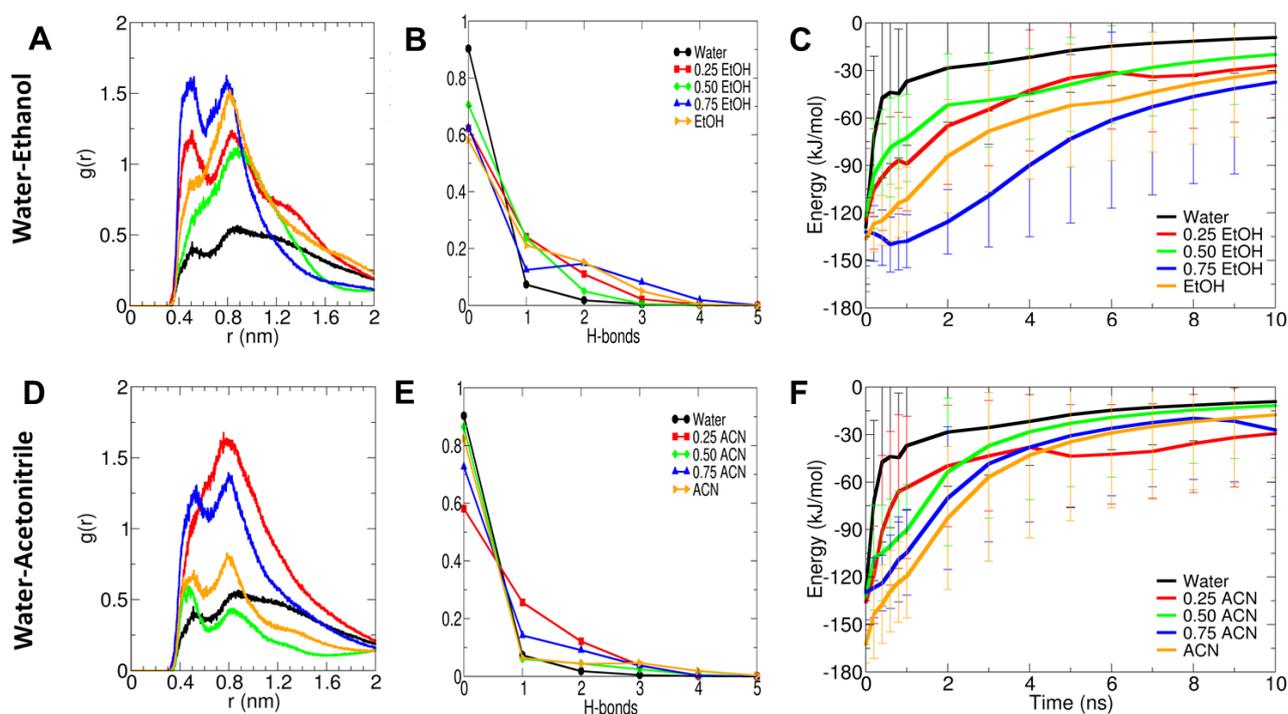


Figure 5.6. (A) xylose-lignin radial distribution function, $g(r)$, between the surface xylosyl monomers and the lignin phenolic rings, in pure and aqueous ethanol mixtures. Only the carbons of xylose and the aromatic ring, respectively, have been used in this calculation. (B) Probability distribution of hydrogen bonds between xylan and lignin tetramers, in pure and aqueous organic solvent mixtures. H-bond intervals are discrete, but lines have been added to describe better the trends. (C) Average xylan-lignin intermolecular energies per lignin monomer, in pure and binary ethanol mixtures. The radial distribution function, $g(r)$, the probability distribution of H-bonds and the average xylan-lignin intermolecular energies in acetonitrile and water-acetonitrile mixtures are reported in (D), (E) and (F), respectively.

We evaluated the dynamics of the xylose tetramer initially in complex with a guaiacyl tetramer (Figure 5.1D) in pure and binary mixtures of water-ethanol and water-acetonitrile. The interaction between the xylan and the lignin tetramers was estimated from the computed xylose-lignin radial distribution functions (i.e. $g(r)$ s), in Figure 5.9A, D), the xylose-lignin H-bond distributions (Figure 5.9B, E), and the average interaction energies during the dynamics (Figures 5.9C, F). The xylan-lignin $g(r)$ s show that it is less likely to find lignin coordinating xylan in pure water, as the hemicellulose becomes rapidly hydrated. The addition of the organic solvent increases the xylose-lignin coordination, and the effect becomes more notorious in the presence of 75% ethanol and 25% and 75% acetonitrile solutions. Very recent NMR spectroscopy and dynamic nuclear polarisation experiments have revealed that lignin has abundant electrostatic interactions with the polar motifs of xylan [19]. Accordingly, the xylose-lignin H-bond distributions (Figure 5.9B, E) show the formation of H-bonds between the xylan and lignin in pure and mixed solvents. Moreover, it indicates that lignin remains unbounded (H-bond =0) to xylan in water during ~90% of the simulation time, followed by pure and 50% acetonitrile, in which ~80-85% of the time the unbounded state is observed. On the contrary, the bounded

state ($H\text{-bond} \geq 1$) is more favoured in pure ethanol, ethanol mixtures, and 25% acetonitrile solution (~40% of the simulation time). It is worth mentioning that we expect the polymer interactions in our simulations to be rapidly affected by the solvent since the simulated systems contain excess solvent.

As follows from the xylose-lignin interaction energies (Figure 5.9 C, F and Table S5-1), the xylose-lignin interactions energies show a non-linear behaviour relative to the organic solvent composition, in agreement with our observations of the H-bond distributions and the xylose-lignin radial distribution functions. In water, a bad solvent for lignin but with high affinity for hemicellulose, the xylose-lignin complex displays the weakest binding energy during all the simulation. In the presence of ethanol, we observe that the strongest adhesion between xylose and lignin is found in 75% ethanol mixture and pure ethanol during all the simulation, while the weakest adhesion is found in 25 % and 50% ethanol-water mixtures. Overall, acetonitrile and acetonitrile mixtures are more effective at disrupting xylan-lignin interactions than ethanol and ethanol mixtures. Nonetheless, we observe an inversion of the xylan-lignin adhesion above 2-ns simulation time in the presence of acetonitrile. The xylan-lignin adhesion below 2-ns MD increases linearly with the acetonitrile concentration; however, it follows the order 25% > 75 % > pure > 50% at the end of the 10 ns dynamics, where we find that the statistical difference between the interactions energies in the presence of the solvent is not significant (Table S5-1). We further followed the evolution of the Lennard-Jones (LJ) and Coulomb (Coul) interactions to the xylan-lignin interaction energies (Figure 5.4 and Table S5-1), and we observe that the relative intensity of the van der Waals interactions (LJ) is correlated to the intensity of the electrostatic Coulomb interactions. Moreover, we find a trend for stronger Coulomb energies than LJ interactions between xylan and lignin (Figure 5.10).

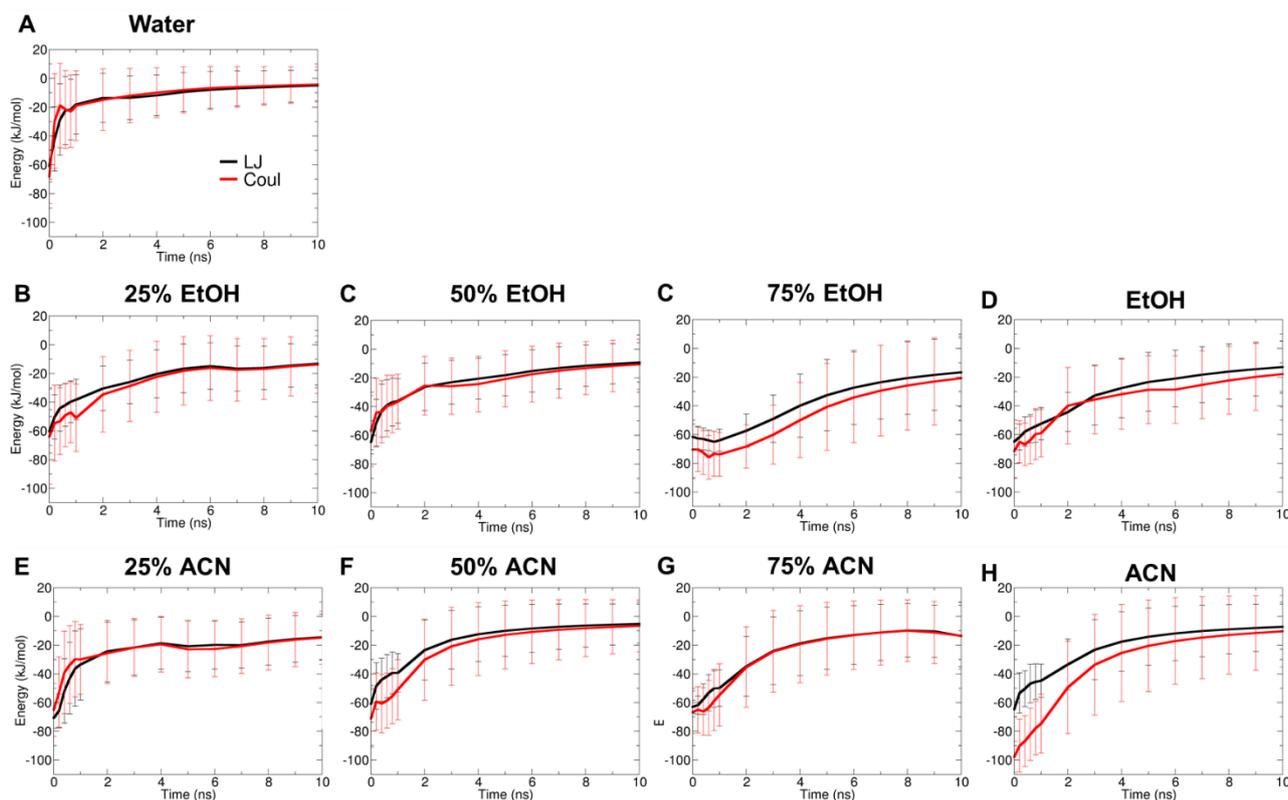


Figure 5.7. Evolution of the Lennard-Jones (LJ, black) and Coulomb (Coul, red) average lignin-xytan energies along the 10-ns MD normalised by the number of lignin dimers, in pure and mixed water-organic solvent mixtures.

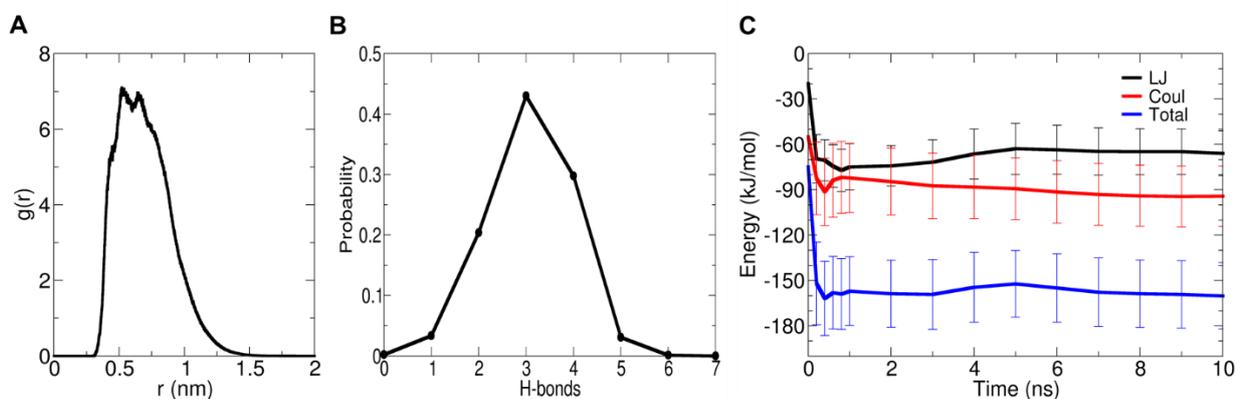


Figure 5.8. Trajectory analysis of xylan-lignin tetramers in the gas phase. (A) xylose-lignin radial distribution function, $g(r)$, between the surface xylosyl monomers and the lignin phenolic rings. Only the carbons of xylose and the aromatic ring, respectively, have been used in this calculation. (B) Probability distribution of hydrogen bonds between xylan and lignin tetramers. H-bond intervals are discrete, but lines have been added to describe better the trends. (C) Average xylan-lignin intermolecular energies.

To understand better the origin of the detachment between the xylan-cellulose complex, we analysed the trajectories in gas phase (Figure 5.11) and the presence of the solvents, and we extracted the xylose-lignin complexes at different simulation times in water, 75% ethanol, and 25% acetonitrile, illustrated in Figure 5.11. First, the initial structure of the xylose-lignin, obtained from the equilibration in gas phase (no solvent) of the

polymers (Fig. 5.1D) shows a folded lignin tetramer wrapped by a bent xylan tetramer. The initial SASA of the xylan and lignin tetramers in this complex are 7.7 and 9.2 nm², respectively. Furthermore, xylan adopts a conformation in which three out of the four xylose monomers are in the α -form (glycosidic torsion angle $\sim 64^\circ$) with the hydroxyl groups on the same side. In lignin, two pairs of rings are stacked with torsion angles equal to -65° and -85° , respectively, whereas the third pair adopts a T-stacked conformation with a torsion angle of -139° . Moreover, the xylan-lignin complex interact both through H-bonding and stacking. The coordination between xylan and lignin rings (Figure 5.11A) is almost 5-fold higher when compared to the xylan-lignin coordination in 75% ethanol (Figure 5.9A). The H-bond distribution falls at 3 H-bonds; therefore, the hydrogen bond interactions are maximised in the absence of the solvent and strongly disrupted upon the solvent addition (Figure 5.9B). Furthermore, the average interaction energies are also almost 4- to 13-times stronger compared to 75% ethanol and water, respectively, (Table S5-1) and are dominated by Coulomb interactions as a result of the higher frequency of H-bonds, as illustrated in Figure 5.9 C. Moreover, the cellulose-lignin LJ interactions without solvent are on average up to 4 to 13 times stronger than in the solvents, but the Coulomb interactions increase even more - up to 4 to 22 times. Thus, the addition of solvent has the strongest effect on the disruption of the electrostatic interactions by disrupting the hydrogen bonding between lignin and cellulose, similarly as found above for cellulose-lignin.

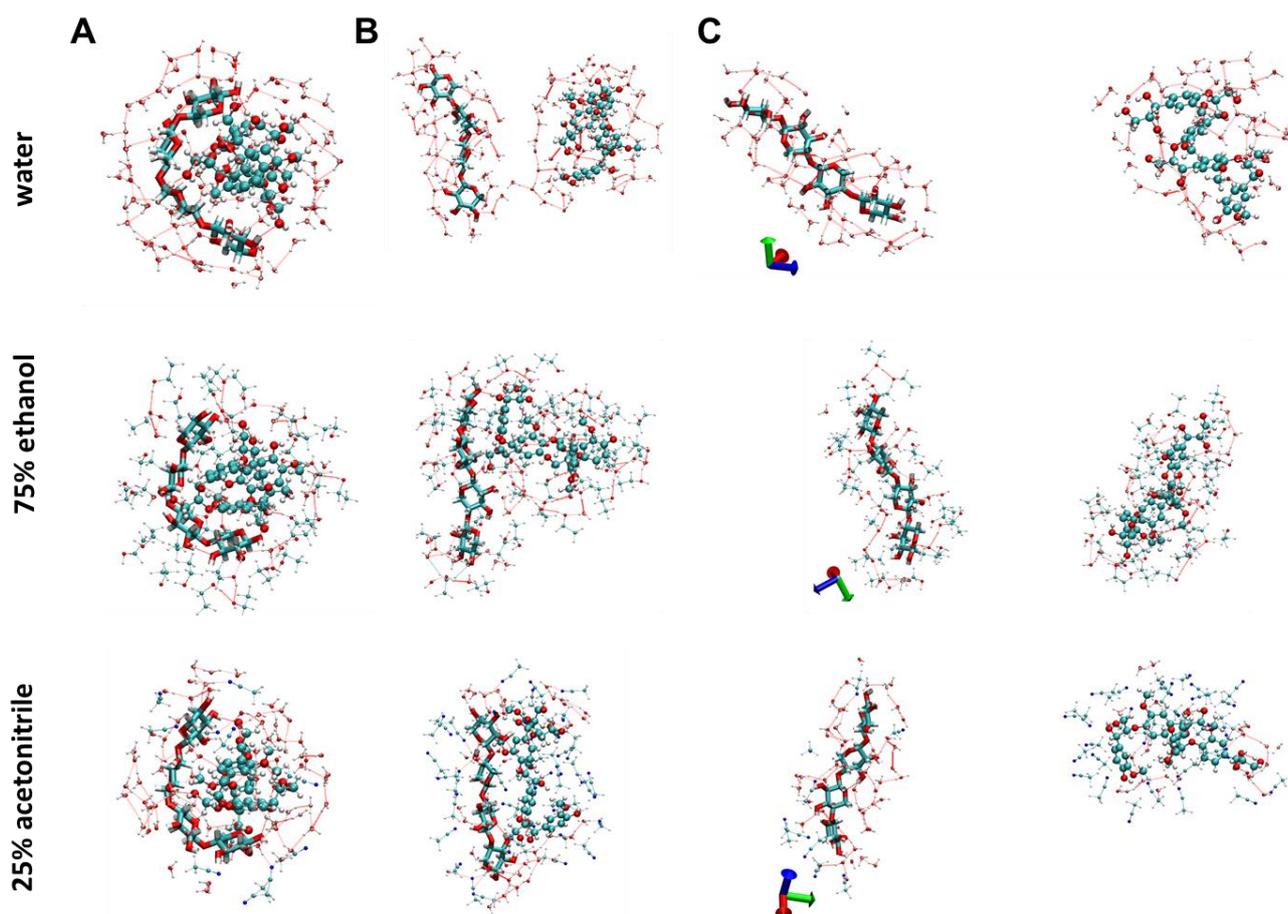


Figure 5.9. MD snapshots of xylan-lignin complexes to illustrate A) the initial xylan-lignin cluster and equilibrated 3.5 nm shell solvent, B) xylan-lignin cluster within 2-ns simulation time and C) detachment between xylose and lignin at the end of the 10-ns dynamics in water, 75% ethanol, and 25% acetonitrile. The lignin tetramer is represented by CPK and the xylose tetramer by Bonds. H-bonds are shown with red dashed lines. Atom colour code: cyan=carbon, white=hydrogen, red=oxygen, and blue=nitrogen.

Furthermore, we also analysed the trajectories in the presence of the solvents to understand the factors responsible for the xylan-lignin detachment in our simulations. We extracted the xylose-lignin complexes at different simulation times in water, 75% ethanol, and 25% acetonitrile, illustrated in Figure 5.12. Figure 5.12A shows the solvent-equilibrated xylan-complex along with a 3.5 nm solvent shell at $t=0$ ns. The structures of the xylan-lignin complex within the first 2-ns of simulation (Figure 5.12 B) show that xylan rapidly undergoes a conformation change. In water, a hydrated and detached xylan molecule from lignin is observed, whereas in 75 % ethanol and 25% acetonitrile the xylan-lignin complex remains. Finally, in Figure 5.12 C, the xylan detaches from lignin in all solvent mixtures at the end of the 10-ns MD, accordingly with the interactions energies shown in Figure 5.9 C, F.

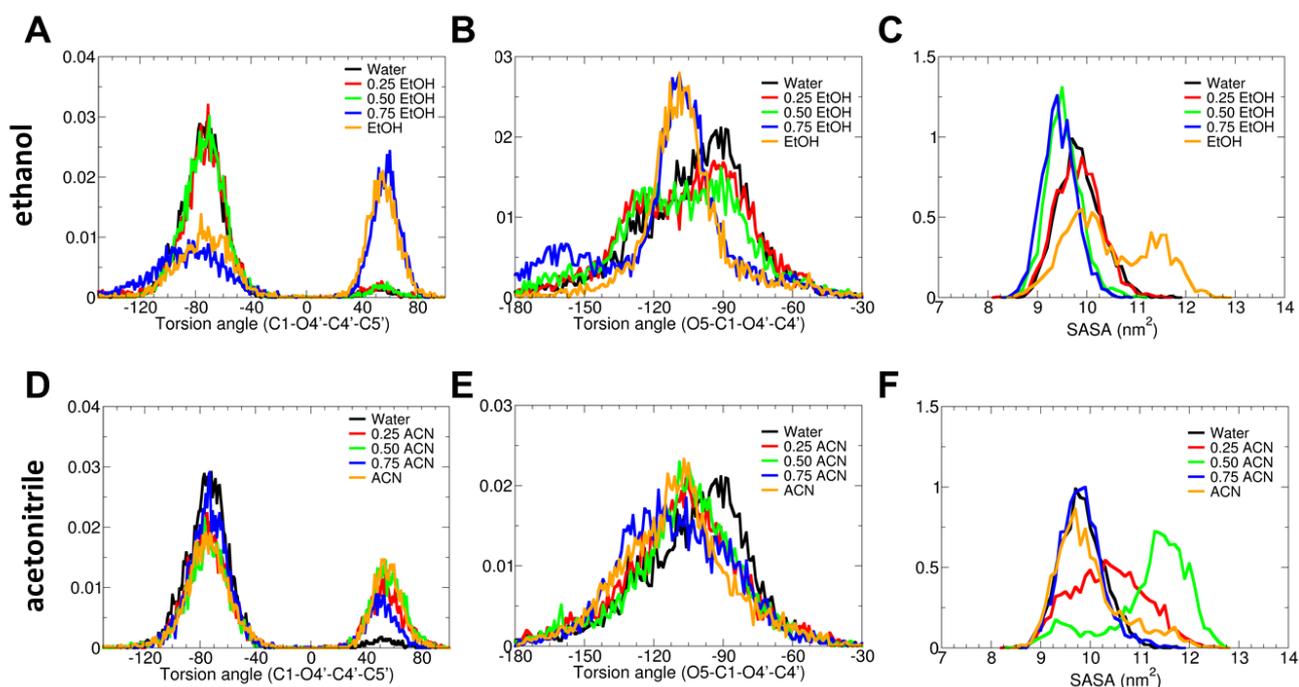


Figure 5.10. Average distributions of glycosidic torsional angles (A) θ (O5-C1-O4'-C4') and (B) Ψ (C1-O4'-C4'-C5') and (C) Average distributions of solvent accessible surface area (SASA) of lignin tetramers at different ethanol concentrations. (D-F) idem as (A-C) in water-acetonitrile mixtures.

We analysed the glycosidic torsional angles of xylan and the SASA of lignin within the first 2-ns MD to understand the origin of the detachment, and the results are summarised in Figure 5.13. The distribution of the torsion angle θ (O5-C1-O4'-C4') during the first 2-ns (Figure 5.13A, D) falls mostly at $\sim -74^\circ$ recovering the β -(1 \rightarrow 4) bond form. This conformational change occurs almost spontaneously in the presence of water, whereas the somewhat higher frequency of α -(1 \rightarrow 4) bonds is an indication that this conformational change is slowed down in the presence of the organic solvents and the mixtures, particularly at pure and 75% ethanol, in which the adhesion is stronger. Thus, the first factor responsible for the detachment of the xylan-lignin tetramer is the rearrangement of the xylan conformation upon solvation, without the restraints imposed during the equilibration. Furthermore, the distribution of the torsion angle Ψ (C1-O4'-C4'-C5', Figure 5.13B, E) in water falls near $\sim 90^\circ$, whereas the distributions in pure and 75% ethanol are narrower and fall near $\sim 110^\circ$, possibly describing the twofold conformation, as we reported in Chapter 4. Falcoz-Vigne et al. [22] have observed that the interaction between xylan and cellulose is stronger when xylan is in the twofold conformation than when it is in the threefold one. Our results have shown that xylose adopts the twofold conformation in the presence of the organic solvent. Thus, our results suggest that the interaction between xylan and lignin is also enhanced in the presence of the organic solvent arising from a hydrophobic interaction, which results in the stabilisation of lignin in the presence of xylan. Our results are in qualitative agreement with the experimental and theoretical observations of Smith et al. [57] who showed that the presence of tetrahydrofuran (THF) in THF-water mixtures delays the xylan's solubility when compared to the solubility in water.

Moreover, the SASA distribution of lignin during the first 2-ns show a rather compact structure, that falls near $\sim 10 \text{ nm}^2$ in water, acetonitrile, 75% acetonitrile, and 25% ethanol, but SASA falls below ~ 9.2 in 50 and 75% ethanol. The SASA distributions in pure ethanol and 25 and 50% acetonitrile are broader and show more extended conformations. Thus, another factor responsible for the detachment of the xylan-lignin tetramers is the conformational change of lignin upon solvation, particularly at 50% acetonitrile. Furthermore, as we discussed in Chapter 4, the decrease of the xylan mobility upon addition of ethanol, and the gel-like behaviour could enhance the lignin adhesion to xylose, by slowing down also conformational changes of lignin upon solvation. Moreover, our results in chapter 4 suggest that acetonitrile enhance xylan mobility when compared to ethanol. Thus, this can contribute to the overall reduced adhesion between xylan and lignin in acetonitrile than in ethanol.

4. Conclusions

Molecular dynamics simulations were employed to describe at the atomistic level the effect of the solvents on the adhesion of lignin to cellulose and hemicellulose. The considered solvents are first pure water, ethanol and acetonitrile, and second, the binary mixtures of water with the two organic solvents. We evaluated the cellulose-lignin and xylose-lignin interactions in the absence of the solvent as well. The polymer-polymer adhesion was investigated through radial distribution functions, hydrogen bonding, and intermolecular interaction energies at the classical level.

We followed the adhesion behaviour between xylose and lignin tetramers and the radial distribution analysis show that it is less likely to find lignin coordinating xylan in pure water, as the hemicellulose becomes rapidly hydrated. The addition of the organic phase, however, increases the xylose-lignin coordination and the effect becomes more notorious in the presence of the organic solvent. The xylose-lignin interactions energies show a non-linear behaviour to the organic solvent composition, in agreement with our observations of the H-bond distributions and the xylose-lignin radial distribution functions. In water, a bad solvent for lignin but with a high affinity for hemicellulose, the xylose-lignin complex displays the weakest binding energy, followed by 50% acetonitrile. The strongest adhesion between xylose and lignin is observed at 75% ethanol mixture, followed by 25% and 75% acetonitrile mixture, which we have reported to be good solvents for lignin as the SASA distribution of lignin tetramers reaches a maximum.

Furthermore, at these concentrations, the presence of the organic phase seems to slow down the xylose conformational changes and, thus, the detachment to lignin. Concerning the adsorption behaviour of lignin dimer on the cellulose hydrophobic and hydrophilic surfaces, we found that lignin could be stabilised in the presence of water by stacking interactions between the phenolic groups and the saccharide groups. Furthermore, lignin dimers were found to preferentially interact with the hydrophilic chains of cellulose, whose less polar aliphatic hydrogen atoms of the glucose rings are exposed. The computed radial distribution functions $g(r)$, and average number of hydrogen bonds were analysed concerning cellulose and lignin

conformational changes and cellulose-lignin interaction energies in the mono-component and 50, and 75wt% water-organic solvents. Water is the medium that favours mostly the adhesion of lignin on cellulose, followed by pure ethanol and acetonitrile solvents.

In the mixed solvents, the cellulose-lignin adhesion is strongly disrupted. The specific organisation of the solvent molecules in the binary mixtures around the cellulose-lignin complex is identified as the most crucial factor affecting the cellulose-lignin conformations and their interaction energies. Moreover, the preferential interactions of the two organic components in the presence of water depending on their chemical nature and concentration. The weakest (or no adhesion) of lignin on cellulose is found in the 50% acetonitrile-water solvent, and at 75% ethanol-water solvent. Despite the relatively small cellulose-lignin model, our MD results are in a qualitative agreement with recent studies that show an increase of lignin solubility in water-ethanol mixtures[73].

References

1. Lambert, E., et al., *Real Time and Quantitative Imaging of Lignocellulosic Films Hydrolysis by Atomic Force Microscopy Reveals Lignin Recalcitrance at Nanoscale*. *Biomacromolecules*, 2019. **20**(1): p. 515-527.
2. Li, M.-F., S. Yang, and R.-C. Sun, *Recent advances in alcohol and organic acid fractionation of lignocellulosic biomass*. *Bioresource Technology*, 2016. **200**: p. 971-980.
3. Lali, A.M., J.S. Varavadekar, and P.C. Wadekar, *Process for fractionation of biomass* 2011.
4. Grande, P.M., et al., *Fractionation of lignocellulosic biomass using the OrganoCat process*. *Green Chemistry*, 2015. **17**(6): p. 3533-3539.
5. Carneiro, A.P., O. Rodríguez, and E.A. Macedo, *Dissolution and fractionation of nut shells in ionic liquids*. *Bioresource Technology*, 2017. **227**: p. 188-196.
6. Yoo, C.G., *Pretreatment and fractionation of lignocellulosic biomass for production of biofuel and value-added products*. 2012, Iowa State University.
7. Mussatto, S.I. and G.M. Dragone, *Chapter 1 - Biomass Pretreatment, Biorefineries, and Potential Products for a Bioeconomy Development*, in *Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery*. 2016, Elsevier: Amsterdam. p. 1-22.
8. Romero, I., E. Ruiz, and E. Castro, *Chapter 10 - Pretreatment With Metal Salts A2 - Mussatto, Solange I*, in *Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery*. 2016, Elsevier: Amsterdam. p. 209-227.
9. Santos, J.C., et al., *Chapter 9 - Biomass Pretreatment With Oxalic Acid for Value-Added Products A2 - Mussatto, Solange I*, in *Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery*. 2016, Elsevier: Amsterdam. p. 187-208.
10. Shuddhodana, et al., *Chapter 23 - Enzymatic Hydrolysis of Lignocellulosic Residues A2 - Mussatto, Solange I*, in *Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery*. 2016, Elsevier: Amsterdam. p. 543-560.

11. Wikandari, R., R. Millati, and M.J. Taherzadeh, *Chapter 12 - Pretreatment of Lignocelluloses With Solvent N-Methylmorpholine N-oxide A2 - Mussatto, Solange I*, in *Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery*. 2016, Elsevier: Amsterdam. p. 255-280.
12. Xu, J.K. and R.C. Sun, *Chapter 19 - Recent Advances in Alkaline Pretreatment of Lignocellulosic Biomass A2 - Mussatto, Solange I*, in *Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery*. 2016, Elsevier: Amsterdam. p. 431-459.
13. Zhang, C., et al., *Chapter 22 - Sulfite Pretreatment to Overcome the Recalcitrance of Lignocelluloses for Bioconversion of Woody Biomass A2 - Mussatto, Solange I*, in *Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery*. 2016, Elsevier: Amsterdam. p. 499-541.
14. Domínguez-Robles, J., et al., *Aqueous acetone fractionation of kraft, organosolv and soda lignins*. *International Journal of Biological Macromolecules*, 2018. **106**: p. 979-987.
15. Yang, Y. and F. Zhang, *Ultrasound-assisted extraction of rutin and quercetin from *Euonymus alatus* (Thunb.) Sieb.* *Ultrasonics Sonochemistry*, 2008. **15**(4): p. 308-313.
16. Aguilera-Segura, S.M., et al., *Synergistic sorption of mixed solvents in wood cell walls: experimental and theoretical approach*. *Macromolecular symposia*, 2019. **386**(1): p. 1900022.
17. Wierzbicki, M.P., et al., *Xylan in the Middle: Understanding Xylan Biosynthesis and Its Metabolic Dependencies Toward Improving Wood Fiber for Industrial Processing*. *Frontiers in Plant Science*, 2019. **10**(176).
18. Petridis, L. and J.C. Smith, *Molecular-level driving forces in lignocellulosic biomass deconstruction for bioenergy*. *Nature Reviews Chemistry*, 2018. **2**(11): p. 382-389.
19. Kang, X., et al., *Lignin-polysaccharide interactions in plant secondary cell walls revealed by solid-state NMR*. *Nature Communications*, 2019. **10**(1): p. 347.
20. Zhang, Y., et al., *Recent progress in theoretical and computational studies on the utilization of lignocellulosic materials*. *Green Chemistry*, 2019. **21**(1): p. 9-35.
21. Donev, E., et al., *Engineering Non-cellulosic Polysaccharides of Wood for the Biorefinery*. *Frontiers in Plant Science*, 2018. **9**(1537).
22. Falcoz-Vigne, L., et al., *Quantification of a tightly adsorbed monolayer of xylan on cellulose surface*. *Cellulose*, 2017. **24**(9): p. 3725-3739.
23. Kumar, R., et al., *Cellulose–hemicellulose interactions at elevated temperatures increase cellulose recalcitrance to biological conversion*. *Green Chemistry*, 2018. **20**(4): p. 921-934.
24. Mazeau, K. and L. Charlier, *The molecular basis of the adsorption of xylans on cellulose surface*. *Cellulose*, 2012. **19**(2): p. 337-349.
25. Yang, H., et al., *Quantum Calculations on Plant Cell Wall Component Interactions*. *Interdisciplinary Sciences: Computational Life Sciences*, 2018.
26. Almond, A. and J.K. Sheehan, *Predicting the molecular shape of polysaccharides from dynamic interactions with water*. *Glycobiology*, 2003. **13**(4): p. 255-264.

27. Houtman, C.J. and R.H. Atalla, *Cellulose-Lignin Interactions (A Computational Study)*. Plant Physiology, 1995. **107**(3): p. 977-984.
28. Youssefian, S. and N. Rahbar, *Molecular Origin of Strength and Stiffness in Bamboo Fibrils*. Scientific Reports, 2015. **5**: p. 11116.
29. Silveira, R.L., et al., *Plant Biomass Recalcitrance: Effect of Hemicellulose Composition on Nanoscale Forces that Control Cell Wall Strength*. Journal of the American Chemical Society, 2013. **135**(51): p. 19048-19051.
30. Durbeej, B. and L.A. Eriksson, *A Density Functional Theory Study of Coniferyl Alcohol Intermonomeric Cross Linkages in Lignin - Three-Dimensional Structures, Stabilities and the Thermodynamic Control Hypothesis*, in *Holzforschung*. 2003. p. 150.
31. Durbeej, B., Y.-N. Wang, and L.A. Eriksson, *Lignin Biosynthesis and Degradation — a Major Challenge for Computational Chemistry*, in *High Performance Computing for Computational Science — VECPAR 2002: 5th International Conference Porto, Portugal, June 26–28, 2002 Selected Papers and Invited Talks*, J.M.L.M. Palma, et al., Editors. 2003, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 137-165.
32. Besombes, S. and K. Mazeau, *Molecular dynamics simulations of a guaiacyl β -O-4 lignin model compound: Examination of intramolecular hydrogen bonding and conformational flexibility*. Biopolymers, 2004. **73**(3): p. 301-315.
33. Besombes, S. and K. Mazeau, *The cellulose/lignin assembly assessed by molecular modeling. Part 1: adsorption of a threo guaiacyl beta-O-4 dimer onto a Ibeta cellulose whisker*. Plant physiology and biochemistry : PPB, 2005. **43**(3): p. 299-308.
34. Besombes, S. and K. Mazeau, *The cellulose/lignin assembly assessed by molecular modeling. Part 2: seeking for evidence of organization of lignin molecules at the interface with cellulose*. Plant Physiology and Biochemistry, 2005. **43**(3): p. 277-286.
35. Charlier, L. and K. Mazeau, *Molecular Modeling of the Structural and Dynamical Properties of Secondary Plant Cell Walls: Influence of Lignin Chemistry*. The Journal of Physical Chemistry B, 2012. **116**(14): p. 4163-4174.
36. Vermaas, J.V., et al., *Automated Transformation of Lignin Topologies into Atomic Structures with LigninBuilder*. ACS Sustainable Chemistry & Engineering, 2019. **7**(3): p. 3443-3453.
37. Vermaas, J.V., et al., *Systematic parameterization of lignin for the CHARMM force field*. Green Chemistry, 2019. **21**(1): p. 109-122.
38. Vermaas, J.V., et al., *Mechanism of lignin inhibition of enzymatic biomass deconstruction*. Biotechnology for Biofuels, 2015. **8**(1): p. 217.
39. Langan, P., et al., *Common processes drive the thermochemical pretreatment of lignocellulosic biomass*. Green Chemistry, 2014. **16**(1): p. 63-68.
40. Sangha, A.K., et al., *Molecular simulation as a tool for studying lignin*. Environmental Progress & Sustainable Energy, 2012. **31**(1): p. 47-54.
41. Schulz, R., et al., *Scaling of Multimillion-Atom Biological Molecular Dynamics Simulation on a Petascale Supercomputer*. Journal of Chemical Theory and Computation, 2009. **5**(10): p. 2798-2808.

42. Lindner, B., et al., *Solvent-Driven Preferential Association of Lignin with Regions of Crystalline Cellulose in Molecular Dynamics Simulation*. *Biomacromolecules*, 2013. **14**(10): p. 3390-3398.
43. Woods group. Complex Carbohydrate Research Center, U.o.G., Athens, GA. (<http://glycam.org>). *GLYCAM Web*. 2005-2019.
44. Pingali, S.V., et al., *Morphological changes in the cellulose and lignin components of biomass occur at different stages during steam pretreatment*. *Cellulose*, 2014. **21**(2): p. 873-878.
45. Chandra, R.P., V. Arantes, and J. Saddler, *Steam pretreatment of agricultural residues facilitates hemicellulose recovery while enhancing enzyme accessibility to cellulose*. *Bioresource Technology*, 2015. **185**: p. 302-307.
46. Leeﬂang, B.R., et al., *A 1H-NMR and MD study of intramolecular hydrogen bonds in methyl β -cellobioside*. *Carbohydrate Research*, 1992. **230**(1): p. 41-61.
47. Sawada, D., et al., *Structure and dynamics of a complex of cellulose with EDA: insights into the action of amines on cellulose*. *Cellulose*, 2013. **20**(4): p. 1563-1571.
48. Smith, M.D., et al., *Cosolvent pretreatment in cellulosic biofuel production: effect of tetrahydrofuran-water on lignin structure and dynamics*. *Green Chemistry*, 2016. **18**(5): p. 1268-1277.
49. Smith, M.D., et al., *Enhanced sampling simulation analysis of the structure of lignin in the THF-water miscibility gap*. *Physical Chemistry Chemical Physics*, 2016. **18**(9): p. 6394-6398.
50. Vasudevan, V. and S.H. Mushrif, *Insights into the solvation of glucose in water, dimethyl sulfoxide (DMSO), tetrahydrofuran (THF) and N,N-dimethylformamide (DMF) and its possible implications on the conversion of glucose to platform chemicals*. *RSC Advances*, 2015. **5**(27): p. 20756-20763.
51. Mostofian, B., et al., *Local Phase Separation of Co-solvents Enhances Pretreatment of Biomass for Bioenergy Applications*. *Journal of the American Chemical Society*, 2016. **138**(34): p. 10869-10878.
52. Smith, M.D., et al., *Organosolv-Water Cosolvent Phase Separation on Cellulose and its Influence on the Physical Deconstruction of Cellulose: A Molecular Dynamics Analysis*. *Scientific Reports*, 2017. **7**(1): p. 14494.
53. Umemura, M. and Y. Yuguchi, *Solvation of xyloglucan in water/alcohol systems by molecular dynamics simulation*. *Cellulose*, 2009. **16**(3): p. 361-371.
54. Yuguchi, Y., et al., *Gelation of xyloglucan in water/alcohol systems*. *Cellulose*, 2004. **11**(2): p. 203-208.
55. Yamanaka, S., et al., *Gelation of tamarind seed polysaccharide xyloglucan in the presence of ethanol*. *Food Hydrocolloids*, 2000. **14**(2): p. 125-128.
56. Chen, P., et al., *Translational Entropy and Dispersion Energy Jointly Drive the Adsorption of Urea to Cellulose*. *The Journal of Physical Chemistry B*, 2017. **121**(10): p. 2244-2251.
57. Smith, M.D., et al., *Temperature-dependent phase behaviour of tetrahydrofuran–water alters solubilization of xylan to improve co-production of furfurals from lignocellulosic biomass*. *Green Chemistry*, 2018. **20**(7): p. 1612-1620.
58. Nishiyama, Y., P. Langan, and H. Chanzy, *Crystal Structure and Hydrogen-Bonding System in Cellulose I β from Synchrotron X-ray and Neutron Fiber Diffraction*. *Journal of the American Chemical Society*, 2002. **124**(31): p. 9074-9082.

59. Gomes, T.C.F. and M.S. Skaf, *Cellulose-Builder: A toolkit for building crystalline structures of cellulose*. Journal of Computational Chemistry, 2012. **33**(14): p. 1338-1346.
60. Pettersen, R.C., *The Chemical Composition of Wood*, in *The Chemistry of Solid Wood*. 1984, American Chemical Society. p. 57-126.
61. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers*. SoftwareX, 2015. **1–2**: p. 19-25.
62. Bekker, H., et al. *Gromacs: A parallel computer for molecular dynamics simulations*. in *Physics computing*. 1993.
63. Berendsen, H.J.C., D. van der Spoel, and R. van Drunen, *GROMACS: A message-passing parallel molecular dynamics implementation*. Computer Physics Communications, 1995. **91**(1): p. 43-56.
64. Van Der Spoel, D., et al., *GROMACS: Fast, flexible, and free*. Journal of Computational Chemistry, 2005. **26**(16): p. 1701-1718.
65. Páll, S., et al., *Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS*, in *Solving Software Challenges for Exascale: International Conference on Exascale Applications and Software, EASC 2014, Stockholm, Sweden, April 2-3, 2014, Revised Selected Papers*, S. Markidis and E. Laure, Editors. 2015, Springer International Publishing: Cham. p. 3-27.
66. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water*. The Journal of Chemical Physics, 1983. **79**(2): p. 926-935.
67. Dick, T.J. and J.D. Madura, *Chapter 5 A Review of the TIP4P, TIP4P-Ew, TIP5P, and TIP5P-E Water Models*, in *Annual Reports in Computational Chemistry*. 2005, Elsevier. p. 59-74.
68. Mark, P. and L. Nilsson, *Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K*. The Journal of Physical Chemistry A, 2001. **105**(43): p. 9954-9960.
69. Guvench, O., et al., *Additive empirical force field for hexopyranose monosaccharides*. J Comput Chem, 2008. **29**.
70. Guvench, O., et al., *CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses*. J Chem Theory Comput, 2009. **5**.
71. Petridis, L. and J.C. Smith, *A molecular mechanics force field for lignin*. Journal of Computational Chemistry, 2009. **30**(3): p. 457-467.
72. van der Spoel, D., P.J. van Maaren, and C. Caleman, *GROMACS molecule & liquid database*. Bioinformatics, 2012. **28**(5): p. 752-753.
73. Goldmann, W.M., et al., *Solubility and fractionation of Indulin AT kraft lignin in ethanol-water media*. Separation and Purification Technology, 2019. **209**: p. 826-832.

Chapter 6. Discussion: the solvent effect on lignocellulose interactions

Contents

1. Cellulose-lignin adhesion: correlation between interaction energies and the solvent's preferential interaction	202
2. Cellulose-solvent interactions: the water exclusion and the entropic effect	204
3. Solvent effects on molecular wood structures: possible implications in lignocellulose adhesion and wood swelling	205
4. Assessment of the achieved scope of this work, the reliability of model size and simulation time scales, and future perspectives.	207
References	210

Abstract

In this chapter, we discuss the possible implications of the results reported in Chapters 3-5. We discuss (i) the correlation between the interaction energies and the preferential interactions of the solvents; (ii) the driving forces controlling the solvent-cellulose interactions; (iii) the solvent effects on molecular wood structures and their possible implication in lignocellulose interactions; and (iv) the scope of this work and limitations of the chosen models.

1. Cellulose-lignin adhesion: correlation between interaction energies and the solvent's preferential interaction

The cellulose-lignin interaction energies and their evolution along dynamics simulations, presented in Chapter 5 (Figs. 5.4 and 5.5,) demonstrate that the introduction of the solvent results in a substantial disruption of cellulose-lignin interactions, particularly the Coulomb electrostatic ones, predominantly associated with H-bonds interactions. Moreover, the energy profiles indicate different solvent effects for water, ethanol, acetonitrile, and the solvent mixtures on the lignin-cellulose adhesion. In water, the hydrophobic lignin-water interactions enhance the lignin-cellulose association, which results in stronger LJ interactions. Moreover, this association facilitates the occurrence of hydrogen bonding between lignin and cellulose, and long-lasting H-bonds are observed. In ethanol, the cellulose-lignin coordination decreases significantly compared to water (2-times) because of the amphiphilic nature of ethanol. However, its presence enhances the Coulomb electrostatic interactions of the complex that results in non-effective disruption of the cellulose-lignin interactions. On the contrary, the presence of acetonitrile results in a more effective disruption of the cellulose-lignin interactions compared to water or ethanol, due to stronger disruption of both LJ and Coulomb interactions. Furthermore, our results show that the lowest adhesion between lignin and cellulose is found in the presence of binary water-organic solvent mixtures. Moreover, the interaction energies pass through a minimum interaction in 50% acetonitrile-water and 75% ethanol-water mixtures.

To understand the origin of the cellulose-lignin adhesion, we compare the results obtained for the cellulose-solvent and lignin-solvent $g(r)$ s, studied in Chapter 4 and the cellulose-lignin $g(r)$ in Chapter 5. The $g(r)$ s between the solvents and the individual wood components in Chapter 4 (Fig. 6 and Fig. 16), reveal well-defined solvent coordination shells around the cellulose and lignin models, in pure and water-ethanol and water-acetonitrile mixtures. For comparison, we studied the solvent organisation around the lignin-cellulose complex studied in Chapter 5 (Figs. S6.1 to S6.5). The cellulose-solvent and lignin-solvent $g(r)$ s in pure, 50 wt%, and 75% cosolvent concentration reasonably reproduce the trends of the profiles obtained in Chapter 4. The detailed analysis of the coordination of these solvents to particular oxygen sites in cellulose (O2, O3, O4, O5 and O6) and lignin (O α , O β , O γ and OMe) indicates different preferred coordinations of water, ethanol

and acetonitrile. A schematic summary of preferred coordination sites of water, ethanol, acetonitrile and lignin to the cellulose oxygens, is presented in Fig. 6.1

In cellulose, the $O_{\text{cell}}-O_{\text{water}}$ $g(r)$ s show that water preferentially coordinates O6 (hydroxymethyl group), which makes this oxygen site less accessible to other molecules. Thus, the strong $O6_{\text{cell}}-O_{\text{water}}$ coordination governs the lignin interaction towards O3 and O2 in the pure water, as shown from $O_{\text{cell}}-O_{\text{lig}}$ $g(r)$ in Figure 5.7. The amphiphilic nature of ethanol limits the lignin interaction with the O6 and O2 sites, but the same amphiphilic behaviour results in the non-effective disruption of the cellulose-lignin interactions in pure ethanol. Moreover, comparing the $O_{\text{cell}}-X_{\text{sol}}$ preferential coordination, the solvent coordination of O6 sites in cellulose is mostly electrostatically driven, whereas coordination of less exposed O2 or O3 sites require higher affinity with the less polar aliphatic hydrogen atoms of the glucose rings. Thus, acetonitrile can limit the lignin interaction with the O2 and O3 sites and reduce the LJ and Coulomb interactions better than ethanol due to its less polar character. However, ACN is less effective to entirely disrupt the cellulose-lignin interaction with the cellulose O6 site in pure acetonitrile due to its aprotic character, resulting in residual electrostatic Coulomb lignin-cellulose. In the water-organic mixtures, acetonitrile preserve up to the largest extent its preferred coordinations towards the cellulose sites found in the pure organic solvents; contrary to ethanol, which changes its preferential coordination in both cellulose surfaces as a result of water-ethanol competition.

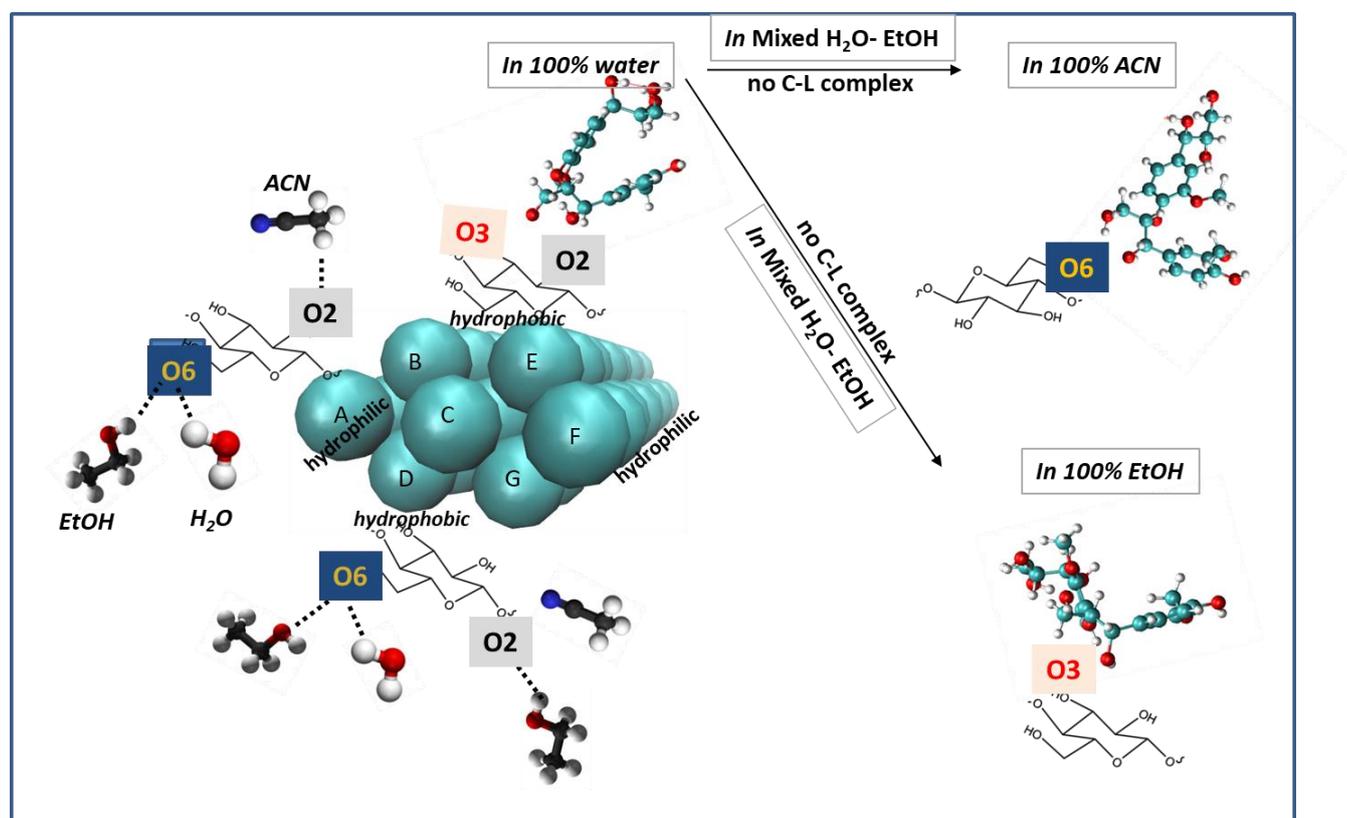


Figure 6.1 Preferred coordination of water (H₂O), ethanol (EtOH) and acetonitrile (ACN) to cellulose O-sites, and preferred coordination of lignin to cellulose O-sites in 100% water, ethanol and acetonitrile solvents. C-L denotes cellulose-lignin complex, which is not observed in the mixed water-organic solvents.

In lignin, the $O_{\text{lig}}-X_{\text{sol}}g(r)$ s show that all pure water, ethanol, and acetonitrile solvents display the highest coordination towards O_{γ} and O_{α} . The slightly stronger preferential interaction of the solvents with the O_{γ} site in lignin limits its interaction with the cellulose sites. Nonetheless, Fig. 5.7 displays a significant $O_{\gamma_{\text{lig}}}-O_{2_{\text{cell}}}$ coordination in the presence of water. This can be easily explained due to the enhanced lignin-cellulose association in order to reduce the hydrophobic interactions with water. Moreover, in the absence of the intramolecular H-bond between $O_{\gamma}-O_{\alpha}$ sites, lignin stabilises its conformation by participating in hydrogen bonds with cellulose. Furthermore, Figs. S4-5 show that the water-lignin and water-ethanol site coordination order is preserved. On the contrary, the acetonitrile coordination in water-acetonitrile mixtures of 25 wt% and 50 wt% displays a higher preference for the O_{α} site than O_{γ} . Thus, our results suggest that the preferential interactions of water and acetonitrile for different sites in both lignin and cellulose results in a more effective disruption of the lignin-cellulose interactions in water-acetonitrile solvents than in ethanol-water mixtures, and the pure solvents.

2. Cellulose-solvent interactions: the water exclusion and the entropic effect

Furthermore, the presence of two solvent components in the mixtures modifies the number of coordinated solvent molecules depending on the chemical nature of the organic component. As demonstrated by the cumulative numbers (Figs. 4.7 and S6-6), the addition of EtOH and ACN in water leads to the exclusion of H_2O from the solvation shell of the cellulose and lignin. The ratio of water molecules between hydrophilic and hydrophobic surfaces is ~ 1.2 - 1.3 in water-ethanol, whereas it is 1.5 - 1.8 in water-acetonitrile mixtures (Figs. 4.7 and S6-6). Moreover, the ratio of cosolvent molecules between hydrophilic and hydrophobic surfaces is 0.9 - 1 in water-ethanol, whereas it is 0.6 - 0.8 in water-acetonitrile. Thus, our results indicate that there is an accumulation of acetonitrile molecules around the hydrophobic surface, whereas the substantial exclusion of water from the hydrophobic surface results in more hydrated hydrophilic surfaces, compared to the water-ethanol mixtures. Moreover, a saturation-like behaviour of acetonitrile and water molecules around the hydrophobic surface is observed.

We showed in Chapter 4 (Fig. S4.7-9) that Lennard-Jones interactions are the driving force for cellulose-acetonitrile interactions at all concentrations, whereas the cellulose-water interactions are always electrostatically driven. On the other hand, ethanol-cellulose are mainly contributions from Coulomb electrostatics in the pure solvent, and they are further equivalent to LJ in the presence of water. Chen et al. [1] demonstrated that the primary molecular mechanisms for urea absorption on cellulose are the Lennard-Jones dispersion energies as well as the gain in the translational entropy of water expelled from the cellulose surface, in water-urea mixtures. Moreover, the authors show that the decrease in the entropy of urea adsorbed on the cellulose surface is always lower than the increase in the entropy of water.

In this work, we do not study explicitly the entropy contributions caused by the solvents and their interactions with the wood compounds. However, we can qualitatively compare our hydrogen bond and cumulative numbers analysis, since the H-bond lifetimes implicitly describe the freedom of water molecules in the first solvation shell of cellulose. The cellulose-water H-bond lifetimes' analyses from the autocorrelation function in Chapter 6 (Table 4-2) indicate that cellulose-water H-bonds last longer in the presence of ethanol than in acetonitrile. The addition of the organic phase to water results in a reduction of the translational freedom of water molecules. This is because the water molecules participate in hydrogen bonding with ethanol and acetonitrile (Table 3.3), which results in longer-lasting water-water hydrogen bond lifetimes in the mixture than in bulk water (Table 3.4). Moreover, the higher frequency of H-bonds in water-ethanol than in water-acetonitrile mixtures results in a higher restriction of water motion in the presence of water than in acetonitrile. Correspondently, the cellulose-water H-bonds are also longer lasting in the presence of ethanol than in acetonitrile (Table 4.2), restricting the motion of water in the first solvation shell. As outlined by Chen et al. [1], the water exclusion from the first solvation shell of cellulose results in a higher entropic gain than the decrease in the entropy the cosolvent adsorbed on the cellulose surface. Thus, the water exclusion at the hydrophobic surface described above is expected to result in an entropic gain, particularly in the presence of acetonitrile than in ethanol because of the stronger water exclusion from the hydrophobic cellulose surface.

Furthermore, the variety of the polymers present in wood, with different amphiphilic behaviours, demands almost different solvent characteristics to optimise and stabilise the different wood polymers. Hemicellulose and the hydrophilic domains of cellulose and lignin (the propyl alcohol groups) will interact preferentially with water. Whereas, the presence of the organic solvent can stabilise the most hydrophobic domains of cellulose and lignin. Thus, the water-organic solvent mixture act together to optimise the interactions with the wood polymers. While the presence of water is useful to disrupt the electrostatic interactions within and between the polymers, its hydrophobic nature probably causes a less potent effect with the less polar domains of cellulose and lignin; therefore, enhancing the cellulose-cellulose, lignin-lignin, and cellulose-lignin interactions. The presence of the organic phase, however, reduces the hydrophobic response of lignin and cellulose to the mixed water-organic solvent by destabilising their stacking interactions. Thus, this results in a reduction of the stacking forces between the cellulose and lignin, and an increase of the surface area of lignin and the hydrophobic domain of cellulose. Therefore, a synergetic effect to disrupt both electrostatic and stacking interactions is achieved in the presence of both water and the organic solvent.

3. Solvent effects on molecular wood structures: possible implications in lignocellulose adhesion and wood swelling

The average solvent-accessible surface area (SASA) of cellulose (Figure 3.3) revealed an increase of the surface upon solvation compared with the initial bulk-crystal, particularly in the mixed solvent region. Furthermore, a higher increase of the cellulose SASA is observed in acetonitrile than in ethanol mixtures,

whereas this trend is inversed in the pure organic solvents. This behaviour correlates qualitatively with the hyper-swelling profiles of Chang et al. [2] (Figure 1.7), which show a nonlinear relation between the volume swollen and the concentration of the organic solvent. Thus, our results suggest that the increase of the surface area of cellulose in mixed water-organic solvents could have some implications on the observed swelling in wood. Moreover, aside from the changes in the hydroxymethyl group conformations and the H-bond network, the cellulose nanocrystal studied in this work remains intact in all the studied solvents.

The conformation changes of the lignin dimers and tetramers studied through the SASA, dihedral angles, and stacking (Figures 4.15 and 4.16) suggest that water-organic mixtures increase the lignin solubility. Our results are in a qualitative agreement with recent studies that show an increase of lignin solubility in water-ethanol mixtures[3]. The effect of water-solvent mixtures on the cellulose-lignin interaction is concomitant with the changes from stacked to more extended lignin conformations most likely resulting from the decreased interaction with cellulose chains. A broader lignin SASA distribution observed for the 4 dimers in complex with cellulose (Fig S6.7) compared to the SASA distribution of a single lignin dimer in water (Fig. 4.16) shows that cellulose stabilises lignin at reducing the hydrophobic interactions, which allows the exposure of a more extended lignin surface when interacting with cellulose than with water alone. Nonetheless, the interactions of lignin with water remain highly hydrophobic, the inclusion of the organic component lowers these hydrophobic interactions while disrupting the cellulose-lignin complex.

Notably, the surface area of the lignin models, here studied, vary in a wide range (particularly when compared with the surface of the cellulose nanocrystal), as both experimental and theoretical studies suggest that β -O-4 structures are flexible molecules that can adopt a large number of conformations [4-6]. Thus, our results suggest that solubilization of lignin-rich domains and the detachment of polymer-lignin interfaces can influence the wood swelling, in agreement with previous experimental and computational observations [7-11]. Indeed, it has been evidenced that morphological changes of lignocellulosic biomass as exposed to pure and mixed solvents plays an essential role in the increase the swelling, therefore, promoting the fractionation of the material. Langan et al.[7, 8] showed in steam explosion pretreatment of poplar chips that cellulose dehydration (of the microfibrillar bundles) and the separation of the lignin-hemicellulose complex, followed by pressure of lignin globules to separate, seem to cause swelling and will expand the size of small voids by a few nanometers. Furthermore, other authors have observed changes in the wood structure thanks disruption of the compound middle lamella, thus suggesting a partial dissolution of lignin from the matrix [9, 10].

However, the presence of the organic phase can also cause undesirable effects on the interactions between wood compounds. In this work, we have shown that the presence of the organic solvent slows down the detachment between xylan and lignin, particularly in the presence of ethanol, by decreasing xylan's mobility and showing a gel-like behaviour[12-14]. The presence of acetonitrile at low concentration also tends to have a similar effect, despite our results show that it increases the xylan's mobility when compared to water. Moreover, acetonitrile-water mixtures seem to perform overall better at disrupting the cellulose-lignin and the

xylan-lignin interactions than ethanol-water mixtures. Furthermore, our results have shown that xylose adopts the twofold conformation in the presence of the organic solvent. Falcoz-Vigne et al.[15] have observed that the interaction between xylan and cellulose is stronger when xylan is in the twofold conformation than when it is in the threefold one. Thus, our results suggest that the interaction between xylan and lignin is also enhanced in the presence of the organic solvent arising from a hydrophobic interaction, which results in the stabilisation of lignin in the presence of xylan. Therefore, the outcome of swelling in pure and mixed solvents depends on the synergetic interactions in the macromolecular arrangement of the different wood polymers, and their interaction with the solvents, which we have shown to follow a non-linear behaviour.

4. Assessment of the achieved scope of this work, the reliability of model size and simulation time scales, and future perspectives.

The variety and structural heterogeneity of the lignocellulose biomass is too complex and not yet entirely resolved. Moreover, it varies with the origins of biomass resources [16]. This makes the complete understanding of lignocellulosic interactions far from trivial for both experimental and theoretical efforts. Nonetheless, providing detailed and systematic descriptions of pieces of lignocellulosic components, and in particular, at the atomistic (micro) scale, is highly desirable to unveil fundamental Physico-chemical properties being yet largely in-understood despite recent efforts [17-19].

For all these reasons, the systematic modelling of inter-atomic interactions between the main lignocellulose compounds can provide insight into the fundamental interactions on the cell wall. MD or even QM (or QM/MM) simulations of the pairwise interactions between lignocellulose compounds in explicit solvent mixtures can be useful to gain understanding at the molecular level of the interactions between wood components in the presence of pure and water-organic solvent mixtures. The lignocellulose interactions evaluated in this work were assessed using rather simple models of cellulose, hemicellulose, and lignin. We do not consider that these models may wholly correlate the entire complex behaviour in the biomass structure. Nevertheless, we estimate that the properties and trends, we assessed here and established for the first time, are not biased by the size of the models, because of their predominantly local, pairwise atomistic character. For example, the distinct molecular groups in cellulose, lignin and xylan, revealed by our studies to undergo solvent-dependent conformational changes, allowing to form or deform solvent-dependent intra- or intermolecular H-bonds, and the characteristic oxygen sites competing for H-bond bindings with a specific type of solvent molecule in the binary water mixtures are well captured with the sizes of the models considered here. The sizes are also limited by the fact that we wanted to study the phenomena with all-atom MD models, in order to describe the H-bonds correctly. The absolute values of SASA or the interaction energies are evidently model-size dependent, but the established trends are not expected to alter with the size of the models. This is indeed what we established comparing our results, for instance about lignin ring conformations, with a recent study of lignin in THF –water solvent of [20] with much larger lignin polymer

representation. The good agreement between our conclusion with the conclusions of the authors in ref. [20] [17] demonstrates that indeed, the trend is captured even lignin is considered as a dimer. This is not surprising and can be anticipated keeping in mind that the primary interactions in polymers structures are largely determined by the interactions between adjacent individual molecular units (polymer building blocks). The higher-order structuring and properties, such as polymer chains organisations, cellulose crystallinity and its mechanical properties, among others, will undoubtedly need to envisage much larger size and time scales modelling with coarse grain or finite element based theoretical methods. In any case, such rather macro-scale properties are behind the scope of the present work and to our humble opinion are better studied with the today available experimental spectroscopy techniques. Here, we aim to capture local interatomic effects, hardly accessible even with the modern experimental techniques.

Further on, we note that the cellulose nanocrystal consists of seven strands in our model, while cellulose in primary walls contains 18-36 strands. The much smaller model size, considered by us, does not prohibit understanding the solvent-induced conformational freedom of cellulose and the specific solvent-cellulose pairwise interactions at a short-range. We admit that smaller size models could suffer to estimate precisely the true vdW interactions; however as demonstrated previously from DFT augmented with empirical London dispersion term [], the dispersion interactions in alkane chains is beared by CH_2 unit, being of about 3 kcal/mol and increases linearly with the increase of CH_2 number in the alkane chains. Therefore, the conclusions, reported in Chapter 5 are valid lower limit of vdW interactions, which also makes valid our conclusions about the vdW trends in the different solvents and the effect of solvents.

Moreover, we built the smallest possible cellulose model to represent the solvent-accessible polar groups, in both hydrophilic and hydrophobic surfaces. As it is demonstrated in Chapter 4, the aliphatic groups are only surface accessible in cellulose hydrophobic chains (BDEG), which confers the more hydrophobic character to this surface than to cellulose chains AF. Thus, cellulose model can capture the more hydrophilic and hydrophobic characters of the different surfaces correctly. Again, our results are in line with those available for larger models and more extended dynamics simulation: the cosolvent accumulation at the hydrophobic surfaces and the water accumulation at the hydrophilic surface of our 7-strand nano-crystal model in the presence of water and acetonitrile reasonably reproduce the observations of Smith et al. in water-tetrahydrofuran mixtures [20]. Moreover, the spontaneous phase separation comes as a result of the observed microheterogeneity reported in Chapter 3 and in agreement with previous studies of binary water-acetonitrile solvents, in which a non-uniform mixing of water-acetonitrile results on isolated water-rich and acetonitrile-rich domains.

To be considered is that the cellulose-lignin and xylan-lignin adhesion were studied with guaiacyl dimer and tetramer, respectively, but not with longer lignin polymers. The size and complexity of the model is an important factor when studying their solubility since such property depends on the molecular weight. Here, we did not study the solubility or the cellulose-lignin adhesion in true wood materials. Again, this was beyond

our aims and is beyond our modelling interests. Moreover, the lignin polymer is composed of a variety of lignin monomers and linkages, and we estimate that any model considered will be very limited to describe observed macro-scale effects. We therefore, assessed the effect of pure water and organic solvents as well as their mixtures on the primary lignin dimer on cellulose adsorption and provided estimates about the intermolecular energy evolution in few tenths of ns time-scale in order to capture the nature of inter-atomic forces keeping together lignin to cellulose and the effect on these forces of solvent compositions. The analysis of our results about the more macroscopic understandings as the hydrophobic or hydrophilic interactions logically confirmed that the intra-molecular interactions are driving the well accepted in chemistry hydrophilic/hydrophobic behaviours of macromolecules. Thus, our results do not fully capture all the features that dominate the lignocellulose interactions, but they clearly provide a well-founded picture about the intrinsic, atomistic trends of the solvent-dependent inter- and intramolecular interactions, not previously reported

It is essential to comment on the simulation time lengths. The phenomena we studied (torsion angle molecular conformational changes, H-bond formations, non-bonded interactions) typically occur in the *ps* to *ns* time scale. As most of the results were achieved with parameters based methods, as are the MD simulations, to obtain a reasonably good ensemble description we performed the simulations within few tenths of ns time scale for each lignocellulosic component in a distinct solvent composition. As shown from the energy error bars in Chapter 5, the confidence in the energies of the adsorbed lignin dimer to cellulose is rather satisfactory. Besides, the well-emphasised differences in terms of interaction energies and structures, between in gas-phase adsorbed lignin on cellulose and in the considered solvents, clearly supports the conclusion the methodology used here in terms of force-field parameters, size of the models, and the chosen simulation time lengths are appropriate. Indeed, we found the same tendency as those reported in few works devoted to similar investigations but applying larger size and longer simulation time scales. Besides, we note that the results, presented and analysed in Chapters 4 and 5, were issued from more than 1.2 μ s of total simulation production time, which could have been sufficient to sample the conformational space of more complex and more significant MD systems, but in significantly fewer solvents, meaning that we could not capture the trends of solvent composition effects. Instead, we performed shorter production runs that could lead to the equilibrium of the solvent molecules around lignocellulose components. Moreover, the smaller models allow to reduce the correlation times and observe some configurational changes faster than in a bigger complex. Thus, the considered simulation time of our production runs lead to the equilibrium of the solvent molecules. Moreover, it adequately samples the conformational space of lignocellulose components as follows from the results evidencing the evolution over time and the relatively small error bars of the lignin dimer-cellulose interaction energies. The latter results show that the current simulation reasonably converged for the simulation times studied.

Our results have been obtained utilising the CHARMM force field, the solvent parameters within the force field, and the CHARMM-compatible lignin force field and, which have been recurrently used in the modelling of lignocellulose interactions [21-23]. The MD simulations of liquids can be more convenient if using the OPLS FF since this has been optimised for simulation of liquids. However, the results in Chapter 3 show that the CHARMM force field reproduce reasonable well the thermodynamic and structural properties of interest for this work. Moreover, CHARMM atom charges are parameterised based on the optimisation of H-bond networks, which can be advantageous in the analysis of H-bonds and the coordination of the solvent oxygens in the solvent. Furthermore, the computational modelling of lignocellulose is a promising field, and better and more accurate methods and parameters are developing every day [23, 24]. Moreover, the perspective of this work goes further in the QM/MM direction, where DFT techniques can model the solute, whereas the solvent can be modelled by the at classical level by the OPLS force field, as illustrated in Chapter 4.

References

1. Chen, P., et al., *Translational Entropy and Dispersion Energy Jointly Drive the Adsorption of Urea to Cellulose*. The Journal of Physical Chemistry B, 2017. **121**(10): p. 2244-2251.
2. Chang, S., et al., *Solvent polarity and internal stresses control the swelling behaviour of green wood during dehydration in organic solution*. Bioresources, 2012. **7**(2).
3. Goldmann, W.M., et al., *Solubility and fractionation of Indulin AT kraft lignin in ethanol-water media*. Separation and Purification Technology, 2019. **209**: p. 826-832.
4. Besombes, S. and K. Mazeau, *Molecular dynamics simulations of a guaiacyl β -O-4 lignin model compound: Examination of intramolecular hydrogen bonding and conformational flexibility*. Biopolymers, 2004. **73**(3): p. 301-315.
5. Stomberg, R. and K. Lundquist, *Stereochemistry of lignin structures of the β -O-4 type*, in *Nordic Pulp & Paper Research Journal*. 1994. p. 37.
6. Besombes, S., et al., *Molecular Modeling of Lignin β -O-4 Model Compounds. Comparative Study of the Computed and Experimental Conformational Properties for a Guaiacyl β -O-4 Dimer*, in *Holzforschung*. 2003. p. 266.
7. Langan, P., et al., *Common processes drive the thermochemical pretreatment of lignocellulosic biomass*. Green Chemistry, 2014. **16**(1): p. 63-68.
8. Pingali, S.V., et al., *Morphological changes in the cellulose and lignin components of biomass occur at different stages during steam pretreatment*. Cellulose, 2014. **21**(2): p. 873-878.
9. MEIER, P., et al., *Multiple Swelling of Pinewood (Pinus Sylvestris) in Binary and Ternary Mixtures of Ethanol, Acetone and Water* MATERIALS SCIENCE (MEDŽIAGOTYRA), 2006. **12**(1).
10. Meiera, P., et al., *Mechanical properties of pinewood (Pinus Sylvestris) swollen in organic liquids* Proc. Estonian Acad. Sci. Eng., 2006. **12**(2): p. 125-133
11. Aguilera-Segura, S.M., et al., *Synergistic sorption of mixed solvents in wood cell walls: experimental and theoretical approach*. Macromolecular symposia, 2019. **386**(1): p. 1900022.

12. Yuguchi, Y., et al., *Gelation of xyloglucan ID water/alcohol systems*. Cellulose, 2004. **11**(2): p. 203-208.
13. Yamanaka, S., et al., *Gelation of tamarind seed polysaccharide xyloglucan in the presence of ethanol*. Food Hydrocolloids, 2000. **14**(2): p. 125-128.
14. Umemura, M. and Y. Yuguchi, *Solvation of xyloglucan in water/alcohol systems by molecular dynamics simulation*. Cellulose, 2009. **16**(3): p. 361-371.
15. Falcoz-Vigne, L., et al., *Quantification of a tightly adsorbed monolayer of xylan on cellulose surface*. Cellulose, 2017. **24**(9): p. 3725-3739.
16. Parham, R.A. and R.L. Gray, *Formation and Structure of Wood*, in *The Chemistry of Solid Wood*. 1984, American Chemical Society. p. 3-56.
17. Zhang, Y., et al., *Recent progress in theoretical and computational studies on the utilization of lignocellulosic materials*. Green Chemistry, 2019. **21**(1): p. 9-35.
18. Lambert, E., et al., *Real Time and Quantitative Imaging of Lignocellulosic Films Hydrolysis by Atomic Force Microscopy Reveals Lignin Recalcitrance at Nanoscale*. Biomacromolecules, 2019. **20**(1): p. 515-527.
19. Kang, X., et al., *Lignin-polysaccharide interactions in plant secondary cell walls revealed by solid-state NMR*. Nature Communications, 2019. **10**(1): p. 347.
20. Smith, M.D., et al., *Cosolvent pretreatment in cellulosic biofuel production: effect of tetrahydrofuran-water on lignin structure and dynamics*. Green Chemistry, 2016. **18**(5): p. 1268-1277.
21. Smith, M.D., et al., *Enhanced sampling simulation analysis of the structure of lignin in the THF-water miscibility gap*. Physical Chemistry Chemical Physics, 2016. **18**(9): p. 6394-6398.
22. Smith, M.D., et al., *Organosolv-Water Cosolvent Phase Separation on Cellulose and its Influence on the Physical Deconstruction of Cellulose: A Molecular Dynamics Analysis*. Scientific Reports, 2017. **7**(1): p. 14494.
23. Vermaas, J.V., et al., *Systematic parameterization of lignin for the CHARMM force field*. Green Chemistry, 2019. **21**(1): p. 109-122.
24. Vermaas, J.V., et al., *Automated Transformation of Lignin Topologies into Atomic Structures with LigninBuilder*. ACS Sustainable Chemistry & Engineering, 2019. **7**(3): p. 3443-3453.

Chapter 7. First-Principles Simulations of Two-Dimensional Electronic Spectroscopy of lignin model monomers and their β -O-4 linked dimer

Contents

Abstract	213
1. Introduction	213
2. Computational methods.....	216
2.1 Selected structures.....	216
2.2 Excited-State calculations of monomers and dimers	217
2.3 Simulations of 2D Electronic Spectra	217
3. Results and discussion.....	218
3.1 Electronic spectra of anisole and lignin monomers	218
3.2 CAS vs RAS 2D electronic spectra of lignin monomers.....	223
3.2 2D Electronic spectra of unstacked and stacked lignin dimer conformers	224
4. Conclusions	229

Abstract

In this work, we studied the electronic structure and conformations in water of a guaiacyl lignin dimer employing 2D electronic spectroscopy in the UV-VIS region, using the SOS//QM/MM methodology, with QM = CASSCF and RASSCF approaches. First, vertical excitation energies and 2D-UV/VIS excitation spectra were obtained for all lignin model monomers (guaiacol, creosol, homoveratrole) in gas phase, including all π and lone-pair electrons and orbitals in the active space. They were further compared against anisole and previously studied benzene and phenol, to aid in excited state assignments. We found an excellent agreement between the 2DEF obtained by both QM schemes for the monomers. Besides, we studied the 2D electronic spectra of stacked and unstacked lignin conformers in solution taken from MD simulations of the lignin dimer in water studied in Chapter 4. The 2DUV spectra show that charge transfer (CT) states become bright upon coupling, allowing to distinguish between two conformers with different degrees of stacking.

1. Introduction

Lignin is the most abundant natural aromatic polymer on earth, being one of the significant cell-wall components of wood and grass species, after cellulose [1]. Thus, its availability renders it the most abundant renewable source of aromatics and platform chemicals [2-8]. It is a three dimensional, highly branched, and polyphenolic molecule complex structure with high molecular weight, as illustrated in Figure 7. 1A, with a variety of lignin units and linkages. The phenylpropanoid unit is the basic structure of lignin, consisting of an aromatic ring and a 3-C side chain [9]. The monolignol subunits are derived from coniferyl, *p*-coumaryl, and sinapyl alcohols (Figure 7. 1B), and are referred to as, after polymerisation, *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) residues, respectively [10]. Two types of inter-unit linkages (C-O and C-C linkages) link several aromatic units, being the β -O-4 ether bond the predominant inter-unit linkage in the structure of lignin (Figure 7. 1C) [11].

Characterisation is essential because the chemical structure and composition of lignin varies with source, type of lignin, and isolation method. Chemical structure characterisation can be useful in determining lignin sources, degradation of lignin samples, and the relationship between physical and thermal properties, among others [12]. However, the study of lignin is a considerable challenge for experimental techniques due to the lack of a regular and ordered structure. The β -O-4 linkage represents the predominant inter-unit linkage in lignin, and it is an excellent model for studying major conformation features such as H-bonding and flexibility [13]. Thus, concerning the complexity of the lignin network, we have chosen here to study the guaiacyl β -O-4 dimer model compound. Nonetheless, the study of such model compounds can lead to the understanding of the behaviour of larger and more complex molecular systems.

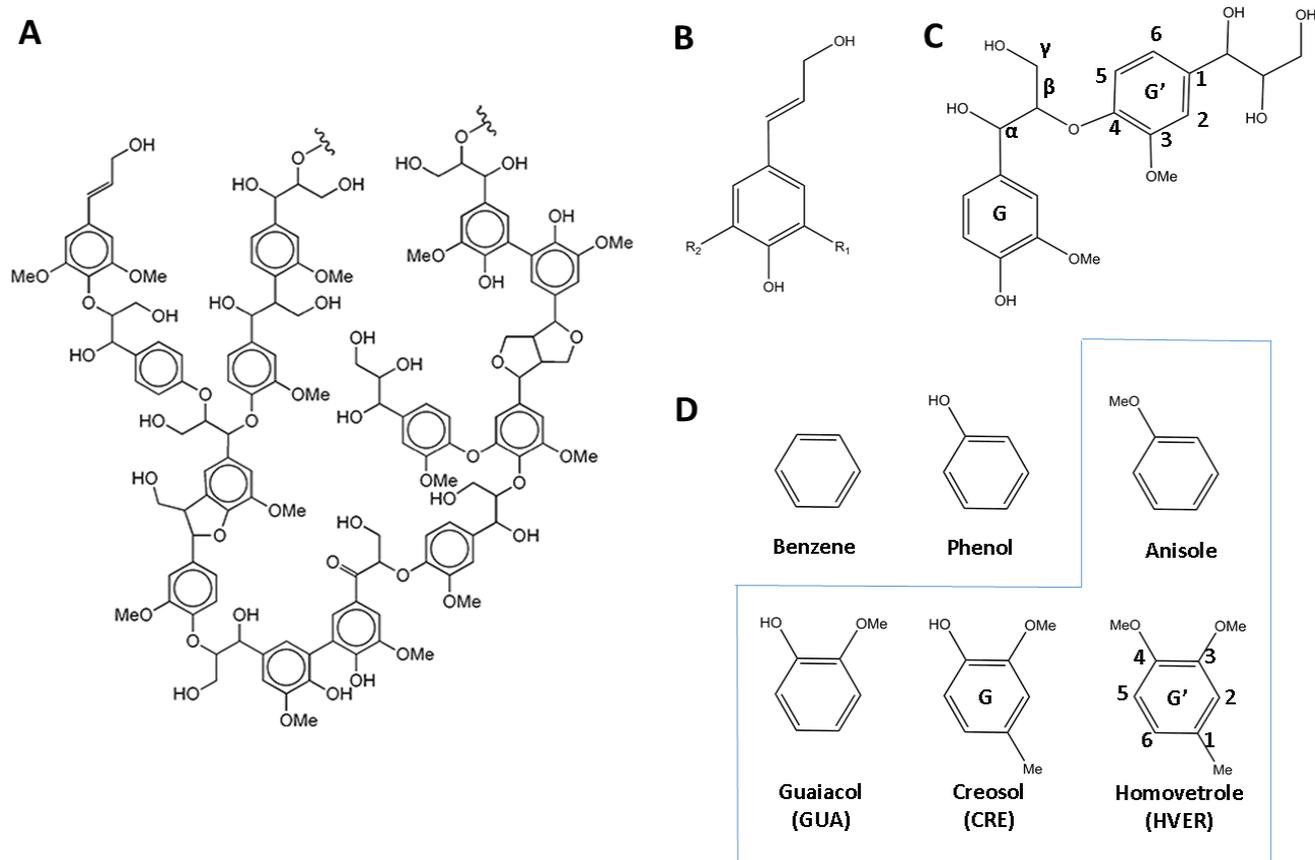


Figure 7. 1. (A) Chemical structure of a lignin polymer with different linkages and monomer units. Taken from ‘ACS molecule of the week’, July 2012 [10]. (B) The three monolignols before polymerisation: p-coumaryl ($R_1=R_2=H$), coniferyl ($R_1=H$, $R_2=OMe$), and sinapyl ($R_1=R_2=OMe$) alcohols. (C) Lignin dimer studied in this work, formed by two guaiacyl (G, G’) units linked with a β -O-4 ether bond. (D) Chemical structure of guaiacol (GUA), creosol (CRE), and homoveratrole (HVER) model lignin monomers studied in this work (in box), and benzene, phenol, and anisole molecules, to which they were compared.

Molecular modelling holds great potential for studying lignin and can provide valuable conformational information that remains experimentally unavailable. Both experimental and theoretical studies suggest that β -O-4 structures are flexible molecules that can adopt a large number of conformations [13-15]. We have shown that the interaction of lignin dimers in water remains highly hydrophobic. The lignin dimer adopts a stacked conformation, whereas it adopts a more unstacked conformation in pure ethanol and water-ethanol mixtures (Chapter 4) [16]. Nonetheless, our MD simulations of lignin dimers in complex with cellulose have shown that cellulose stabilises lignin allowing a more extended lignin surface when interacting with cellulose than in water alone (Chapter 5). Therefore, characterisation of the chemical structure of lignin conformers can be useful for determination of different degrees of aggregation in wood cells walls or when dissolved in solvents.

The aromatic nature of lignin makes it a great candidate for UV-VIS spectra studies as each of its aromatic rings provides native local probes for tracking lignin dynamics in solution. Two-dimensional (2D) UV electronic spectroscopy (2DES) [17-21] is a recently developed technique, which can disentangle signals

arising from different intermolecular interactions that remain hidden in the 1D pump-probe spectra. 2DUV electronic spectroscopy can target the π - π^* transitions of aromatic residues, and can be employed for tracking all sorts of localised, delocalised, energy and charge-transfer photoinduced phenomena; however, their interpretation proves to be challenging. Alternatively, computational simulation of 2DES can be used to separate the different peaks into their contributing specific electronic transitions, thus providing a tool to recognise, understand, and interpret the observed primary fingerprints [22]. In this approach, an explicit mixed quantum mechanics/molecular mechanics (QM/MM) methodology is employed for the evaluation of electronic excited state energies and transition dipole moments, in which different photochemical outcomes are associated with specific molecular conformations.

The Sum-Over-States (SOS) approach [23] and the QM/MM scheme (SOS//QM/MM) [22, 24] have shown great potential to simulate 2DES and to characterise different conformations in DNA and proteins. It has been shown that 2DES obtained with this approach can distinguish between two configurations of a peptide containing both unstacked (non-interacting) and π or T-stacked (interacting) peptide motifs containing phenolic side chains in proteins or DNA/RNA [25-28]. Multiconfigurational and multireference perturbation methods, such as complete active space self-consistent field (CASSCF) and second-order multireference perturbation theory (CASPT2) techniques, can be applied to calculate the electronic properties of multichromophoric system reliably.

The UV-VIS spectra of lignin polymers [12] and the lignin model compounds [29-32] have been studied recently, but the number of studies remain scarcely available. Experimentally, only the first and second transition states of lignin model monomers has been obtained, and their electronic nature and some spectroscopic signatures have been described at different levels of theory [32]. The GS-Lb transition of lignin dimer with a β -O-4 ether bond has also been studied; however, the electronic structure characterisation of lignin remains widely unexplored. In this work, we studied the electronic structure and conformations in water of a guaiacyl lignin dimer (Figure 7.1C) employing 2D electronic spectroscopy in the UV-VIS region, using the SOS//QM/MM approach.

First, we obtained the vertical excitation energies and 2D-UV/VIS excitation spectra at the CASSCF(10,8)/6-31G* level for all lignin monomers (Figure 7. 1D) in the gas phase. We further compared them against anisole and previously studied benzene and phenol to aid in excited state assignments. Furthermore, we compare the computed spectra and excitation energies with the calculations at the RASSCF(4,5|0,0|4,3)/6-31G* level, a less computational demanding scheme in comparison with the complete active space calculations. We found an excellent agreement between the 2DEF obtained by both schemes for the monomers. Furthermore, we studied the 2D electronic spectra of stacked and unstacked lignin conformers in solution taken from previously reported MD simulations. We were able to track spectroscopic fingerprints between them, showing that 2D electronic spectroscopy has excellent potential as a novel diagnostic tool to track the folding state of different polymer types.

2. Computational methods

2.1 Selected structures

Guaiacol (GUA), shown in Figure 7. 1D, has been recently used as a minimal chromophore monomer of guaiacyl [32]. In order to mimic better the para-alkylation in lignin, we have also studied creosol (CRE). Notably, guaiacyl rings in our lignin dimer model are not identical. The lignin ring G' (Figure 7. 1C), which is more like a monomer found in native lignin, includes the ether bond rather than the hydroxyl group. Thus, to include both the para-alkylation and the β -O-4 ether bond effect, we studied 3,4-dimethoxytoluene, also known as homoveratrole (HVER). We built the GUA, CRE, and HVER monomers with Molden and their geometry was initially optimised in the gas phase at the PBE/6-31G* level. Subsequently, MP2/6-31G* level optimisation using Gaussian was achieved. An active space of ten electrons and eight orbitals (i.e. CASSCF (10,8)) was further used for the QM calculations, using MOLCAS.

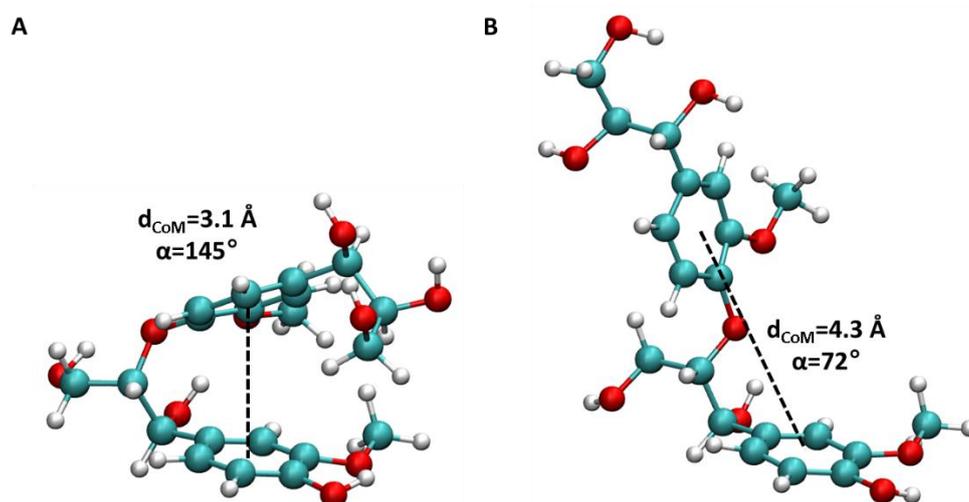


Figure 7. 2. Lignin dimer structures of stacked (A) and unstacked (B) conformers in solution from MD simulations (water molecules are not shown). The distance between the centres of mass of the two-chromophore rings, d_{CoM} , is shown as an indicator of stacking between the chromophores. The angle between the planes of the two chromophore rings, α , is also measured to differentiate between co-planarity ($\alpha \sim 0^\circ$, $\alpha \sim 180^\circ$) and T-stacking ($\alpha \sim 90^\circ$).

Lignin dimer structures (stacked and unstacked) were selected from the previously reported MD of lignin in water (Chapter 4) [16]. We performed a root mean square deviation (RMSD)-based cluster analysis using the analysis tool of GROMACS. First, the RMSD of atom positions between all pairs in the lignin dimer structure was determined. We varied the RMSD cut-off between 0.10 to 0.15 nm in 0.01 nm increments and performed cluster analysis for each RMSD value. For all RMSD clusters, we found the same representative structure as the centroid of the most populated clusters, having a stacked conformer, as shown in Figure 7. 2A. An unstacked dimer, a less populated cluster structure (Figure 7. 2B), was chosen for comparison and for testing the method accuracy concerning the monomer calculations. After selection, MD snapshots were extracted

containing the lignin dimers, along with a 15-Å radius water drop. Snapshot refinement was performed at the QM/MM level with the Cobramm package. The lignin dimer was included in the QM layer, whereas the water atoms were treated classically and were kept frozen. As with the monomers, PBE/6-31G* followed by MP2/6-31G* level optimisations was performed. An active space of twenty electrons and sixteen orbitals, i.e., CASSCF (20,16), was used for the QM calculations at the QM/MM level with the Cobramm package, using MOLCAS (QM) and Amber (MM).

2.2 Excited-State calculations of monomers and dimers

Multiconfigurational calculations were performed using state-average (SA)-CASSCF or -RASSCF methodology as implemented in the Molcas 8.1 code, including up to 30 and 50 states in the state-averaging procedure for monomers and dimers, respectively. SA-CASSCF and SA-RASSCF calculations were followed by single state (ss)-CASPT2 calculations to account for dynamic correlations. Transition dipole moments (TDMs) were calculated at SA-CASSCF or SA-RASSCF level using the RASSI routine of the Molcas code. Cholesky decomposition was used to accelerate the computations of two-electron integrals.

Calculations of GUA, CRE, and HVER monomers in the gas phase were performed at the SA50-ss-CASPT2//CASSCF(10,8)/6-31G* level including all valence π -electrons and the lone oxygen pairs, and the n and π -orbitals of the monomer chromophores. Furthermore, we compare the computed spectra and excitation energies with the calculations at the SA50-ss-CASPT2//RASSCF(4,5|0,0|4,3)/6-31G* level, a less computational demanding scheme in comparison with the CAS calculations. In the RAS scheme, three virtual orbitals were included in the RAS3 space, where up to quadruple transitions are allowed. Similarly, five occupied orbitals were included in the RAS1, where a maximum number of four holes is allowed.

As with the monomers, reference SA10-ss-CASPT2//CASSCF(20,16)/6-31G* calculations for the unstacked dimer in water were performed to establish a frame of comparison with the SA50-ss-CASPT2//RASSCF(4,10|0,0|4,6)/6-31G* scheme. The 6-31G* basis set was used with the following contraction scheme: C, O/[3s2p1d] and H/[2s].

2.3 Simulations of 2D Electronic Spectra

The transition dipole moments (TDMs) and energies were obtained at the multiconfigurational (CASSCF and RASSCF) and multireference level (PT2). The SOS//QM/MM approach has been already used for resolving the 2DUV/VIS spectra of a CFYC tetrapeptide solvated in water [22, 27]. The theory and methods have been summarised in the Theoretical Background and can be found in detail in refs [22, 25-28].

3. Results and discussion

3.1 Electronic spectra of anisole and lignin monomers

In this section, we compare the calculations at the *ss*-CASPT2/CASSFC(10,8)/6-31G* level of lignin monomers guaiacol (GUA), creosol (CRE) and homoveratrole (HVER) (Figure 7. 1D) below and above their ionization limits with the available experimental data and with reference CASSFC(8,7) electronic spectra calculations of anisole (this work), and the benzene and phenol monomers [28]. All valence π -electrons and lone pairs, and n and π -orbitals of the monomer chromophores have been included in the active space of each molecule monomer, as Illustrated in Figure 7. 3.

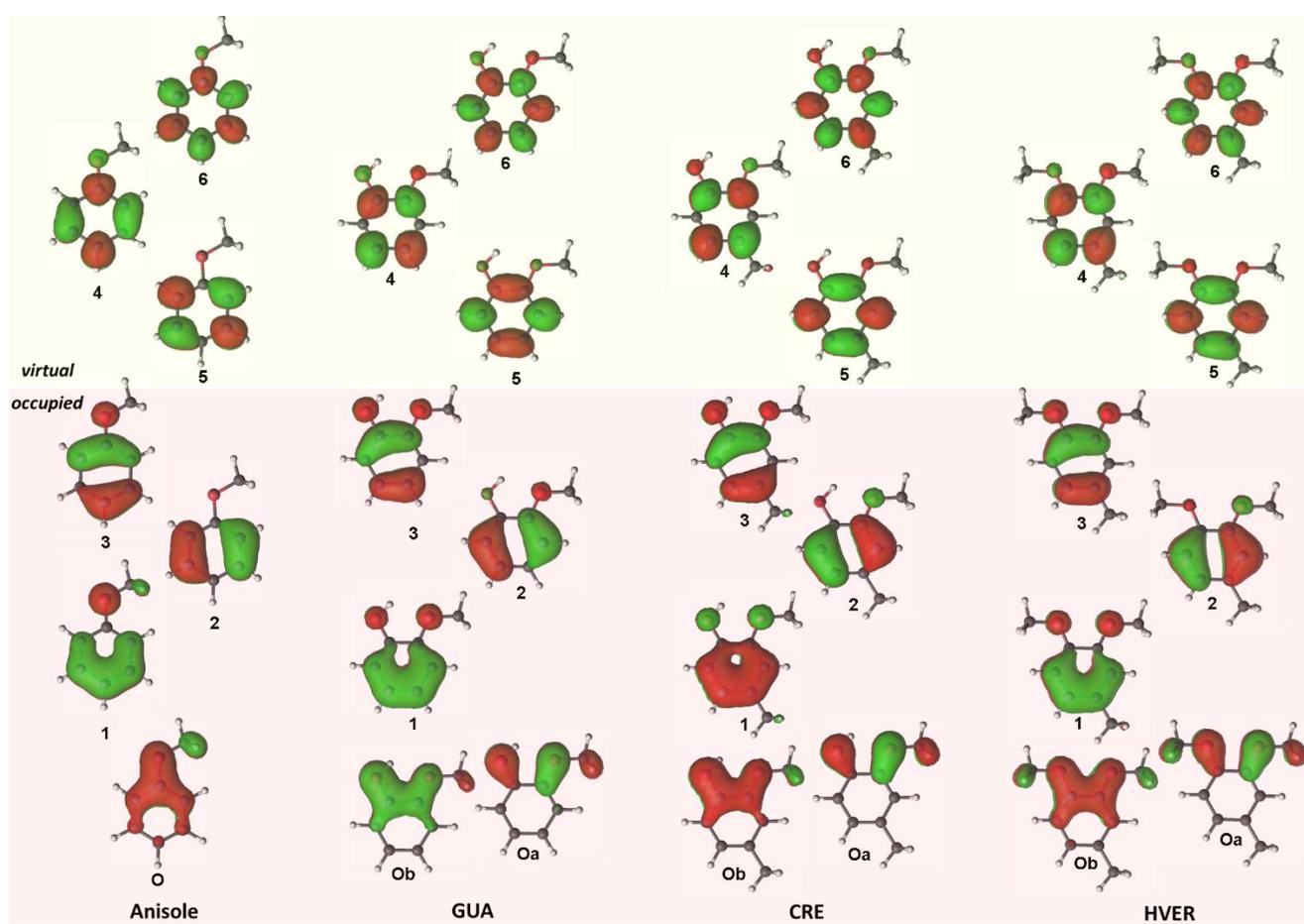


Figure 7. 3. Active space orbitals for the SA30-CAS(8,7) calculations of anisole and the SA30-CAS(10,8) calculations of GUA, CRE, and HVER monomers in the gas phase, organised by occupancy number at the CAS level. As expected, the molecular orbitals (MOs) of anisole and the lignin monomers GUA, CRE, and HVER in the active space share similar features, differing mainly in their symmetry. Anisole has a single oxygen lone pair orbital (O) localized in the oxygen and coupled with the aromatic ring, whereas GUA, CRE, and HVER have two delocalized oxygen lone pairs orbitals; the first has a bonding character (Ob) and displays a similar geometry as in anisole, and the second (Oa) is anti-bonding orbital. Furthermore, we notice that MO 5 in anisole resembles closely MO4 in GUA, CRE and HVER.

The first valence states below the first ionization energy of anisole, GUA, CRE and HVER monomers (labels 1-6, Table 7.1) have a similar nature to those of benzene and phenol [28] and for this reason we use the D_{6h} benzene notation to refer to them, although phenol, anisole and the lignin monomers have a different symmetry. The two lowest energy (π - π^*) excited states of GUA, labelled ${}^1B_{2u}$ and ${}^1B_{1u}$, have been experimentally detected at 4.44-4.59 eV [29-35] and 5.74-5.79 eV [33, 34], respectively. Furthermore, the $GS \rightarrow {}^1B_{2u}$ excited state transition of CRE has been experimentally reported at 4.35 eV [32]. As shown in Table 1, our ss-CASPT2/CASSCF(10,8)/6-31G* level calculation is in good agreement with these experimental data with a minor overestimation of the vertical transitions by ~ 0.13 eV. Furthermore, the experimental $GS \rightarrow {}^1B_{2u}$ transition energy red-shift difference of about 0.1 eV from GUA to CRE [32] is reasonably reproduced in our calculation (~ 0.08 eV). In the case of HVER, there is a moderate blue-shift (~ 0.01 eV) from the $GS \rightarrow {}^1B_{2u}$ transition of CREO due to the absence of the intramolecular hydrogen bond when replacing the $-OH$ by a $-OMe$ group, as it is the case of previously reported for catechol in comparison with GUA [31].

Table 7.9. Vertical $GS \rightarrow SN$ Excitation Energies (eV) and Transition Dipole Moments Magnitudes (au) in the anisole, guaiacyl, creosol, and homoveratrole. Anisole calculations were obtained at the CAS (7,8), while the rest of the monomers were obtained at level.

State	Anisole					Guaiacol					label
	E_E	Trans.	Coeff	GS	Lb	E_E	Trans.	Coeff	GS	Lb	
${}^1B_{2u}$	4.66	3->5	0.67	0.29		4.58	3->5	-0.67	0.32		1
	(4.59)	2->4	0.52			(4.45)	2->4	0.53			
${}^1B_{1u}$	6.07	3->4	-0.67	0.68	0.18	5.86	3->4	-0.71	0.62	0.15	2
	(5.78)	2->5	0.41				2->5	-0.38			
		3->5	0.27				1->4	-0.25			
		1->4	-0.24				3->6	0.22			
${}^1E_{1u}$	6.64	2->5	0.76	2.27	0.11	6.69	2->4	0.5	2.18	0.18	3
	(6.68)	3->4	0.46				2->5	-0.44			
							3->5	0.36			
							3->4	0.32			
${}^1E_{1u}$	6.96	2->4	-0.66	1.89	0.38	6.7	2->5	-0.51	1.99	0.25	4
	(6.9)	3->5	0.53				3->4	0.43			
		0->5	-0.24				2->4	-0.39			
							3->5	-0.32			
${}^1E_{2g}$	7.2	1->5	0.53	0.55	0.34	7.16	1->5	-0.55	0.89	0.31	5
		2,3->5	-0.42				2->5	-0.36			
		2->5,4	0.32				2->6	0.3			
		2->6	0.3				2->4	-0.27			
		2->4	-0.26				3->5	-0.27			
		3->4	0.24				2,3->5,4	0.24			
${}^1E_{2g}$	7.67	1->4	-0.45	0.2	0.1	7.22	2->5	0.47	0.24	0.16	6
		3->4	0.44				1->4	-0.42			
		3->6	0.35				3->5,4	0.37			
		3->4	0.26				3->6	0.32			
		2,3->5,4	0.26				2,3->4	0.28			
	Creosol					Homoveratrole					
${}^1B_{2u}$	4.5	3->5	-0.67	0.37		4.51	3->5	-0.68	0.39		1
	(4.35)	2p->4	-0.51				2->4	0.54			
${}^1B_{1u}$	5.83	3->4	0.68	0.66	0.12	5.65	3->4	0.73	0.72	0.18	2
		2p->5	-0.39				2->5	0.38			
		3->5	0.27				1->4	0.27			
		1->4	-0.25								
${}^1E_{1u}$	6.51	2p->4	-0.54	2.28	0.13	6.42	2->5	-0.62	2	0.34	3

${}^1E_{1u}$		2p->5	0.42			3->4	0.49			
		3->5	0.37			2->4	-0.24			
		3->4	0.31							
	6.52	2p->5	0.5	1.99	0.24	6.49	2->4	-0.58	2.14	0.18
${}^1E_{2g}$		3->4	0.47			3->5	-0.45			
		2p->4	0.4			2,3->5,4	0.29			
		3->5	-0.31			2->5	0.25			
	7.09	1->5	0.57	0.65	0.27	6.95	1->5	0.56	0.93	0.27
${}^1E_{2g}$		2p->5	0.35			2->5	-0.36			
		2p->6	-0.28			2->4	0.28			
		2p->4	0.25			2->6	-0.25			
		2p,3->5	-0.23			3->5	0.23			
		2p,3->5,4	0.22							
	7.09	1->4	0.45	0.45	0.19	7.02	1->4	0.42	0.51	0.21
	2p->5	-0.44			2->5	-0.42				
	3->5,4	-0.36			3->5,4	0.37				
	3->6	0.3			2,3->4	0.31				

^aDominant configurations (referred to orbitals in Figure 7. X) and corresponding CI coefficients. States nomenclature refers to the ideal D6h symmetry of benzene (Table 7.X). Experimental transition energies are reported, when available, in parentheses.

Using a minimal active space only, our ss-CASPT2/CASSFC(10,8)/6-31G* calculations achieve excellent agreement with the most widely accepted experimental GS \rightarrow ${}^1B_{2u}$ transition values (i.e., 4.45 eV for GUA and 4.35 for CREO) in comparison with previous theoretical studies at the TD-DFT B3LYP/6-31G(d,p) level, which report the first vertical excitation energy of GUA and CREO at 4.70-5.29 eV and 4.78-5.18eV, respectively [31, 32]. In agreement with the molecular orbitals in the active space in our calculations, the ground state (GS) \rightarrow ${}^1B_{2u}$ transition of GUA involves primarily a HOMO-LUMO+1 (MOs 3 \rightarrow 5, Figure 7. 3) excitation with a minor contribution from HOMO-1 to LUMO (MOs 2 \rightarrow 4), which is a classical $\pi\pi^*$ excitation of a ${}^1B_{2u}$ state [32]

In the energy region below the GUA ionization limit (\sim 7.93 eV), where the ${}^1B_{2u}$, ${}^1B_{1u}$, and the pairs of ${}^1E_{1u}$ and ${}^1E_{2g}$ states are detected (Table 7.1), red-shifted absorptions and almost no splitting of the ${}^1E_{1u}$ and ${}^1E_{2g}$ doublets is observed for GUAI and CREO, with respect to the benzene absorption energies. HVER, on the contrary, displays a somewhat higher splitting than the substituted phenols studied here, but this splitting is not as dramatic as in phenol or anisole. Furthermore, contrary to benzene, the ${}^1B_{2u}$, ${}^1B_{1u}$ and the 1E_2 states are dipole allowed. The methoxy and methyl functionalization of benzene in the lignin monomers induces an increase in the TDMs of almost all transitions, including those from the GS and the ${}^1B_{2u}$ as also observed for phenol. The oscillator strength of the GUA ${}^1B_{2u}$ is \sim 0.01 (|TDM|=0.32 au) in comparison with an oscillator strength of \sim 0.02 (|TDM|=0.37 au) and \sim 0.02 (|TDM|=0.39 au) for CREO and HVER, respectively.

As we have shown, the energy region below the ionisation limit of phenol, anisole, and the lignin model compounds have a similar nature as benzene and are described by the transitions of the π -orbitals of the monomer chromophores. The region above the ionisation limit, however, shares some common features but signature bands of each molecule define it (Table 7.S1 and Figure 7. 4). First, we summarise the excited state transitions above the ionisation limit previously reported for benzene and phenol [28]. The broad band found experimentally in benzene at 4.60 eV [36] from the ${}^1B_{2u}$ state has been assigned to the ${}^1B_{2u}\rightarrow 2{}^1E_{2g}$ bright

transitions (TMD=0.64) and the symmetry forbidden ${}^1B_{2u} \rightarrow 1D$ transition (the first doubly excited state). Previous reference RAS(0,0/6,6/2,12) calculations of benzene, the ${}^1B_{2u} \rightarrow 2{}^1E_{2g}$ and ${}^1B_{2u} \rightarrow 1D$ transitions were found near ~ 5.02 eV (40490 cm^{-1}). For phenol, similarly, the absorption experimentally found around 9.0 eV can be associated with the primary double 1D excitation and the two single-electron $2{}^1E_{2g}$ excitations, which can be located at 9.03 and 9.1-9.15 eV, respectively. As in benzene, the ${}^1B_{2u} \rightarrow 2{}^1E_{2g}$ states (37300-37700 cm^{-1}) are bright and display large TDMs in the range of 0.46-0.73 au. The electronic structures of benzene and phenol in the 10-11 eV range are characterised by other doubly excited states (denoted as 2D and 3D), but two bands arising from single excitations involving the oxygen lone pair (O) of the hydroxyl group, denoted ${}^1O_{lp}$ and ${}^2O_{lp}$ are signature fingerprints of phenol in the far-UV. The ${}^1B_{2u} \rightarrow {}^1O_{lp}$ and ${}^1B_{2u} \rightarrow {}^2O_{lp}$ transitions are bright (TMD=0.50-0.42) and clearly defined, and are found near 45400 and 48400 cm^{-1} respectively.

The two-colour 2DUV spectra of anisole in Figure 7. 4 obtained from our ss-PT2/CAS(8,7) calculations of anisole show that the ${}^1B_{2u} \rightarrow 1D$ transition lies far from the two single-electron ${}^1B_{2u} \rightarrow 2{}^1E_{2g}$ transitions and can be found at 37590 and 32100-32340 cm^{-1} , respectively. Moreover, we notice that the $2{}^1E_{2g}$ states in anisole are highly mixed with single excitations involving the oxygen lone pair of the methoxy group. As in phenol, two closely lying signature bands arising from single excitations involving the oxygen lone pair of the methoxy group are bright, and the ${}^1B_{2u} \rightarrow {}^1O_{lp}$ and ${}^1B_{2u} \rightarrow {}^2O_{lp}$ transitions can be found at 42990-43550 cm^{-1} . Moreover, the electronic structure of anisole in the 44000-48000 cm^{-1} range is characterised by the presence of other doubly excited states. Thus, both phenol and anisole exhibit the ${}^1O_{lp}$ and ${}^2O_{lp}$ bands, and whereas the ${}^1B_{2u} \rightarrow {}^1O_{lp}$ and ${}^1B_{2u} \rightarrow {}^2O_{lp}$ transition in phenol remain in the far-UV, in anisole they are red-shifted by up to 2000-5000 cm^{-1} and lie close together.

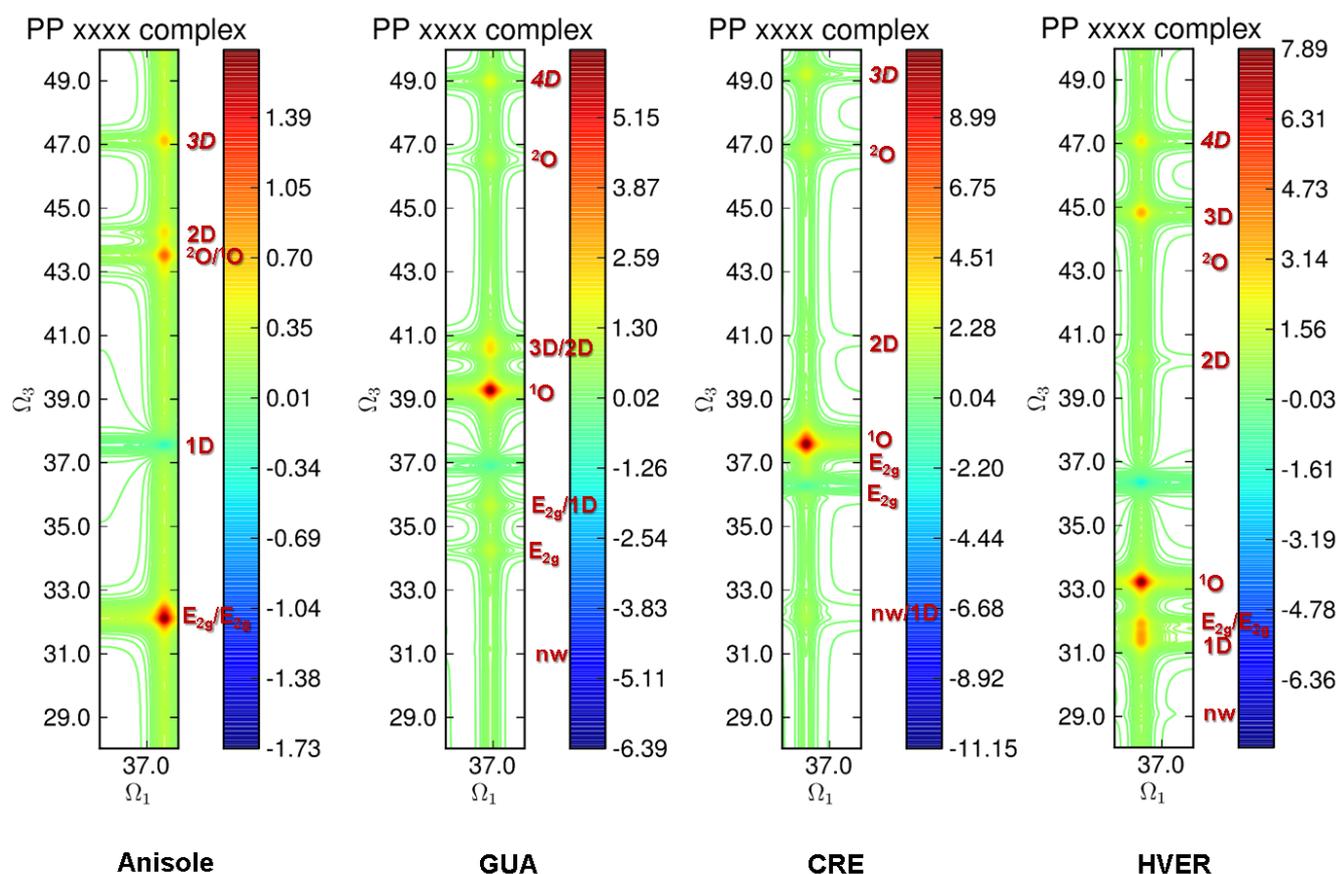


Figure 7. 4. Simulated two-colour 2DUV spectra of anisole, guaiacol, creosol, and homoveratrole at the ss-PT2/CAS(8,7) level. Calculations at the CAS level of theory are shown, highlighting the different positions of the ESAs (positive peaks) above the ionisation energy. Excited-state assignment corresponds to the labels reported in Table 7.S2.

The electronic structure of guaiacol (Table 7.S1 and Figure 7. 4) shares some features of the mono-substituted compounds, but its electronic structure beyond the ionisation energy is greatly affected by the presence of both hydroxyl and methoxy groups that participate in intramolecular H-bonding. A new state just above the ionisation potential (i.e., *nw*) is associated with a single-electron excitation involving the delocalized oxygen lone pair described by an anti-bonding n-orbital (i.e. Oa), as illustrated in Figure 7. 3. This state is also present in the electronic structure of CRE and HVER and can be found at 8.48 and 8.11 eV, respectively, and at 8.44 eV in GUA. The 2^1E_{2g} transitions, as in anisole, are bright and are highly mixed with single excitations involving the same delocalized oxygen lone pair in the anti-bonding Oa orbital. As illustrated in Figure 7. 4, the $^1B_{2u} \rightarrow 2^1E_{2g}$ and $^1B_{2u} \rightarrow 1D$ transition in GUA and HVER are less split than anisole, except for CRE, whose $^1B_{2u} \rightarrow 1D$ is red-shifted below the $^1B_{2u} \rightarrow 2^1E_{2g}$ transitions.

In all lignin monomers, the $^1B_{2u} \rightarrow ^1O_{1p}$ and $^1B_{2u} \rightarrow ^2O_{1p}$ transitions can be associated with single excitations involving the delocalized oxygen lone pair, described by the bonding orbital (Ob) in Figure 7. 3, which closely resembles the orbital involved in the $^1O_{1p}$ and $^2O_{1p}$ states in phenol and anisole. Contrary to anisole,

these states split significantly by up to 8000 in GUA and by up to 10000 cm^{-1} in CRE and HVER. Furthermore, in the three lignin monomers, the $^2\text{O}_{\text{ip}}$ band splits in two transitions arising from the same O_{b} orbital. Only in CRE, splitting of the $^2\text{O}_{\text{ip}}$ band is observed. Notably, the $^1\text{O}_{\text{ip}}$ and $^2\text{O}_{\text{ip}}$ bands are red-shifted in comparison with CRE and GUA. Furthermore, the $^2\text{O}_{\text{ip}}$ states in HVER become dark ($\text{TMD}=0.03$ and 0.04), probably due to the substitution of the $-\text{OH}$ by another $-\text{OMe}$ group, thus, removing the intramolecular H-bond present in GUA and CRE molecules.

3.2 CAS vs RAS 2D electronic spectra of lignin monomers

In the previous section, we have shown how the $\text{ss-CASPT2/CASSFC}(10,8)/6\text{-}31\text{G}^*$ calculations can provide transition energies of the lignin monomers studied here with absolute errors below 0.13 eV, compared to their experimental reference values. However, when dealing with di-chromophore aggregates (i.e., lignin dimers), the use of complete active spaces for each monomer results in unaffordable calculations, due to all the possible configuration state functions (CSFs) in the larger active space. A $\text{CAS}(14,13)$ calculation, such as that for benzene-phenol dimer, requires 736,164 CSFs and such calculation would require a high computational cost. As we will discuss in the next section, we computed a lignin dimer with CREO and HVER monomers using a minimal active space $\text{SA10-CAS}(20,16)$ computation (14'158144 CSFs), which is computationally very demanding. A large number of excited states more significant than a SA10 have to be included in the state-average procedure in order to provide a better picture of the transition energies and dipole moments for simulations of 2DUV spectra, which render these computations unaffordable and unpractical.

In order to find a compromise between the computation cost and the accuracy of the computed transition energies, we now explore the possibility of reducing the size of the multiconfigurational problem. As previously demonstrated in benzene-phenol monomers and dimers, the number of the CFSs can be significantly decreased by restriction of the active space. The wave function analysis of GUA, CREO, and HVER monomer calculations with the $\text{CAS}(10,8)$ active space, and the benzene and phenol ($\text{CAS}(6,6)$ $\text{CAS}(7,6)$, respectively) calculations [28] show that the frontier π -orbitals (the three highest occupied π -orbitals and the three lowest occupied π -orbitals) and n-orbitals for each oxygen lone pair are essential for the description of the singly and doubly excited state manifolds of the chromophores. Therefore, decreasing the size of the active space would not provide accurate results. However, by restricting the total number of simultaneously excited electrons within the active orbitals, the RAS technique reduces the computational cost by reducing the number of CSFs, while maintaining the total of active electrons. Here, we have employed the $\text{ss-CASPT2/RASSCF}(4,5|0,0|4,3)/6\text{-}31\text{G}^*$ level scheme. As summarised in Table S7-2 and Figure 7. 5, the differences between the $\text{RASSCF}(4,5|0,0|4,3)$ (901 CSFs) and $\text{CAS}(10,8)$ (1176 CSFs) transition energies of the monomers do not significantly affect their two-colour 2DUV spectra. Furthermore, the $\text{RASSCF}(4,5|0,0|4,3)$ schemes yield $\text{GS}\rightarrow\text{L}_{\text{b}}$ transition energies that are within 0.01 eV of their corresponding CAS calculations.

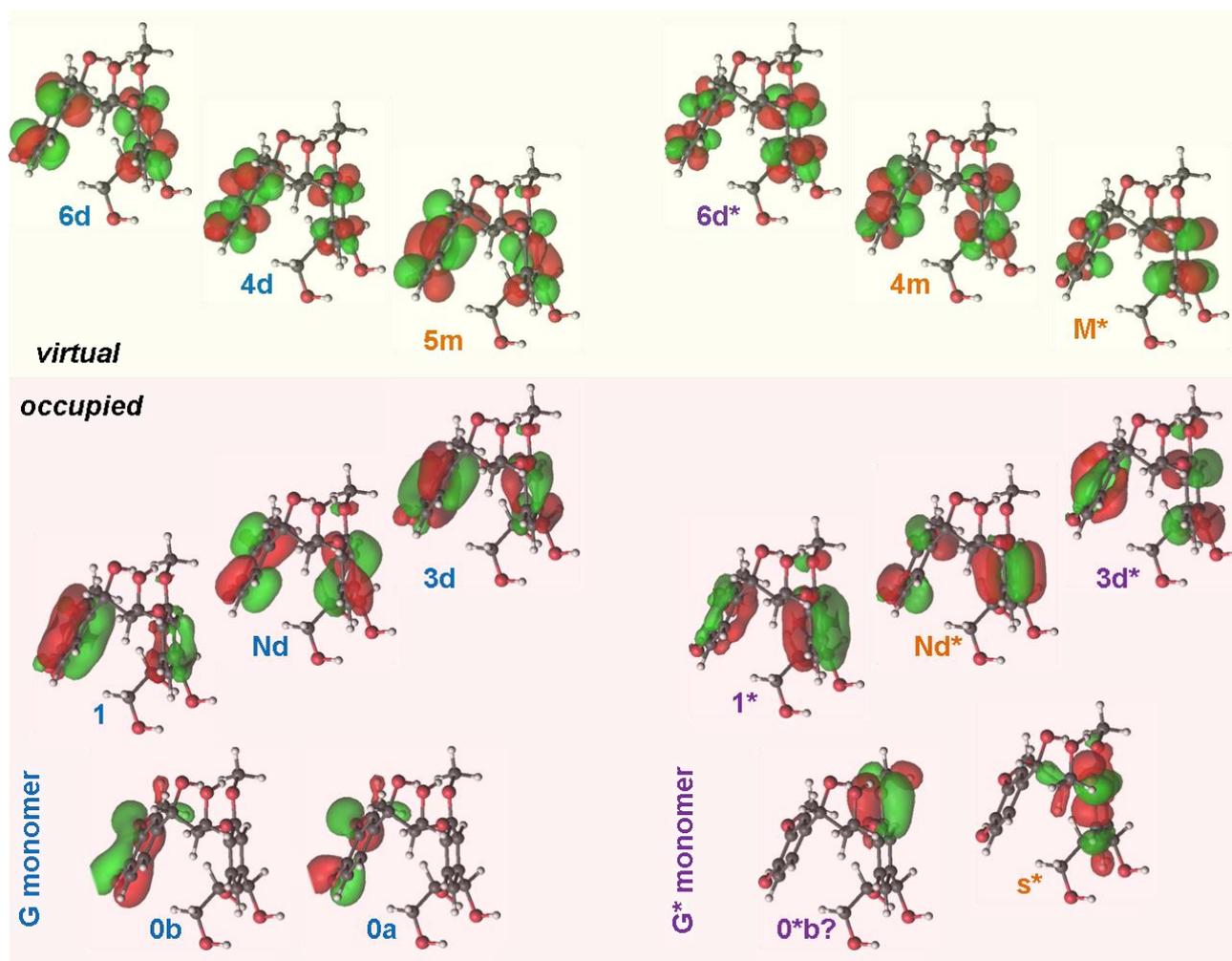


Figure 7. 6. Active space orbitals for the SA50- RASSCF(4,10|0,0|4,6) calculations of the stacked lignin dimer in solution.

As we mentioned in the previous section, we achieved an SA10 ss-CASPT2/CAS(20,16) reference calculation of the unstacked lignin dimer in solution. Because only a small number of states can be sampled for the lignin conformer using the CAS calculations, the computed excited states at the SA10 ss-CASPT2/CAS(20,16) reference calculation of the unstacked lignin dimers in solution will be only used to assess the method accuracy by comparison of the ${}^1B_{2u}$ states of each chromophore with the RAS calculation. Excellent agreement between the RAS and CAS schemes for the computation of the $GS \rightarrow {}^1B_{2u}$ (further on referred to as $GS \rightarrow L_b$) transitions in each monomer is achieved (Table S7-3 and S7-4). The first excited state in the unstacked lignin dimer corresponds to the $GS \rightarrow L_b$ transition in lignin monomer HVER at 4.51 eV at the CAS level vs 4.50 eV at the RAS level. The second excited state, the $GS \rightarrow L_b$ transition in lignin monomer CRE is found at 4.55 eV at both levels of theory [29]. As it has previously reported for phenol and benzene [28], the use of the RAS scheme does not affect the energies of the covalent states (i.e., L_b states), whereas an over stabilization is expected in the energies of the ionic and doubly excited states since the σ - π polarization effect is too significant to be corrected at the perturbation level.

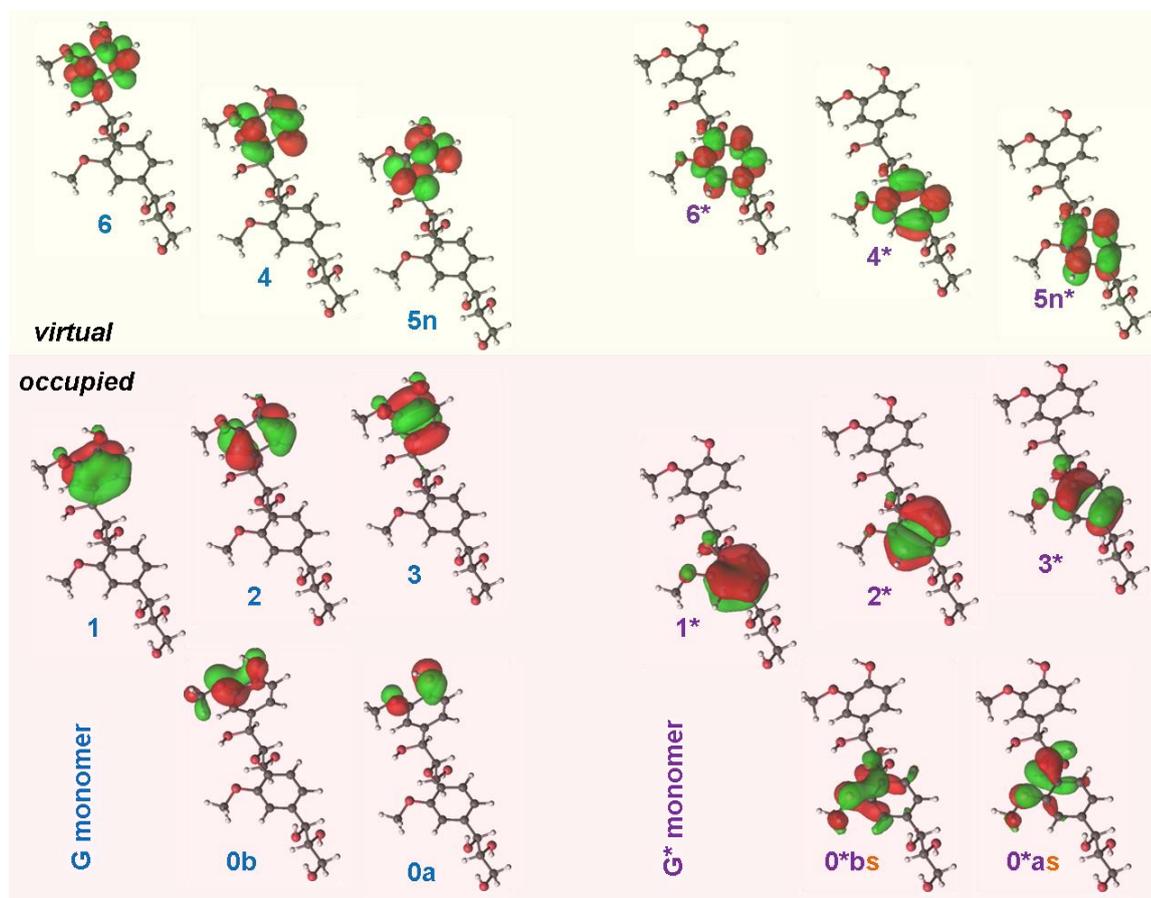


Figure 7. 7. Active space orbitals for the SA50- RASSCF(4,10|0,0|4,6) calculations of the unstacked lignin dimer in solution.

Figure 7.8 and 7.9, respectively, show a comparison between the one colour (UV-UV) and the two-colour (UV-vis) 2D spectra obtained with the explicit SOS//QM/MM approach for the unstacked and stacked lignin dimer, that contain clear spectroscopic fingerprints characterising the different stacking states. The first excited states of CRE and HVER representing the L_b of each ring were selected as targets for the incoming pulses. Probing was performed in two spectral windows: NUV (34500-37500 cm⁻¹, Figure 7.8) and visible (15000-29000 cm⁻¹, Figure 7.9). Probing in this regions resolves correlated transitions and weak quartic couplings in coupled aggregates, whereas probing in the visible region covers the spectral region near the ionisation potential and permits the collection of background-free signals of charge-transfer(CT) transitions in coupled chromophore aggregates[27]. The SOS//QM/MM spectrum of the unstacked dimer with xxxx polarisation (one colour, Figure 7.8A and Figure 7. S7-6 with enhanced dipole) shows the diagonal (negative) bleach signal of the GS→L_b transition in the HVER monomer at ~36300 cm⁻¹ (TDM=0.23). The closely lying diagonal bleach signal (blue-shifted by ~400 cm⁻¹ of the L_b of G') of the GS→L_b transition in the CRE monomer is stronger and more visible, at ~36700 cm⁻¹ (TDM=0.27). Moreover, the diagonal (positive) signals of the L_bCRE →L_bHVER and L_bHVER →L_bCRE (2L_b) transitions are found at ~36500 and ~36100, respectively. However, as shown previously for the cyclic Cysteine-Phenylalanine-Tyrosine-Cysteine (CYTC) tetrapeptide [22], the 2D spectra shows that the excitation from the L_b to the 2L_b states are covered by the intense L_b

bleach (negative) of each monomer. Thus, as expected, the unstacked dimer excited states are mostly localised on the single chromophores. Furthermore, we notice that the TDMs of the Lbs of CRE and HVER in the lignin dimer and solution decrease in comparison with their values in the gas phase (0.37 and 0.39, respectively). The CAS and RAS schemes yield TDMs that are comparable for both for the monomers in the gas phase and the lignin dimers in solution (Figure 7.s S7-1-4). Thus, as signalled previous by Nenov et al. [27], the decrease of signal intensity is an environmental effect related to the conformation and the solvent arrangement.

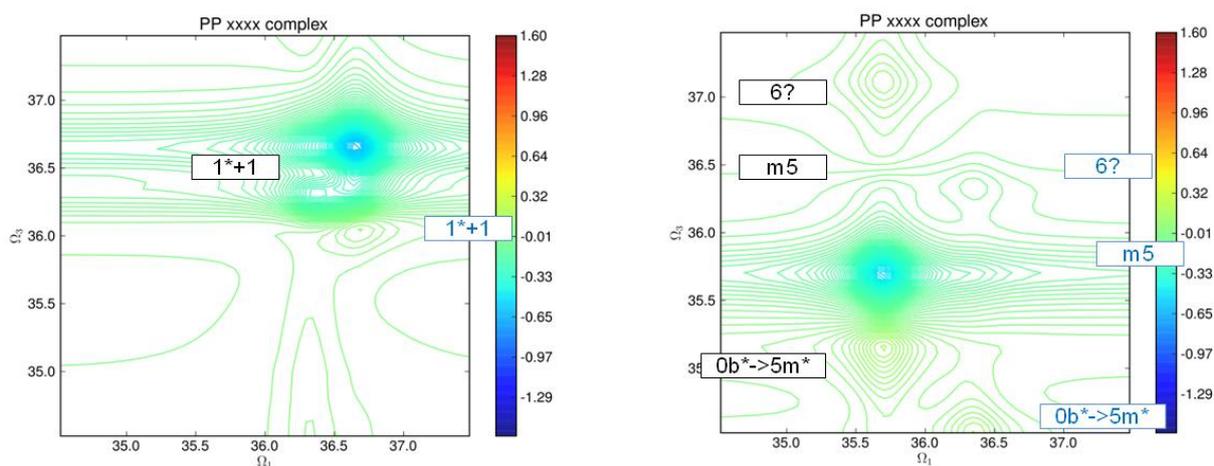


Figure 7. 8. Simulated one-colour 2DUV spectra of unstacked (A) and stacked (B) lignin dimer in solution. Positions of the ESAs and GS bleaching relative to the GS are labelled according to Table 7.1.

Due to the chromophore-chromophore interaction in the stacked configuration, the $GS \rightarrow L_b$ transitions of HVER and CRE are coupled at $\sim 35700 \text{ cm}^{-1}$ (TDM=0.27), red-shifted by about 700 and 800 cm^{-1} in comparison with the $GS \rightarrow L_b$ transitions in CREO and HVER monomers. This observation agrees with the formation of a J-aggregate with a head-to-tail arrangement, which exhibit red-shifted absorptions compared to their free monomers according to the exciton theory [37, 38]. Moreover, the corresponding second pump is found blue-shifted from the combined $GS \rightarrow L_{b\text{HVER, CRE}}$ at $\sim 36400 \text{ cm}^{-1}$ and its signal is much weaker than the $GS \rightarrow L_{b\text{HVER, CRE}}$, as expected (TDM=0.14).

Furthermore, two positive signals at ~ 35200 and 37100 cm^{-1} are revealed. Wavefunction analysis indicates that the later corresponds to what seems a new state due to the mixing of the orbitals in the coupled dimer. The former corresponds to a $n\pi^*$ transition from one of the oxygen lone pair of the HVER ring ($Ob \rightarrow 5$), also present in the lowest optimised HVER monomer in the gas phase, with both methoxy groups in plane with the benzene ring at ~ 33200 . This $n\pi^*$ transition is not present in the one-colour region of the unstacked dimer due to a conformational and solvent arrangement of HVER, where both methoxy and ether bonds groups are found out of plane of the phenolic rings, and this gives rise to shift in this transition, found at ~ 25000 , in the two-colour region, as illustrated in Figure 7. 9 This band assignment is confirmed by analysing the reference

monomer 2D spectra of both HVER monomers (in and out-of-plane) and can be found in Figure S7-5 for reference.

Excitations from local excited states cause off-diagonal contributions to the 2DUV spectrum to high-lying singly excited states, whose molecular orbitals involved in the excitations out of the Lb states can be localized (L), delocalized (D), or they can have a charge transfer character (CT) depending on the permanent dipole moment of the final excited state [22]. From the comparison of the two-colour 2D spectra in Figure 7. 9, the differences between SOS//QM/MM spectra of the stacked and unstacked dimers of lignin is deducible. The most determinant signature of coupled chromophore aggregates is the presence of charge transfer (CT) state signals in the visible probing region of the 2D electronic spectra, as previously demonstrated for the coupling of side groups of a CFYC tetrapeptide in solution [26-28]. In the selected probing window between 15000 and 24000 cm^{-1} , only ESAs involving the 5,6 and X states of the monomer chromophores can be found in the unstacked dimer (See Table S7-2 and S7-3 for label assignment). Wave function analysis shows that new signals that are dominant in this regime arise from CTs or new transitions due to orbital mixing and delocalisation. Upon stacking, CT states can be reached from the Lb of both chromophores, while in non-interacting aggregates the CTs are darks and often inaccessible. When stacked, the overlap of the π orbitals may stabilise the energies by more than 1000 cm^{-1} , thus enhancing the TDM both from the GS and the Lb bands [27]. Therefore, CT states are strongly dependent on the relative geometric orientation, distance of the chromophores, and orbital overlap.

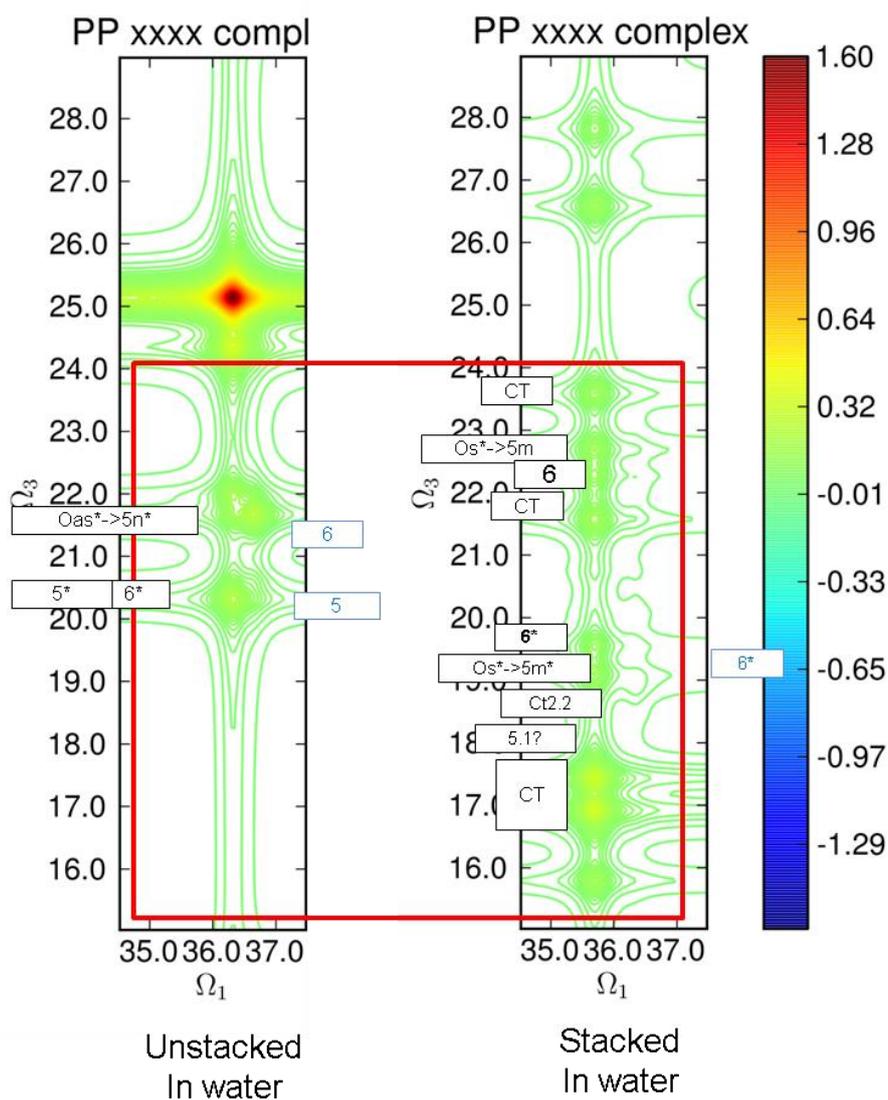


Figure 7. 9. Simulated two-colour 2DUV spectra of unstacked (A) and stacked (B) lignin dimer in solution. The signals are labelled, according to Table 7.X.

Here, we can track spectroscopic fingerprint between two aggregation states of lignin. The results obtained with the SOS//QM/MM methodology show that 2D electronic spectroscopy has excellent potential as a novel diagnostic tool to track the folding state of polymers and DNA. Moreover, they can be used as a guide for conducting 2D spectroscopy experiments of several macromolecules including proteins, and RNA/DNA and, as shown here, aromatic motifs present in complex materials, such as lignin.

4. Conclusions

In this work, we studied the electronic structure and conformations in water of a guaiacyl lignin dimer employing 2D electronic spectroscopy in the UV-VIS region, using the SOS//QM/MM methodology at the CASSCF and RASSCF level of theory. First, vertical excitation energies and 2D-UV/VIS excitation spectra

were obtained for all lignin model monomers (guaiacol, creosol, homoveratrole) in gas phase, including all π and lone-pair electrons and orbitals in the active space. They were further compared against anisole and previously studied benzene and phenol, to aid in excited state assignments. The excited states below the ionization potential correspond to classical $\pi\pi^*$ excitations of a ${}^1B_{2u}$ state. In the energy region below ionisation potential, red-shifted absorptions and almost no splitting of the ${}^1E_{1u}$ and ${}^1E_{2g}$ doublets is observed for GUA and CREO, relative to the benzene absorption energies. HVER, on the contrary, displays a somewhat higher splitting than the substituted phenols studied here, but this splitting is not as dramatic as in phenol or anisole. Furthermore, the ${}^1B_{2u}$, ${}^1B_{1u}$ and the 1E_2 states are dipole allowed. The methoxy and methyl functionalization of benzene in the lignin monomers induces an increase in the TDMs of almost all transitions, including those from the GS and the ${}^1B_{2u}$ as also observed for phenol. Furthermore, we found an excellent agreement between the 2DEF obtained with CAS and RAS schemes for the monomers.

We studied the 1-colour and 2-colour 2D electronic spectra of stacked and unstacked lignin conformers in solution and gas-phase taken from MD simulations of the lignin dimer studied in Chapter 4. Excellent agreement between the RAS and CAS schemes for the computation of GS \rightarrow L_b transitions in each monomer is achieved. The CAS and RAS schemes yield TMDs that are comparable for both for the monomers in the gas phase and the lignin dimers in solution. Thus, as established previously by Nenov et al. [27], the decrease of signal intensity is an environmental effect related to the conformation and the solvent arrangement. Moreover, The 1-colour 2D spectra of the unstacked dimer show that the L_b states of each lignin monomer are uncoupled and localised in each chromophore.

Moreover, the 2D spectra of the stacked dimer shows that the L_b states are coupled and delocalized in the rings. Furthermore, the 2-colour 2DUV spectra show that charge transfer (CT) states become bright upon stacking, allowing to distinguish between two conformers with different degrees of stacking. Thus, our results provide evidence that the use of the SOS//QM/MM approach employed here is suitable for studying 2DUV-vis spectra of different degrees of aggregation of lignin dimers.

REFERENCES

1. Parham, R.A. and R.L. Gray, *Formation and Structure of Wood*, in *The Chemistry of Solid Wood*. 1984, American Chemical Society. p. 3-56.
2. Upton, B.M. and A.M. Kasko, *Strategies for the Conversion of Lignin to High-Value Polymeric Materials: Review and Perspective*. *Chemical Reviews*, 2016. **116**(4): p. 2275-2306.
3. Phillips, M., *The Chemistry of Lignin*. *Chemical Reviews*, 1934. **14**(1): p. 103-170.
4. Bruijninx, P.C.A., R. Rinaldi, and B.M. Weckhuysen, *Unlocking the potential of a sleeping giant: lignins as sustainable raw materials for renewable fuels, chemicals and materials*. *Green Chemistry*, 2015. **17**(11): p. 4860-4861.

5. Sen, S., S. Patil, and D.S. Argyropoulos, *Thermal properties of lignin in copolymers, blends, and composites: a review*. Green Chemistry, 2015. **17**(11): p. 4862-4887.
6. Liu, W.-J., H. Jiang, and H.-Q. Yu, *Thermochemical conversion of lignin to functional materials: a review and future directions*. Green Chemistry, 2015. **17**(11): p. 4888-4907.
7. Wang, Y.-Y., Y.-r. Chen, and S. Sarkanen, *Path to plastics composed of ligninsulphonates (lignosulfonates)*. Green Chemistry, 2015. **17**(11): p. 5069-5078.
8. Yunpu, W., et al., *Review of microwave-assisted lignin conversion for renewable fuels and chemicals*. Journal of Analytical and Applied Pyrolysis, 2016. **119**: p. 104-113.
9. Horwath, W., *Chapter 12 - Carbon Cycling: The Dynamics and Formation of Organic Matter*, in *Soil Microbiology, Ecology and Biochemistry (Fourth Edition)*, E.A. Paul, Editor. 2015, Academic Press: Boston. p. 339-382.
10. ACS. *Molecule of the week*. 2012; Available from: <https://www.acs.org/content/acs/en/molecule-of-the-week/archive/l/lignin.html>.
11. Guadix-Montero, S. and M. Sankar, *Review on Catalytic Cleavage of C–C Inter-unit Linkages in Lignin Model Compounds: Towards Lignin Depolymerisation*. Topics in Catalysis, 2018. **61**(3): p. 183-198.
12. Stark, N.M., D.J. Yelle, and U.P. Agarwal, *4 - Techniques for Characterizing Lignin*, in *Lignin in Polymer Composites*, O. Faruk and M. Sain, Editors. 2016, William Andrew Publishing. p. 49-66.
13. Besombes, S. and K. Mazeau, *Molecular dynamics simulations of a guaiacyl β -O-4 lignin model compound: Examination of intramolecular hydrogen bonding and conformational flexibility*. Biopolymers, 2004. **73**(3): p. 301-315.
14. Stomberg, R. and K. Lundquist, *Stereochemistry of lignin structures of the β -O-4 type*, in *Nordic Pulp & Paper Research Journal*. 1994. p. 37.
15. Besombes, S., et al., *Molecular Modeling of Lignin β -O-4 Model Compounds. Comparative Study of the Computed and Experimental Conformational Properties for a Guaiacyl β -O-4 Dimer*, in *Holzforschung*. 2003. p. 266.
16. Aguilera-Segura, S.M., et al., *Synergistic sorption of mixed solvents in wood cell walls: experimental and theoretical approach*. Macromolecular symposia, 2019. **386**(1): p. 1900022.
17. Selig, U., et al., *Coherent two-dimensional ultraviolet spectroscopy in fully noncollinear geometry*. Optics Letters, 2010. **35**(24): p. 4178-4180.
18. Tseng, C.-h., S. Matsika, and T.C. Weinacht, *Two-Dimensional Ultrafast Fourier Transform Spectroscopy in the Deep Ultraviolet*. Optics Express, 2009. **17**(21): p. 18788-18793.
19. Tseng, C.-h., et al., *Two-Dimensional Fourier Transform Spectroscopy of Adenine and Uracil Using Shaped Ultrafast Laser Pulses in the Deep UV*. The Journal of Physical Chemistry A, 2012. **116**(11): p. 2654-2661.
20. West, B.A., J.M. Womick, and A.M. Moran, *Probing Ultrafast Dynamics in Adenine With Mid-UV Four-Wave Mixing Spectroscopies*. The Journal of Physical Chemistry A, 2011. **115**(31): p. 8630-8637.

21. West, B.A. and A.M. Moran, *Two-Dimensional Electronic Spectroscopy in the Ultraviolet Wavelength Range*. The Journal of Physical Chemistry Letters, 2012. **3**(18): p. 2575-2581.
22. Rivalta, I., et al., *Ab initio simulations of two-dimensional electronic spectra: The SOS//QM/MM approach*. International Journal of Quantum Chemistry, 2014. **114**(2): p. 85-93.
23. Mukamel, S., R. Oszwaldowski, and D. Abramavicius, *Sum-over-states versus quasiparticle pictures of coherent correlation spectroscopy of excitons in semiconductors: Femtosecond analogs of multidimensional NMR*. Physical Review B, 2007. **75**(24): p. 245305.
24. Abramavicius, D., et al., *Coherent Multidimensional Optical Spectroscopy of Excitons in Molecular Aggregates; Quasiparticle versus Supermolecule Perspectives*. Chemical Reviews, 2009. **109**(6): p. 2350-2408.
25. Segarra-Martí, J., et al., *Two-dimensional electronic spectroscopy as a tool for tracking molecular conformations in DNA/RNA aggregates*. Faraday Discussions, 2018. **207**(0): p. 233-250.
26. Nenov, A., et al., *Disentangling Peptide Configurations via Two-Dimensional Electronic Spectroscopy: Ab Initio Simulations Beyond the Frenkel Exciton Hamiltonian*. The Journal of Physical Chemistry Letters, 2014. **5**(4): p. 767-771.
27. Nenov, A., et al., *Tracking Conformational Dynamics of Polypeptides by Nonlinear Electronic Spectroscopy of Aromatic Residues: A First-Principles Simulation Study*. ChemPhysChem, 2014. **15**(15): p. 3282-3290.
28. Nenov, A., et al., *Two-Dimensional Electronic Spectroscopy of Benzene, Phenol, and Their Dimer: An Efficient First-Principles Simulation Protocol*. Journal of Chemical Theory and Computation, 2015. **11**(8): p. 3755-3771.
29. Longarte, A., et al., *IR/UV and UV/UV double-resonance study of guaiacol and eugenol dimers*. The Journal of Chemical Physics, 2005. **122**(16): p. 164304.
30. Wu, R. and B. Brutschy, *Study on the structure and intra- and intermolecular hydrogen bonding of 2-methoxyphenol-(H₂O)_n (n=1,2)*. Chemical Physics Letters, 2004. **390**(1): p. 272-278.
31. Ruiz-Santoyo, J.A., et al., *Intramolecular structure and dynamics of mequinol and guaiacol in the gas phase: Rotationally resolved electronic spectra of their S₁ states*. The Journal of Chemical Physics, 2015. **143**(9): p. 094301.
32. Dean, J.C., et al., *Ultraviolet spectroscopy of fundamental lignin subunits: Guaiacol, 4-methylguaiacol, syringol, and 4-methylsyringol*. The Journal of Chemical Physics, 2013. **139**(14): p. 144313.
33. Dearden, J.C. and W.F. Forbes, *LIGHT ABSORPTION STUDIES: PART XV. THE ULTRAVIOLET ABSORPTION SPECTRA OF ANISOLES*. Canadian Journal of Chemistry, 1959. **37**(8): p. 1305-1314.
34. Dearden, J.C. and W.F. Forbes, *LIGHT ABSORPTION STUDIES: PART XIV. THE ULTRAVIOLET ABSORPTION SPECTRA OF PHENOLS*. Canadian Journal of Chemistry, 1959. **37**(8): p. 1294-1304.
35. Fujimaki, E., et al., *Autoionization-detected infrared spectroscopy of intramolecular hydrogen bonds in aromatic cations. I. Principle and application to fluorophenol and methoxyphenol*. The Journal of Chemical Physics, 1999. **110**(9): p. 4238-4247.
36. Nakashima, N., et al., *Nanosecond laser photolysis of the benzene monomer and eximer*. The Journal of Chemical Physics, 1980. **72**(4): p. 2226-2230.

37. Deng, Y., et al., *H- and J-Aggregation of Fluorene-Based Chromophores*. The Journal of Physical Chemistry B, 2014. **118**(49): p. 14536-14545.
38. Hestand, N.J. and F.C. Spano, *Expanded Theory of H- and J-Molecular Aggregates: The Effects of Vibronic Coupling and Intermolecular Charge Transfer*. Chemical Reviews, 2018. **118**(15): p. 7069-7163.

Chapter 8. Summary of the conclusions and perspectives

Classical molecular dynamics simulations (MD), quantum chemistry calculations (QM-DFT), and combined QM/MM (BOMD/MD) approaches were employed to describe at the atomistic level the effect of the solvents on the structure and dynamics of truncated polymer models of lignocellulose compounds - cellulose, hemicellulose, and lignin.

1. First, the structures of solvent mixtures, their thermodynamic properties and their global chemical hardness, which can be used as estimate of the solvents reactivity, were described at solvents equilibrium dynamics, obtained with MD simulations. A non-linear behaviour with the increase of the organic component in the aqueous mixture was established. The organic components are ethanol, acetonitrile and isopropanol at 0, 25, 50, 75 and 100 wt% organic cosolvent concentrations.

1.1 The solvent structures experience significant changes at all the concentrations. Acetonitrile is not an H-bond donor, and it is inferred that the presence of water and the formation of H-bonds helps it to reduce the strong dipole-dipole interactions while preserving the microheterogeneity of the mixture.

1.2 The proposed by us simple method for the averaged, dynamic, chemical hardness computations revealed that the water-ethanol mixture becomes softer with increasing ethanol content up to 75%, whereas water-acetonitrile mixtures are nearly equally soft at 50 wt% and 75wt% water-acetonitrile mixtures and the pure acetonitrile is even softer.

1.3 As a general trend, a softening of the water solvent by introducing the organic cosolvents is obtained with the only exception for the isopropanol-water mixture at $X = 0.75$. This trend correlates in overall with the deduced from the hydrogen bonds (HB) analysis maximum of water and the organic solvent interactions at 75% mixtures.

1.4 The proposed method for dynamic chemical hardness computations could be further explored to rationalize the behaviour of complex solute systems in mixed solvents, providing dynamic hardness values of the solutes are also assessed.

2. Second, the structure and solvation of the cellulose nanocrystal, lignin dimer and tetramer, and xylose were investigated through solvent accessible surface area (SASA), conformational analysis, H-bonding and H-bond lifetimes, and radial distribution functions. The considered solvents are first pure water, ethanol and acetonitrile, and second, the binary mixtures of water with the two organic solvents at 25, 50 and 75 wt% concentrations. The lignocelluloses-solvent specific interactions were established, depending on the lignocellulose molecular structure and the solvent compositions. In more details:

2.1 The SASA distributions analysis shows the largest cellulose nanocrystal solvent accessible surfaces in the presence of mixed solvents, whereas it has a minimum in the presence of pure acetonitrile, followed by pure ethanol and water. Likewise, lignin displays the highest surface area in the presence of mixed solvents, while water remains a bad solvent for lignin. On the contrary, the hygroscopic nature of xylose results in the lowest surface area in pure acetonitrile and ethanol, whereas a tendency of a more extended surface is observed in pure water and the diluted mixtures.

2.2 The analysis of the hydroxymethyl group conformations in cellulose shows a higher frequency of tg conformers in pure organic solvents indicating lower capacity of hydroxymethyl to form intramolecular H-bonds; following a transition from tg to more solvent exposed gg conformation in the mixed solvents. In the pure water solvent, the solvent exposed gg-hydroxymethyl conformer is predominant.

2.3 The torsion angle analysis of the xylose glycosidic bond has shown that xylose adopts a left-handed threefold conformation in water and diluted mixtures, whereas a twofold conformation is favoured in pure and concentrated organic solvents.

2.4 The analysis of the β -O-4 torsion angle in lignin shows a dihedral conformation distribution near $\theta \sim -90^\circ$ in pure water and diluted ethanol which corresponds to a stacked conformation. In contrast, a more extended T-stacked conformation with a distribution that falls near θ equal to $\sim -150^\circ$ is found in pure and concentrated organic solvents.

2.5 The polymer-solvent correlation analysis using the radial distribution functions show that lignin-water and lignin-ethanol display similar coordination-site trends in pure and mixed solvents, favouring the electrostatic interactions and coordinating preferentially the O_γ sites. In acetonitrile-water mixtures, however, the O_γ site becomes rapidly coordinated by water; therefore, the lignin-acetonitrile coordination is inverted and follows the order $O_\alpha > O_\gamma$, whereas OMe becomes coordinated at diluted acetonitrile content by the methyl group of acetonitrile.

2.6 Likewise, the xylan coordination analysis shows that water and ethanol interact mostly electrostatically with the O2 and O3 sites, whereas apolar interactions between the methyl group of acetonitrile and O2 are preferred. Moreover, the $O_{xylose}-C_{EtOH}$ and the $O_{xylose}-C_{ACN}$ display a coordination-site inversion and follow the order $O3 > O2$ in all water-organic solvent mixtures. On the other hand, the cellulose-water coordination follows the order $O6 > O2 >$ and displays higher number densities in the hydrophilic than the hydrophobic surfaces. Similarly, ethanol follows the order $O6 > O2$ whereas acetonitrile shows higher coordination for O2 than O6. Furthermore, while the water number density is expected to decrease with the addition of organic solvent, the water exclusion in acetonitrile mixtures from the first cellulose solvation shell, especially in the hydrophobic surface, is stronger than

for ethanol mixtures. An increase of the co-solvent phase could yield a more hydrated and more extended water shell around cellulose hydrophilic surfaces, which, enhanced by local microheterogeneity of water-acetonitrile mixtures[1], results in phase-separation when in contact with the different cellulose surfaces. Ethanol mixtures, on the contrary, remain less separated as ethanol-water mixtures are more homogeneous and also share a similar affinity for a specific oxygen site in cellulose.

2.7 Quantum chemistry calculations were used to study the polarization effect of solvents in the partial charges of lignin tetramers and cellulose chains. The Mulliken charges of cellulose oxygens in the presence of water are considerably affected by the solvent in comparison to the gas phase, whereas the effect is significantly reduced in the presence of acetonitrile. The partial charge of the glycosidic oxygen is affected despite the significant lower coordination reported in the RDFs. Similarly, the partial charges of OMe, O β , O γ , and O α oxygens in lignin tetramers show that water strongly interacts with O γ , and O α oxygens, while the others remain less disrupted in the presence of acetonitrile than in the presence of water or ethanol.

2.8 DFT energies of lignin dimers obtained from BOMD/MD dynamics showed that low energy conformations with a low number of intramolecular H-bonds can exist thanks to a stabilization effect of the solvent. The energy gain due to H-bond interactions with the solvent greatly compensates the loss of intramolecular H-bonding, allowing the existence of low energy conformers in the presence of water. Moreover, the lignin-solvent interactions pass through an optimization when adding ethanol to water solvent, as found for the most energetically stable and more solvent-interacting lignin dimer in 75wt%-ethanol-water mixtures. Despite the relatively small lignin models, our results are in a qualitative agreement with recent studies that show an increase of lignin solubility in water-ethanol mixtures[2].

3. Molecular dynamics simulations were employed to describe at the atomistic level the effect of the solvents on the adhesion of lignin to cellulose and to hemicelluloses, while considering the same solvent compositions as for the individual lignocelluloses compounds.

3.1 We followed the adhesion behaviour between xylose and lignin tetramers and the radial distribution analysis show that it is less likely to find lignin coordinating xylan in pure water, as the hemicellulose becomes rapidly hydrated. The addition of the organic phase, however, increases the xylose-lignin coordination and the effect becomes more notorious in the presence of the pure organic solvents.

3.2 The xylose-lignin interactions energies profiles have a non-linear behaviour to the organic solvent composition, in agreement with our observations of the H-bond distributions and the xylose-lignin radial distribution functions. In water, a bad solvent for lignin but with a high affinity for hemicellulose, the xylose-lignin complex displays the weakest binding energy, followed by 50% acetonitrile. The strongest adhesion

between xylose and lignin is observed at 75% ethanol mixture, followed by 25% and 75% acetonitrile mixture, which we have reported to be good solvents for lignin as the SASA distribution of lignin tetramers reaches a maximum. Furthermore, at these concentrations, the presence of the organic phase seems to slow down the xylose conformational changes and, thus, the detachment to lignin.

3.3 Concerning the adsorption behaviour of lignin dimer on the cellulose hydrophobic and hydrophilic surfaces, we found that lignin could be stabilized in the presence of water by stacking interactions between the phenolic groups and the saccharide groups. Lignin dimers were found to preferentially interact with the hydrophilic chains of cellulose, whose less polar aliphatic hydrogen atoms of the glucose rings are exposed.

3.4 Water is the medium that favours mostly the adhesion of lignin on cellulose, followed by ethanol and acetonitrile pure solvents. In the mixed solvents, the cellulose-lignin adhesion is strongly disrupted.

3.5 The specific organization of the solvent molecules in the binary mixtures around the cellulose-lignin complex is identified as the most important factor affecting the cellulose-lignin conformations and their interaction energies. Moreover, the preferential interactions of the two organic components in the presence of water depending on their chemical nature and concentration. The weakest (or no adhesion) of lignin on cellulose is found in the 50% acetonitrile-water solvent, and at 75% ethanol-water solvent. Despite the relatively small cellulose-lignin model, our MD results are in a qualitative agreement with recent studies that show an increase of lignin solubility in water-ethanol mixtures [2].

4. Finally, we studied the electronic structure and conformations in water of a guaiacyl lignin dimer employing 2D electronic spectroscopy in the UV-VIS region, using the SOS//QM/MM methodology at the CASSCF and RASSCF level of theory. First, vertical excitation energies and 2D-UV/VIS excitation spectra were obtained for all lignin model monomers (guaiacol, creosol, homoveratrole) in gas phase, including all π and lone-pair electrons and orbitals in the active space. They were further compared against anisole and previously studied benzene and phenol, to aid in excited state assignments.

4.1 The excited states below the ionization potential correspond to classical $\pi\pi^*$ excitations of a ${}^1B_{2u}$ state. In the energy region below ionisation potential, red-shifted absorptions and almost no splitting of the ${}^1E_{1u}$ and ${}^1E_{2g}$ doublets is observed for GUA and CREO, relative to the benzene absorption energies.

4.2 HVER, on the contrary, displays a somewhat higher splitting than the substituted phenols studied here, but this splitting is not as dramatic as in phenol or anisole. Furthermore, the ${}^1B_{2u}$, ${}^1B_{1u}$ and the 1E_2 states are dipole allowed. The methoxy and methyl functionalization of benzene in the lignin monomers induces an increase in the

TDMs of almost all transitions, including those from the GS and the $^1B_{2u}$ as also observed for phenol. Furthermore, we found an excellent agreement between the 2DEF obtained with CAS and RAS schemes for the monomers.

4.3 We studied the 1-colour and 2-colour 2D electronic spectra of stacked and unstacked lignin conformers in solution and gas-phase taken from MD simulations of the lignin dimer studied in Chapter 4. Excellent agreement between the RAS and CAS schemes for the computation of GS \rightarrow L_b transitions in each monomer is achieved. The CAS and RAS schemes yield TMDs that are comparable for both, the monomers in the gas phase and the lignin dimers in solution. Thus, as established previously by Nenov et al. [3], the decrease of signal intensity is an environmental effect related to the conformation and the solvent arrangement. Moreover, The 1-colour 2D spectra of the unstacked dimer show that the L_b states of each lignin monomer are uncoupled and localised in each chromophore.

4.4 Moreover, the 2D spectra of the stacked dimer shows that the L_b states are coupled and delocalized in the rings. Furthermore, the 2-colour 2DUV spectra show that charge transfer (CT) states become bright upon stacking, allowing to distinguish between two conformers with different degrees of stacking. Thus, our results provide evidence that the use of the SOS//QM/MM approach employed here is suitable for studying 2DUV-vis spectra of different degrees of aggregation of lignin dimers.

Perspectives: This proposed methodologies, results and conclusions from this thesis work open several directions for future investigations of biomass at atomistic levels:

1. The method for dynamic chemical reactivity descriptors, as is the chemical hardness, opens the possibility to explore the principle of *hard-soft-acid-base* relations by comparing the evolutions of the reactivity descriptors a wide range of complex solute molecular systems in solvents, while taking in consideration the dynamics of the solutes and solvent structures. Moreover, this is a simple scheme, and in this first study, we demonstrated its applicability, which opens the possibility for further verification and applications to more complex molecular systems.
2. The understood specific interactions of the here investigated solvents evidences the complexity of the lignocellulose matrices in various solvents, but paved the way for more studies on larger systems or might be used as input knowledge in the machine-learning algorithms. To feed up these algorithms a more extensive calculations on small and simple models of lignocellulosic compounds are needed.
3. The BOMD/MD simulations demonstrated the applicability of the QM/MM scheme and opened the perspectives of increasing the precision of computations on lignocelluloses-explicit solvents by including a less parameter-dependent approach to large size of systems.

4. The generation of MD trajectories of relatively complex systems, as those presented here, and the implementation of QM/MM schemes allow to apply theoretical chemistry not only to averaged reactivity descriptors, but also to various spectroscopic properties in solution. In this way, one can provide additional bases for assignments and interpretation of spectroscopic fingerprints that start to be available from experiments of various lignocellulose compounds in different solvent media.

REFERENCES

1. Aguilera-Segura, S.M., F. Di Renzo, and T. Mineva, *Structures, intermolecular interactions, and chemical hardness of binary water–organic solvents: a molecular dynamics study*. J Mol Model, 2018. **24**: p. 292.
2. Goldmann, W.M., et al., *Solubility and fractionation of Indulin AT kraft lignin in ethanol-water media*. Separation and Purification Technology, 2019. **209**: p. 826-832.
3. Nenov, A., et al., *Tracking Conformational Dynamics of Polypeptides by Nonlinear Electronic Spectroscopy of Aromatic Residues: A First-Principles Simulation Study*. ChemPhysChem, 2014. **15**(15): p. 3282-3290.

Supporting Information

Contents

3.	SI Chapter 3. Structures, intermolecular interactions and chemical hardness of binary water- organic solvents	242
3.1	Thermodynamic properties of pure solvents	242
3.2	Structures of the aqueous binary solvents	244
3.3	Hydrogen bond autocorrelation functions	245
3.4	Dynamic hardness as a function of solvent composition	245
4.	SI Chapter 4. Solvent effect on the structure and dynamics of lignocellulose model compounds	247
4.1	Cellulose-solvent radial distribution functions in mixed solvents	247
4.2	Xylose-solvent radial distribution functions in mixed solvents	249
4.3	Lignin-solvent radial distribution functions in mixed solvents	251
4.4	Polymer-solvent LJ and Coulomb interactions energies	252
5.	SI Chapter 5. Solvent effect on the adhesion between lignocellulose compounds	255
5.1	Cellulose-lignin radial distribution functions	255
5.2	LJ, Coulomb and Total xylan-lignin interaction energies	256
6.	SI Chapter 6. General discussion: the solvent effect on lignocellulose compounds interactions, conformations, and dynamics	257
6.1	Radial distributions of cellulose-solvent and lignin-solvent in cellulose-lignin complex	257
6.2	Number of solvent molecules within 0.7 nm cellulose-solvent shell in cellulose-lignin complex	261
7.	SI Chapter 7. First-Principles Simulations of Two-Dimensional Electronic Spectroscopy of lignin model monomers and their β -O-4 linked dimer	262
7.1	Vertical excitation energies of lignin monomers	262
7.2	Vertical excitation energies of the unstacked lignin dimer	264
7.3	Vertical excitation energies of the stacked lignin dimer	266
7.4	Comparison of 2D spectra of lignin dimer in gas phase and in water	268
7.5	2D spectra of HVER monomer and monomer G' from the unstacked lignin dimer in gas phase	272
		240

7.6	TDM Modified 1-color 2D spectra of unstacked and stacked lignin dimer in solution	273
	References	273

2. SI Chapter 3. Structures, intermolecular interactions and chemical hardness of binary water- organic solvents

3.1 Thermodynamic properties of pure solvents

The simulations for the aqueous mixtures of ethanol, isopropanol, and acetonitrile were performed on 4-nm boxes, using the NPT ensemble at 298.15 K and 1 bar. As follows from the results in **Error! Reference source not found.**, all simulated solvents reproduced closely the reference experimental temperature and pressure conditions, as well as the experimental thermodynamic properties (see main text).

Table S 2-1. Calculated thermodynamic properties from MD simulations: pressure (P), temperature (T), density (ρ), enthalpy of vaporization (ΔH_{vap}), average number of hydrogen bonds per molecule (HB) and diffusivity (D). The number of molecules (N) is reported as well as the experimental values when available. For the references see the reference list in the main paper.

	Water	EtOH	2PrOH	AN
N	2133	657	501	730
^{a)} T (K) ref. 298.15	298.13	300.9	301.9	299.6
^{a)} P (bar) Ref. 1 bar	1.3	1.2	1.4	1.6
ρ (g/L)	986.3	787	786.3	771.8
ρ (g/L) expt.	997 [1, 2]	785 [1, 3]	781.1 [4]	777 [2, 5]
ΔH_{vap} (kJ/mol)	43.64	42.88	45.46	33.63
ΔH_{vap} (kJ/mol) expt.	43.98 [6]	42.32 [6]	45.39 [6]	32.94 [6]
HB	3.54	1.84	1.83	0
D (10^5 cm ² /s)	3.061	0.98	0.49	2.60
D (10^5 cm ² /s) expt.	2.272 [7]	1.075 [3]	0.462 [8]	4.31 [9]

^{a)} The reference temperature and pressure are, respectively, 298.15 K and 1bar.

The ΔH_{vap} was calculated by using equation S1.

$$\Delta H_{\text{vap}} = H_{\text{gas}} - H_{\text{liq}} = (U + PV)_{\text{gas}} - (U + PV)_{\text{liq}} \cong E_{\text{intra}}(g) + RT - E_{\text{intra}}(l) \quad (\text{S1})$$

Assuming the gas behaves as an ideal gas, and that the volume of the liquid phase is negligible in comparison to the volume of the gas phase, the internal energy of liquid can be obtained directly from the potential energy during the simulation and the internal energy of the gas can be computed separately. To this end, one solvent molecule was placed in a cubic box in

vacuum and it was simulated during 20 ns-MD without using cut-off's and with no periodic boundary conditions. Potential energies were computed for each solvent. For all simulated solvent systems, the calculated ΔH_{vap} values agree with the experimental values with errors within 0.4%, 1.3%, 0.2%, and 2.1% for ethanol, isopropanol, and acetonitrile solvents, respectively.

Table S 2-2. Calculated thermodynamic properties from MD simulations: pressure (P), temperature (T), density (ρ), enthalpy of vaporization (ΔH_{vap}), potential energy (E_{pot}) average, and number of hydrogen bonds per molecule (HB).

EtOH				
X_{cosol}	0.25	0.50	0.75	1
N EtOH	201	381	534	657
N Water	1539	973	455	0
N Total	1740	1354	989	657
T (K)	298.62	299.40	300.23	300.9
P (bar)	1.26	1.70	1.35	1.2
ρ (g/L)	941.80	894.53	844.68	787
E_{pot} (kJ/mol)	-40.60	-39.36	-37.26	-33.15
ΔH_{mix} (kJ/mol)	-0.36	-0.48	-0.47	-
HB _{sys}	3.35	3.06	2.60	1.84
HB _{W-W}	2.83	1.93	0.85	-
HB _{EtOH-EtOH}	0.03	0.14	0.50	1.84
HB _{EtOH-W}	0.50	0.98	1.25	-
2PrOH				
X_{cosol}	0.25	0.50	0.75	1
N 2PrOH	153	288	403	501
N Water	1531	960	448	0
N Total	1684	1248	851	501
T (K)	298.87	299.82	300.97	301.9
P (bar)	1.43	1.41	1.20	1.4
ρ (g/L)	941.70	890.83	839.43	786.3
$E_{pot,mix}$ (kJ/mol)	-52.68	-70.46	-100.88	-167.57
ΔH_{mix} (kJ/mol)	-0.24	-0.15	0.08	-
HB _{sys}	3.38	3.12	2.67	1.83
HB _{W-W}	2.97	2.20	1.12	-
HB _{2PrOH-2PrOH}	0.02	0.10	0.39	1.83
HB _{2PrOH-W}	0.39	0.82	1.16	-
AN				
X_{cosol}	0.25	0.50	0.75	1
N AN	224	417	582	730
N Water	1534	950	442	0
N Total	1758	1367	1024	730
T (K)	298.36	298.80	299.11	299.6
P (bar)	0.82	0.71	1.15	1.6
ρ (g/L)	924.33	864.67	813.06	771.8
$E_{pot,mix}$ (kJ/mol)	-44.57	-49.55	-57.11	-70.94
ΔH_{mix} (kJ/mol)	0.38	0.72	0.96	-
HB _{sys}	3.08	2.44	1.50	-
HB _{W-W}	2.83	2.07	1.05	-
HB _{AN-AN}	-	-	-	-

HB_{AN-W}	0.24	0.37	0.45	-
-------------	------	------	------	---

3.2 Structures of the aqueous binary solvents

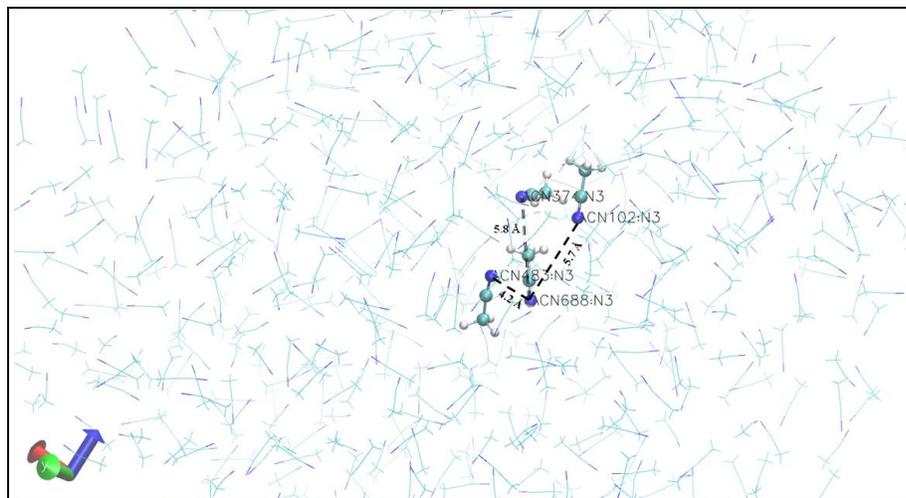


Figure S 2-1. Trajectory snapshot, illustrating the acetonitrile molecular orientations. Ball-and-stick representation is used for a visualization of the parallel and antiparallel intermolecular arrangement. The number of acetonitrile (ACN) residues and the N-N distances are reported in the Figure.

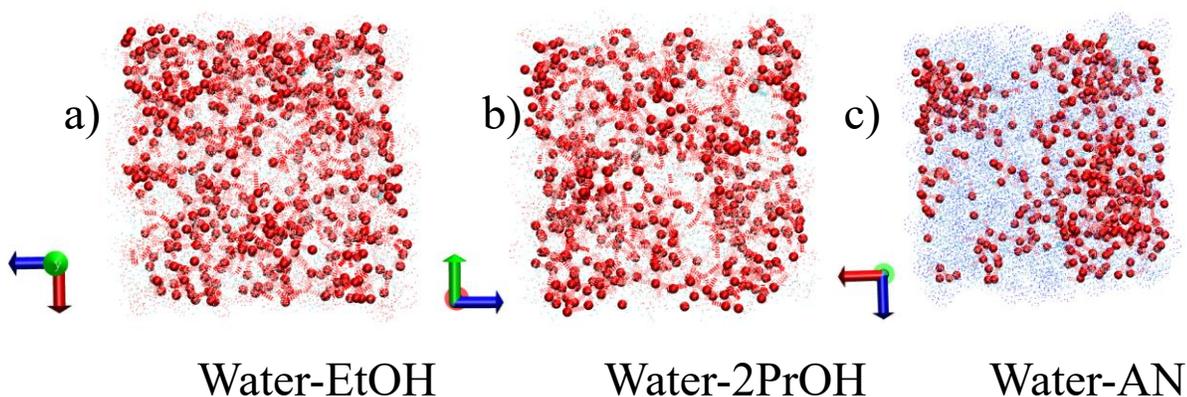


Figure S 2-2. Fig. S1 Trajectory frames for water- a) EtOH, b) 2PrOH, and c) AN mixtures at $X_{0.75}$. Oxygen water molecules are represented by red balls, whereas cosolvent molecules are depicted as a solvent dot mesh. Alcohol mixtures display a uniform distribution of molecules, whereas local phase separation (microheterogeneity) is observed in acetonitrile mixtures.

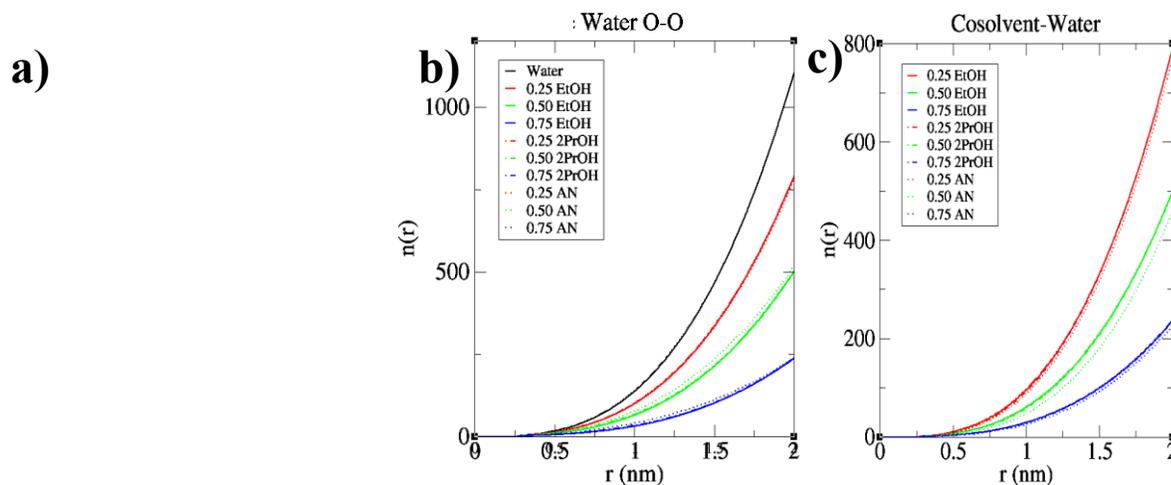


Figure S 2-3. Fig. S2 Cumulative numbers, $n(r)$ in the range of 2nm for the a) cosolvent-cosolvent, b) water-water, and c) cosolvent-water pairs. The water oxygen, ethanol oxygen, isopropanol oxygen, and acetonitrile nitrogen are the reference sites

3.3 Hydrogen bond autocorrelation functions

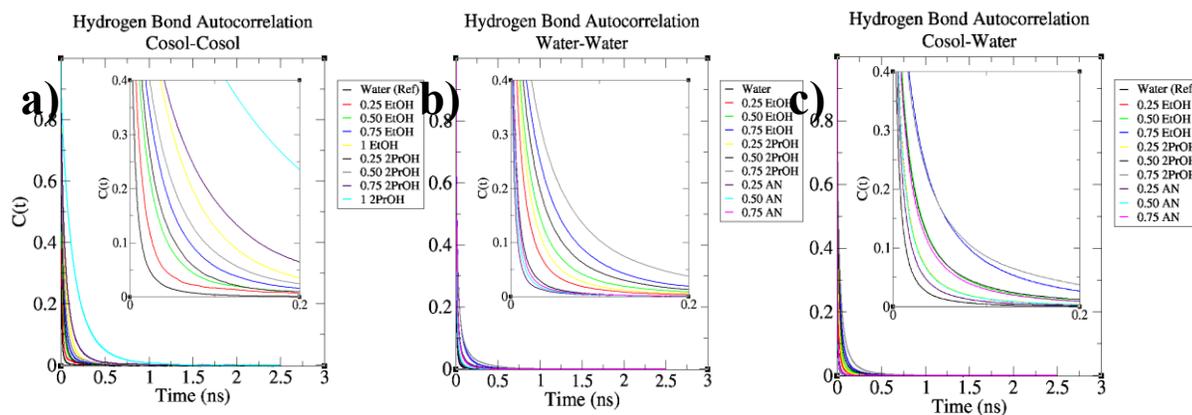


Figure S 2-4. Fig. S3 Hydrogen bond autocorrelation functions, $C(t)$, of the a) cosolvent-cosolvent, b) water-water, and c) cosolvent-water types.

3.4 Dynamic hardness as a function of solvent composition

Table S 2-3. Computed averaged hardness in eV from MD simulations of water-ethanol, water-isopropanol and water-acetonitrile mixtures with different cosolvent fraction X_{cosol} .

X_{cosol}	EtOH	2PrOH	AN
0	8.38	8.38	8.38
0.25	8.34	7.84	7.62
0.50	6.96	7.30	7.16
0.75	6.26	8.94	7.28
1	7.24	8.28	6.76

3. SI Chapter 4. Solvent effect on the structure and dynamics of lignocellulose model compounds

4.1 Cellulose-solvent radial distribution functions in mixed solvents

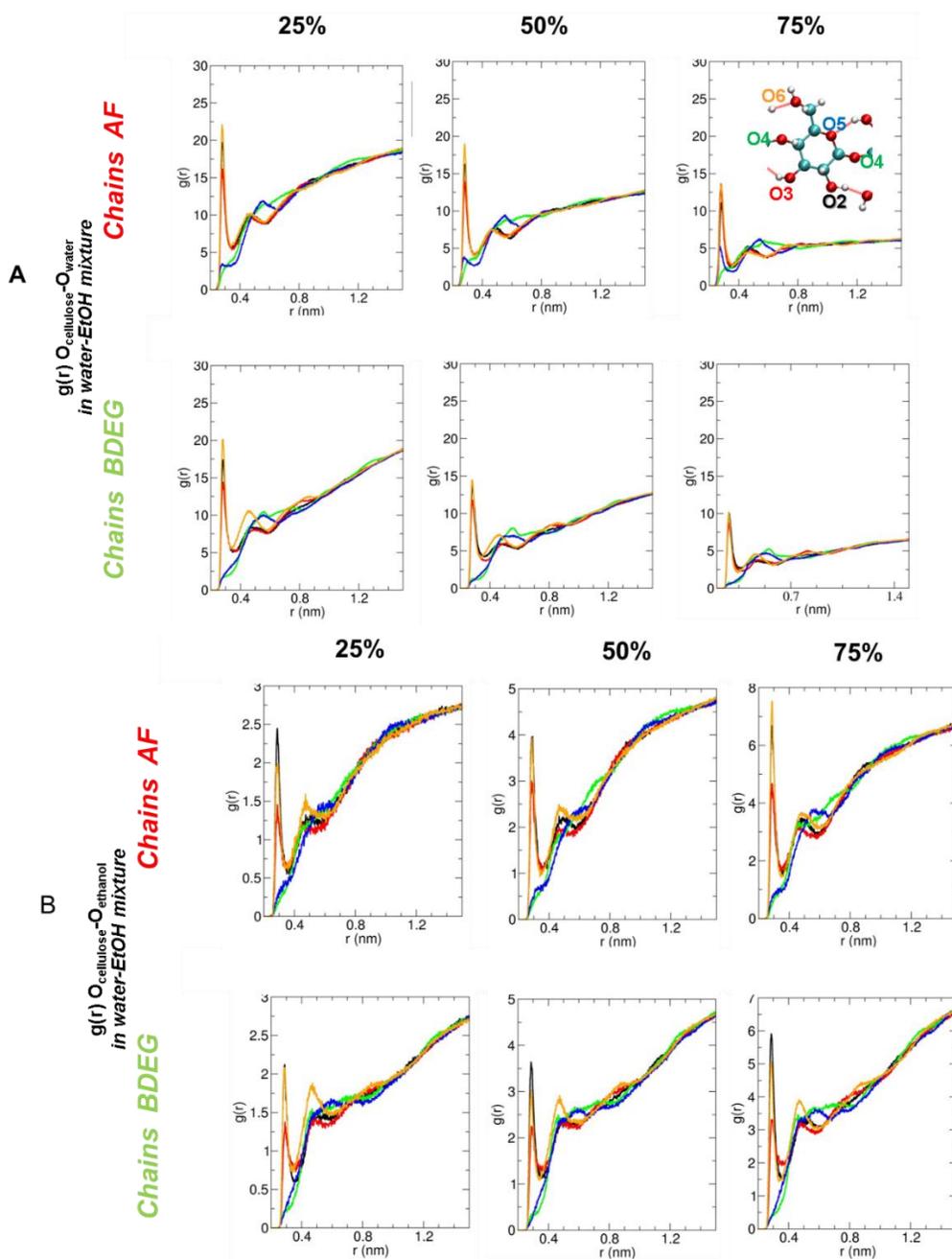


Figure S 3-1. $O_{cell}-O_{water}$ (A) and $O_{cell}-O_{EtOH}$ (B) $g(r)$ s in water-EtOH mixtures (wt%)

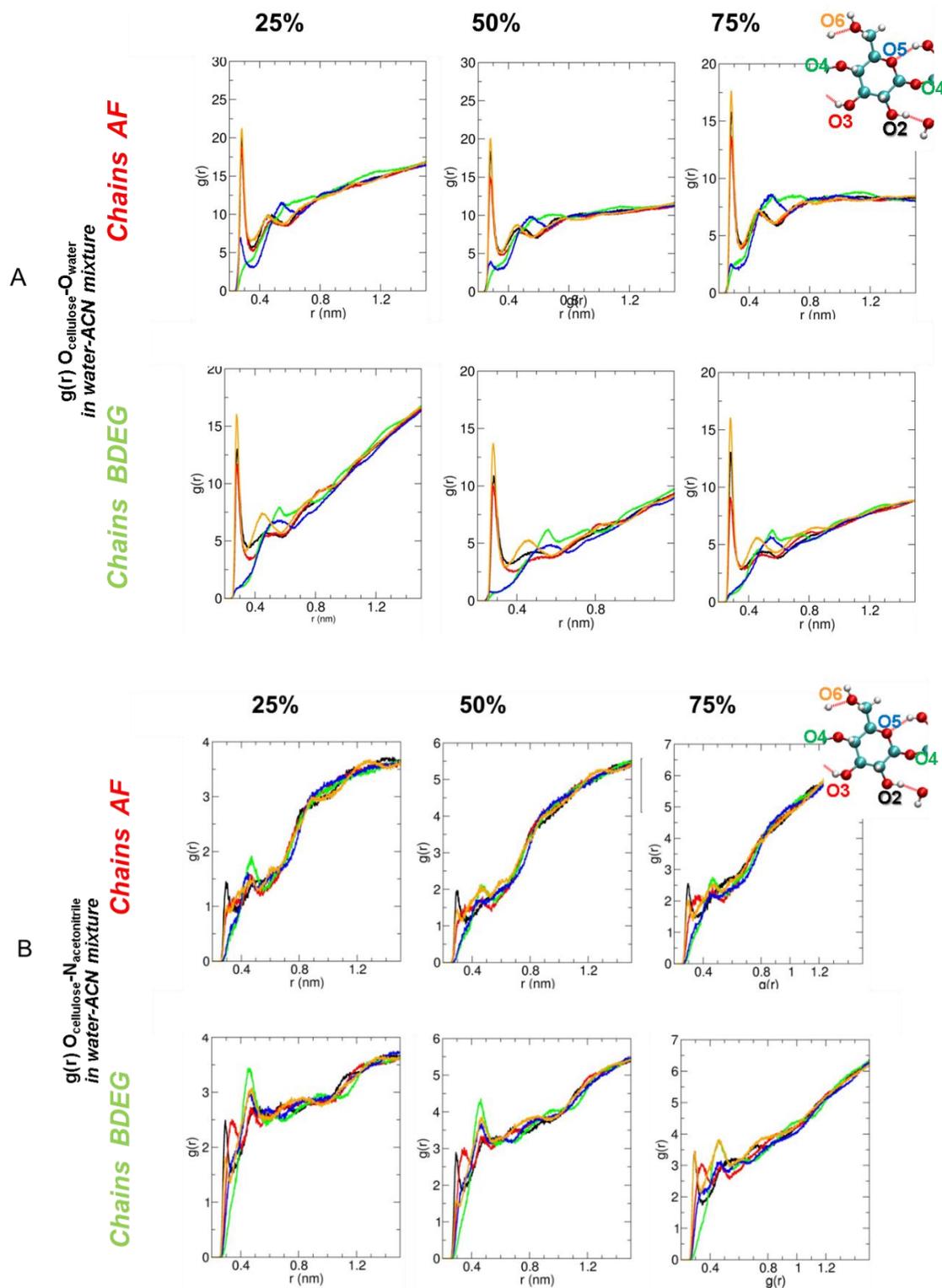
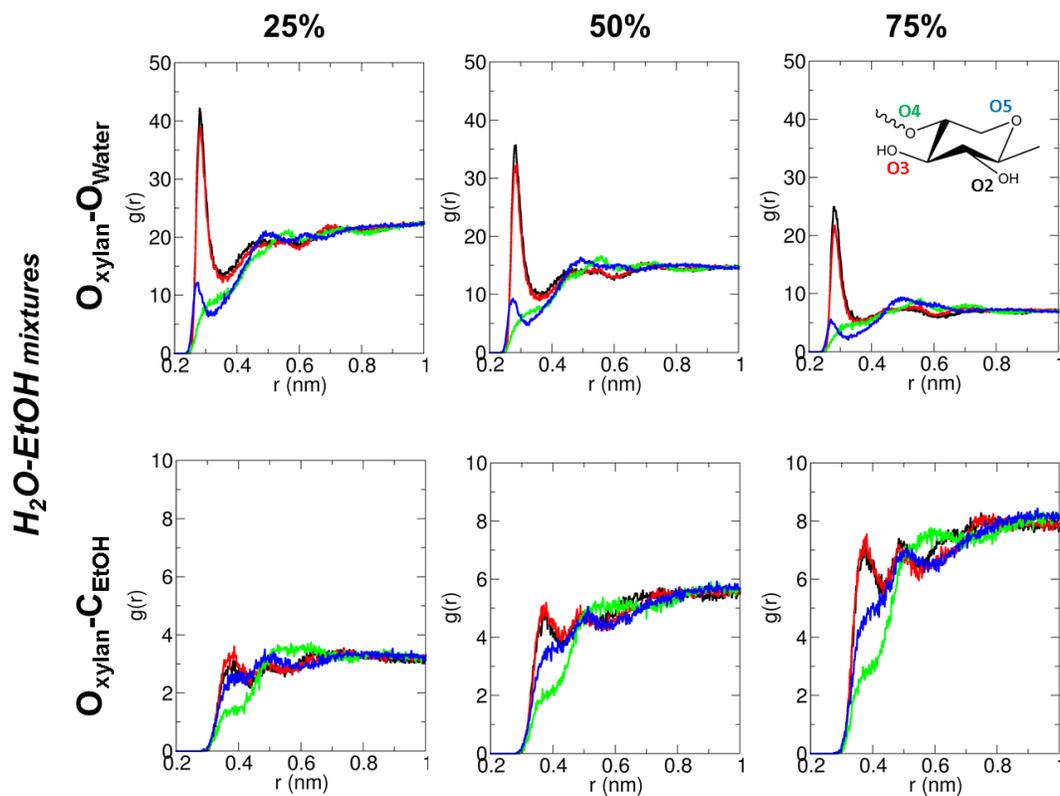


Figure S 3-2. Figure S1. $O_{\text{cell}}-O_{\text{water}}$ (A) and $O_{\text{cell}}-O_{\text{ACN}}$ (B) $g(r)$ s in water-ACN mixtures (wt%)

4.2 Xylose-solvent radial distribution functions in mixed solvents

Figure S 3-3. O_{xylose}-O_{water} and O_{xylose}-C_{EtOH} in water-EtOH mixtures (wt%)

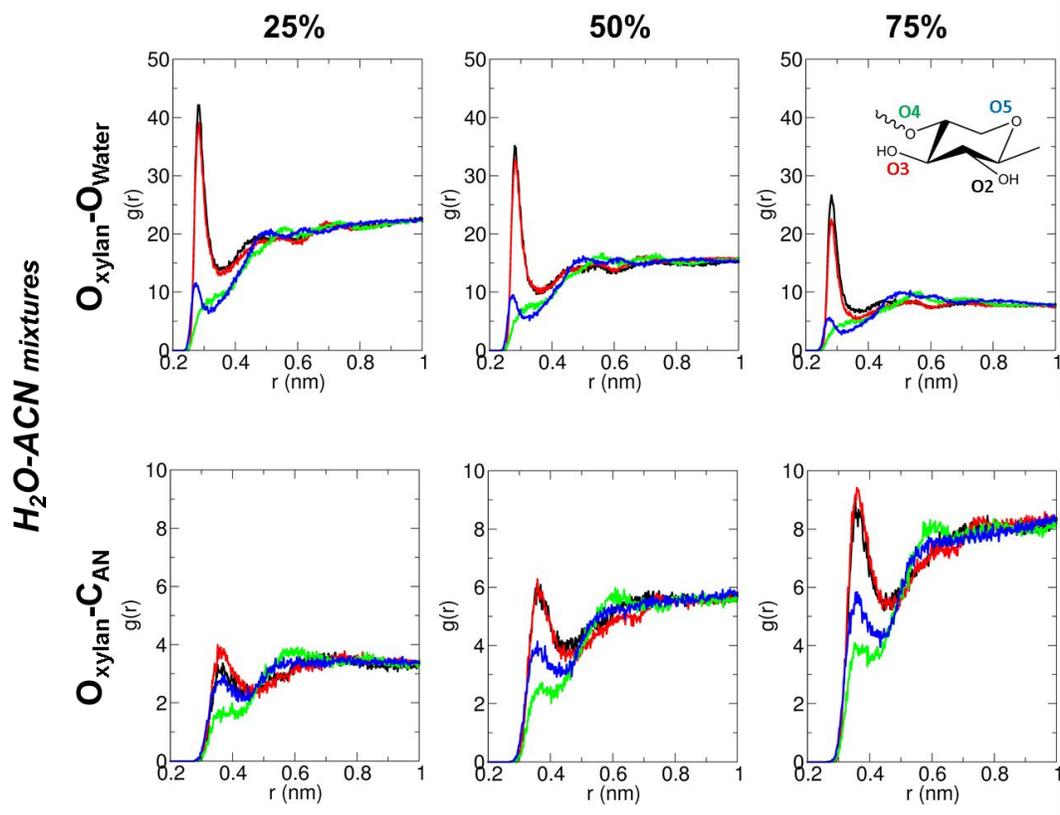
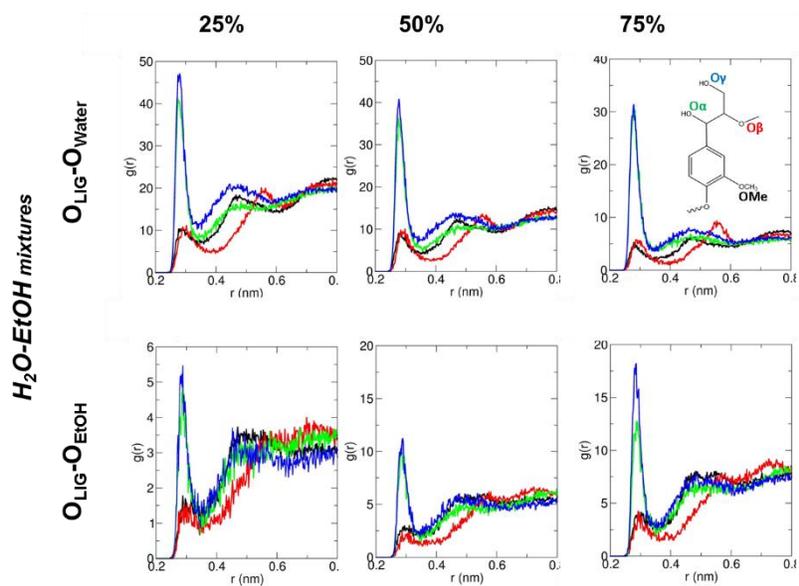
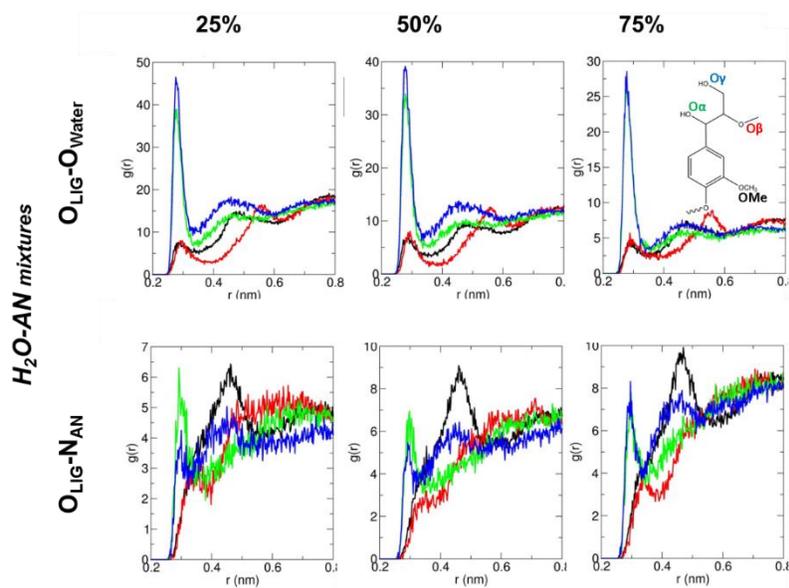


Figure S 3-4. Figure S4. O_{xylose}-O_{water} and O_{xylose}-C_{ACN} in water-ACN mixtures (wt%)

4.3 Lignin-solvent radial distribution functions in mixed solvents

Figure S 3-5. Figure S5. $O_{lig-O_{water}}$ and $O_{lig-O_{EtOH}}$ in water-EtOH mixtures (wt%)Figure S 3-6. $O_{lig-O_{water}}$ and $O_{lig-N_{ACN}}$ in water-ACN mixtures (wt%)

4.4 Polymer-solvent LJ and Coulomb interactions energies

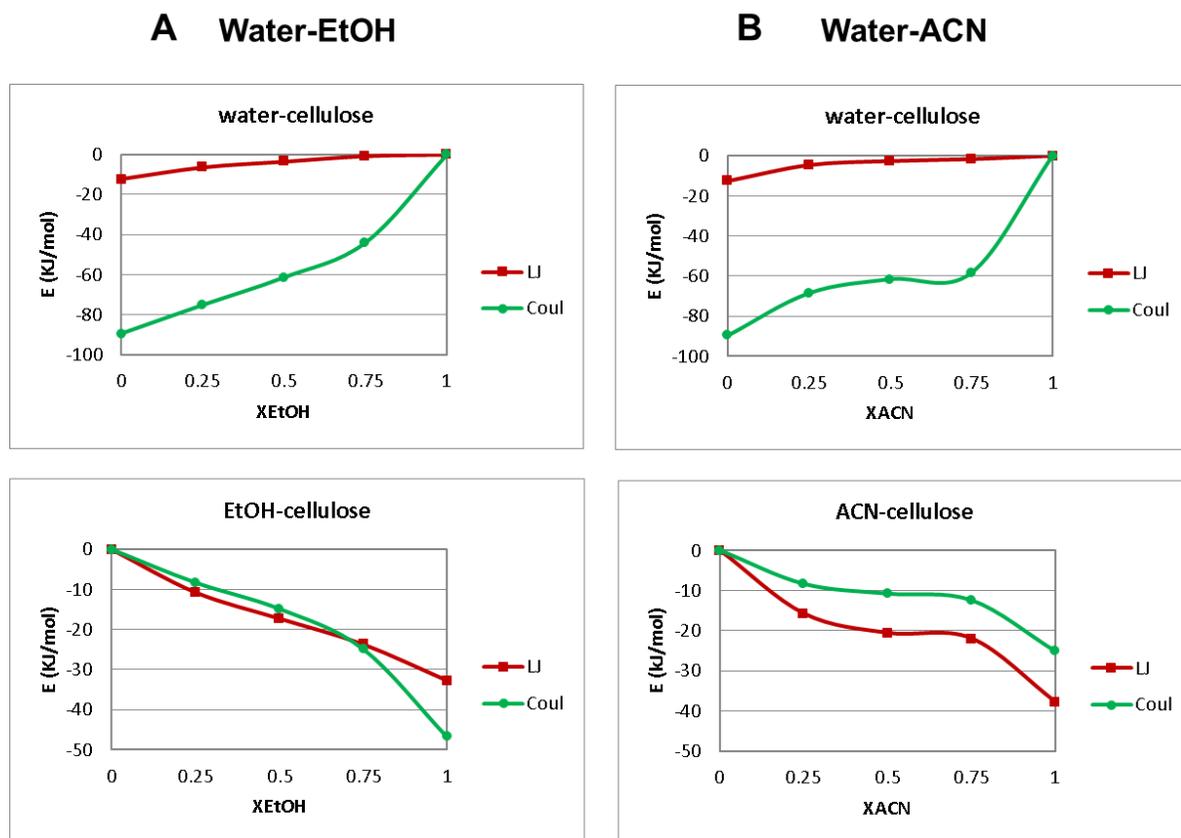


Figure S 3-7. Cellulose-solvent Lennard Jones (LJ) and Coulomb (Coul) energies of interaction as a function of cosolvent fraction

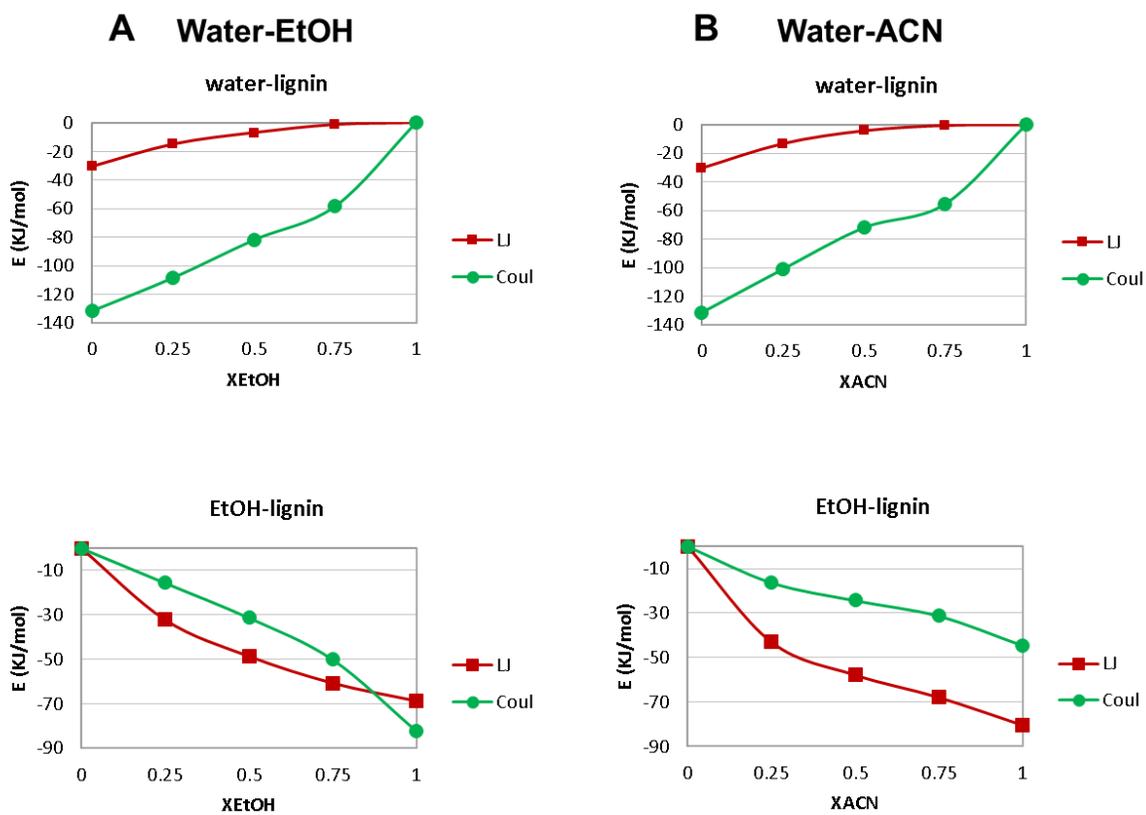


Figure S 3-8. Figure S8. lignin-solvent Lennard Jones (LJ) and Coulomb (Coul) energies of interaction as a function of cosolvent fraction

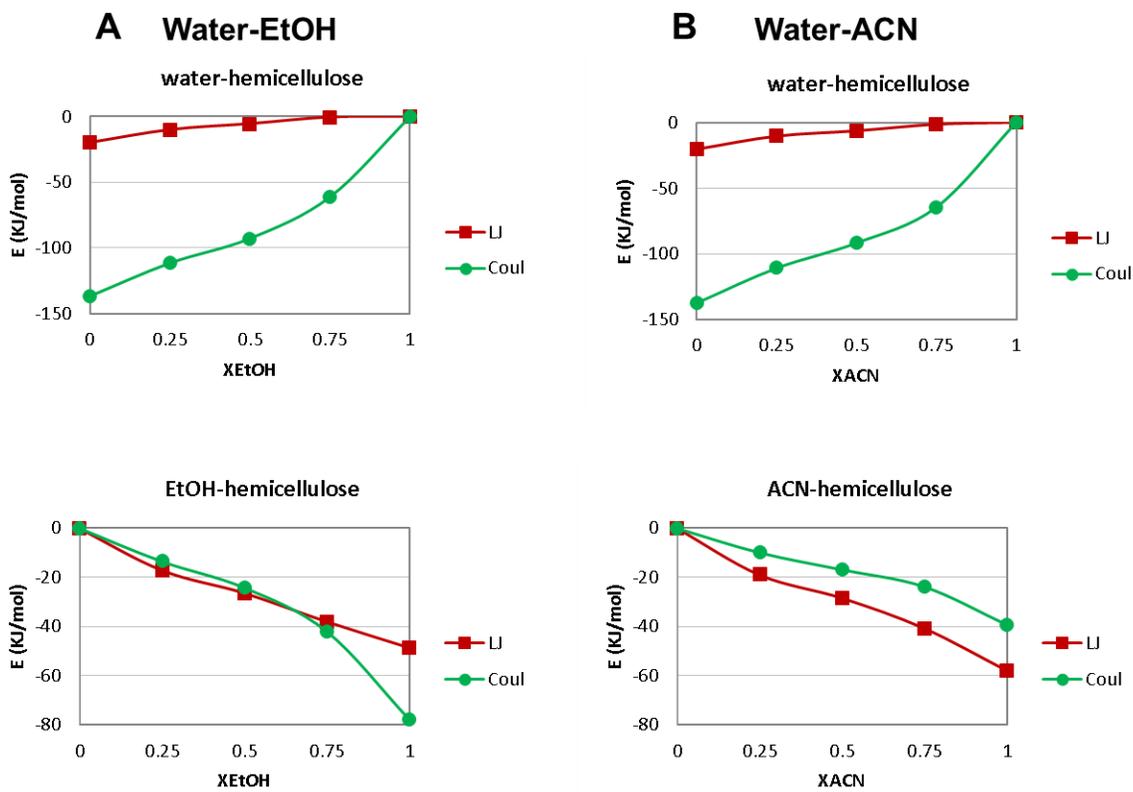


Figure S 3-9. Figure S8. Hemicellulose-solvent Lennard Jones (LJ) and Coulomb (Coul) energies of interaction as a function of cosolvent fraction

4. SI Chapter 5. Solvent effect on the adhesion between lignocellulose compounds

5.1 Cellulose-lignin radial distribution functions

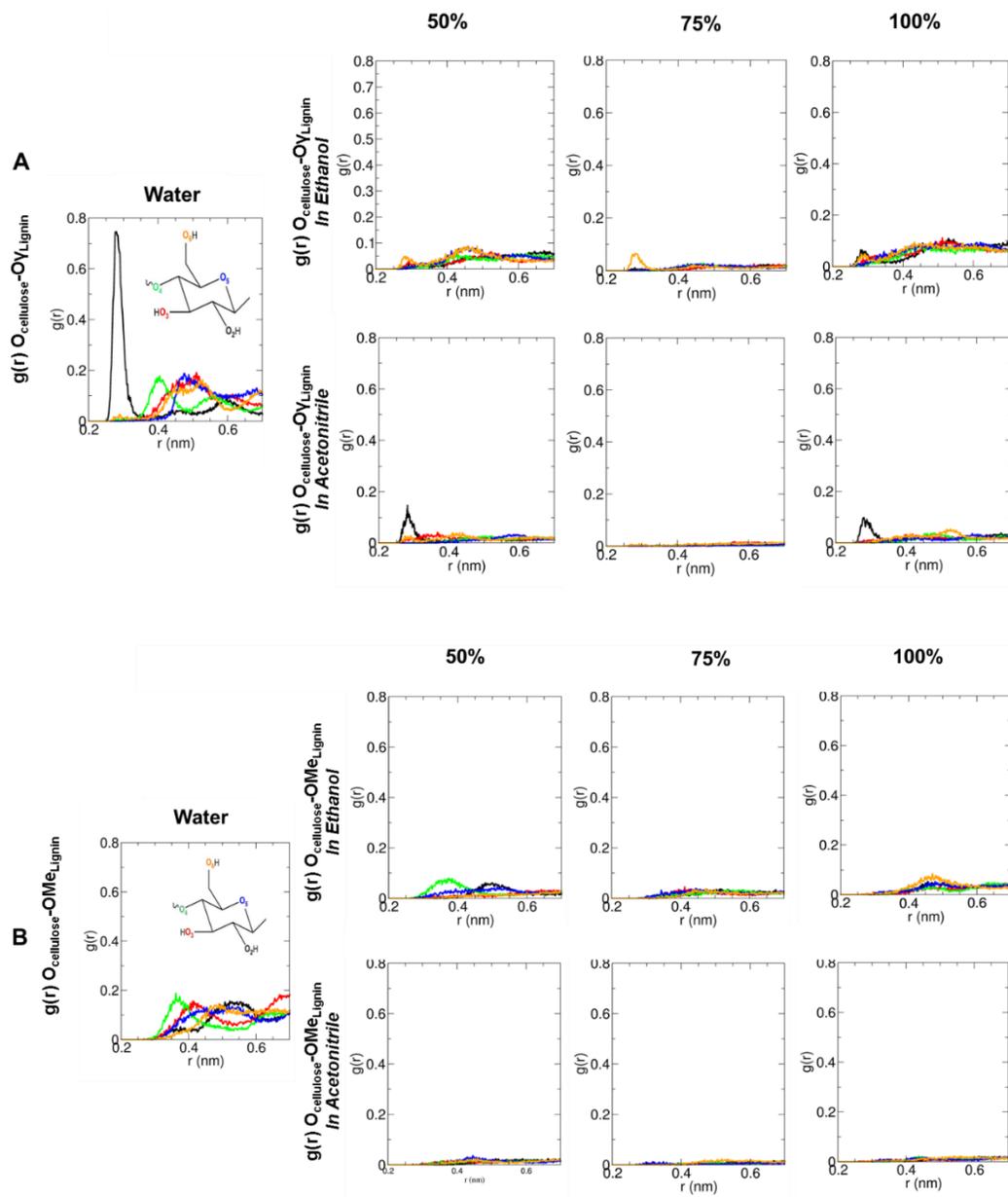


Figure S 4-1. $\text{O}_{\text{cell}}-\text{O}_{\gamma\text{lig}}$ (A) and $\text{O}_{\text{cell}}-\text{OMe}_{\text{lig}}$ (B) in water-EtOH and water-ACN mixtures

5.2 LJ, Coulomb and Total xylan-lignin interaction energies

Table S5-1. Average Lennard-Jones (LJ) , Coulomb (Coul) energies, and Total (LJ+Coul) lignin-cellulose energies in kJ/mol at 2-ns and 10-ns MD.

Average Energies (kJ/mol)						
Solvent	2ns			10 ns		
	LJ	Coul	Total	LJ	Coul	Total
water	-13.5 ± 17.0	-14.8 ± 21.3	-28.3 ± 34.3	-5.0 ± 10.6	-4.2 ± 12.2	-9.2 ± 20.7
0.25 EtOH	-30.4 ± 15.5	-34.6 ± 26.2	-65.0 ± 37.0	-13.1 ± 14.9	-13.8 ± 20.1	-27.0 ± 32.6
0.50 EtOH	-26.3 ± 16.6	-25.5 ± 20.4	-51.8 ± 32.3	-9.2 ± 13.7	-10.6 ± 17.4	-19.8 ± 28.8
0.75 EtOH	-57.3 ± 11.7	-68.3 ± 15.1	-125.6 ± 20.3	-16.6 ± 24.1	-20.7 ± 29.7	-37.3 ± 52.7
EtOH	-44.2 ± 13.5	-39.9 ± 26.6	-84.1 ± 35.8	-13.0 ± 18.3	-17.9 ± 22.9	-30.8 ± 37.4
0.25 ACN	-24.3 ± 21.4	-25.5 ± 21.3	-49.8 ± 38.3	-14.4 ± 16.0	-14.9 ± 18.5	-29.3 ± 30.9
0.50 ACN	-23.5 ± 20.9	-30.1 ± 28.3	-53.6 ± 46.5	-5.1 ± 13.7	-6.5 ± 18.1	-11.6 ± 30.9
0.75 ACN	-34.8 ± 20.8	-35.3 ± 28.0	-70.1 ± 45.1	-13.6 ± 21.4	-13.4 ± 23.2	-27.1 ± 42.0
ACN	-33.4 ± 17.7	-49.4 ± 32.1	-82.8 ± 45.2	-7.2 ± 15.8	-10.4 ± 24.9	-17.5 ± 39.5
Vacuo	-74.1 ± 13.4	-84.6 ± 22.2	-158.7 ± 22.2	-65.9 ± 14.9	-94.3 ± 20.0	-160.1 ± 22.0

5. SI Chapter 6. General discussion: the solvent effect on lignocellulose compounds interactions, conformations, and dynamics

6.1 Radial distributions of cellulose-solvent and lignin-solvent in cellulose-lignin complex

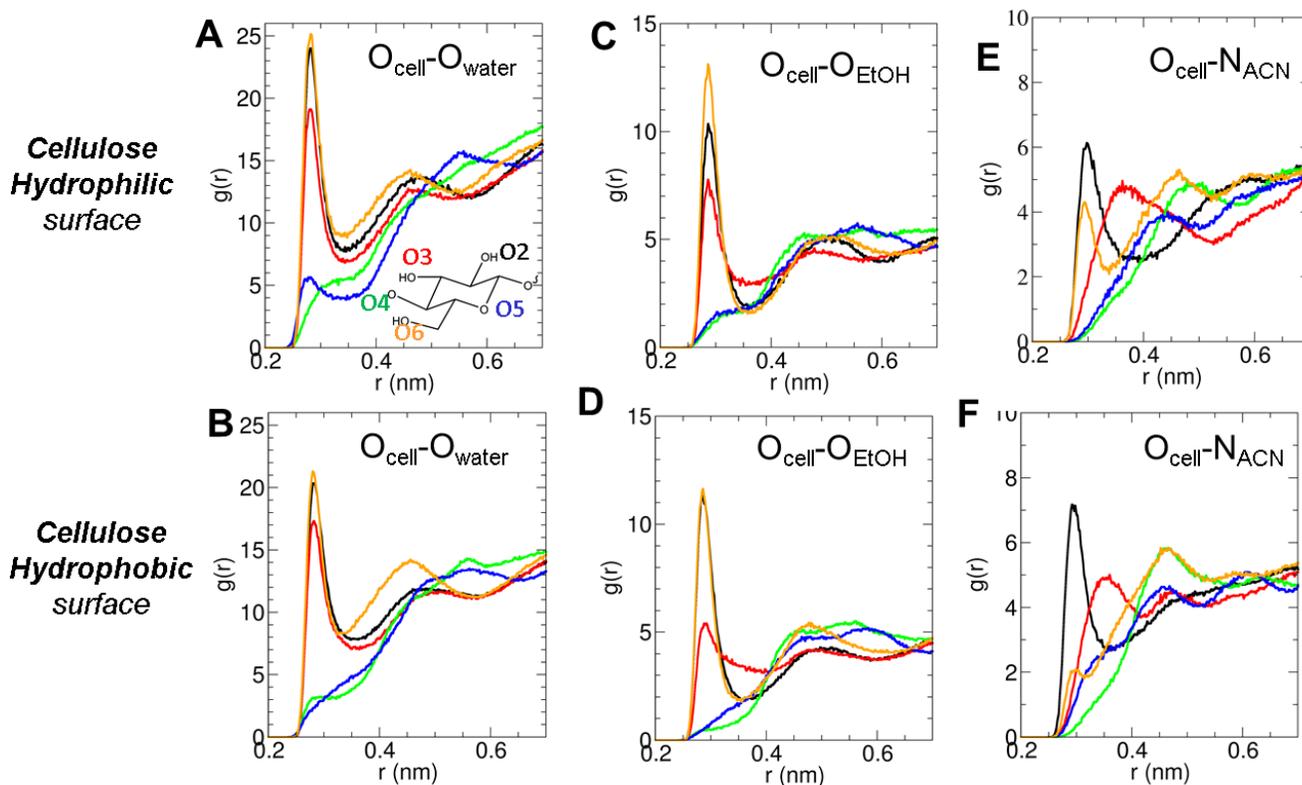
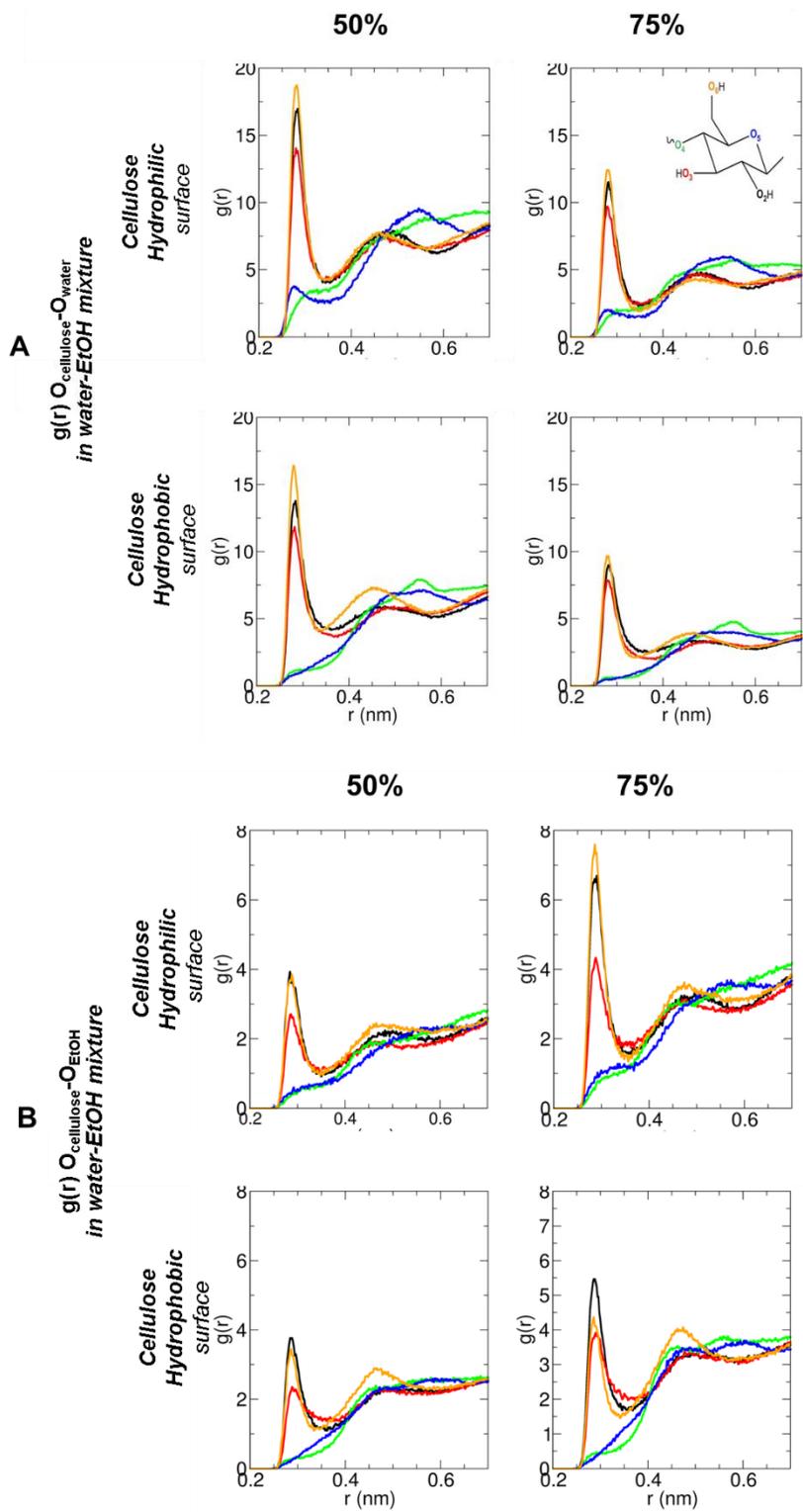
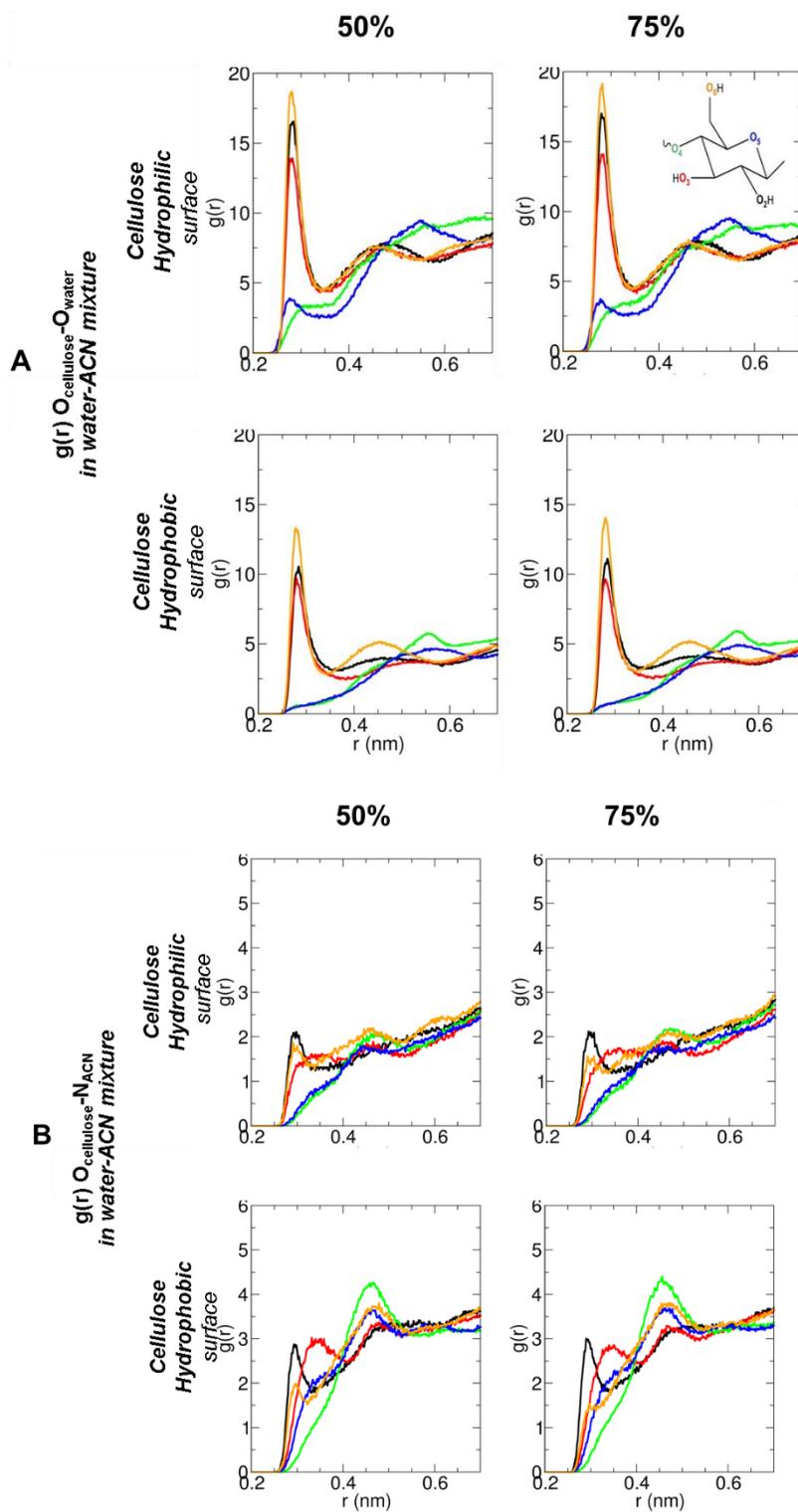


Figure S5-1-1- Solvation of cellulose in pure solvents. Radial distribution functions, $g(r)$, of the Ocellulose-Xsolvent pairs, for hydrophilic (chains AF) and hydrophobic (chains BDEG) surfaces. Ocellulose colour code is O2 (black), O3 (red), O4 (green), O5 (blue), and O6 (orange), see Figure 2 for oxygen notation. The reference site, X, in the solvents are water oxygen, EtOH oxygen, and ACN nitrogen.

Figure S6-2. $\text{O}_{\text{cell}}\text{-O}_{\text{water}}$ (A) and $\text{O}_{\text{cell}}\text{-O}_{\text{EtOH}}$ (B) $g(r)$ s in water-EtOH mixtures (wt%)

Figure S6-3. $O_{\text{cell}}-O_{\text{water}}$ (A) and $O_{\text{cell}}-O_{\text{ACN}}$ (B) $g(r)$ s in water-ACN mixtures (wt%)

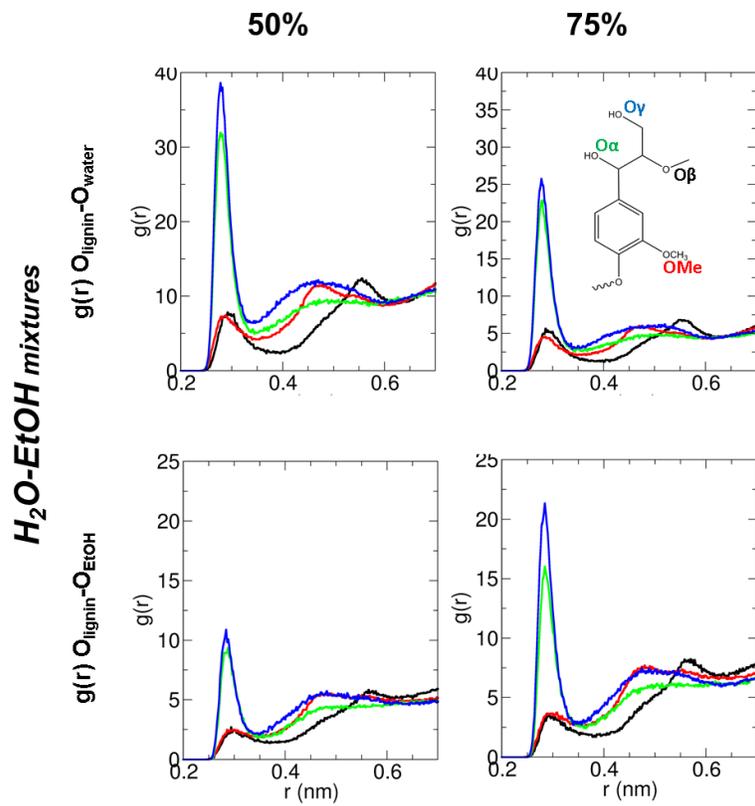


Figure S6-4. $O_{\text{lig}}-O_{\text{water}}$ and $O_{\text{lig}}-O_{\text{EtOH}}$ in water-EtOH mixtures (wt%)

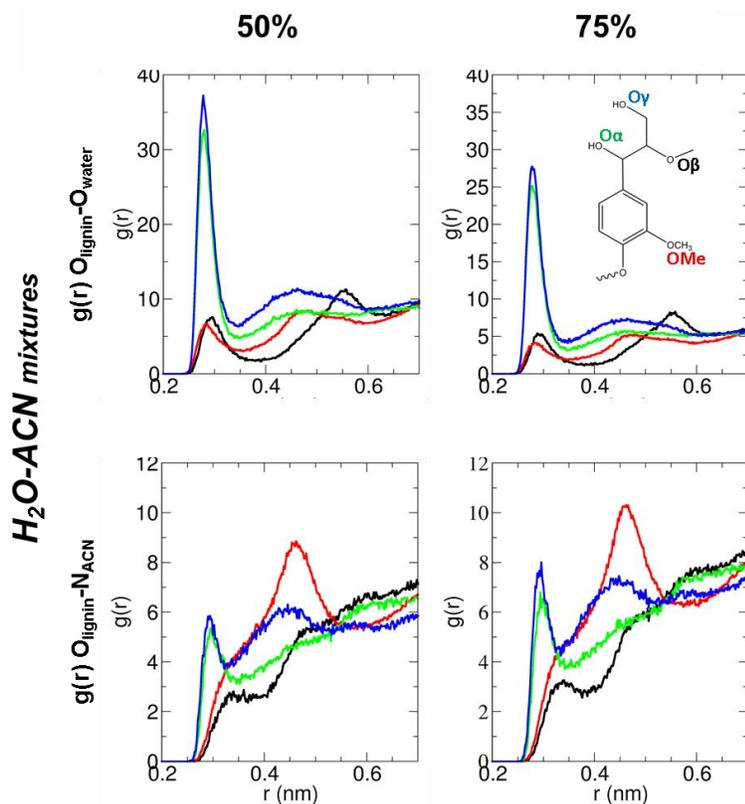


Figure S6-5. $O_{\text{lig}}-O_{\text{water}}$ and $O_{\text{lig}}-O_{\text{ACN}}$ in water-ACN mixtures (wt%)

6.2 Number of solvent molecules within 0.7 nm cellulose-solvent shell in cellulose-lignin complex

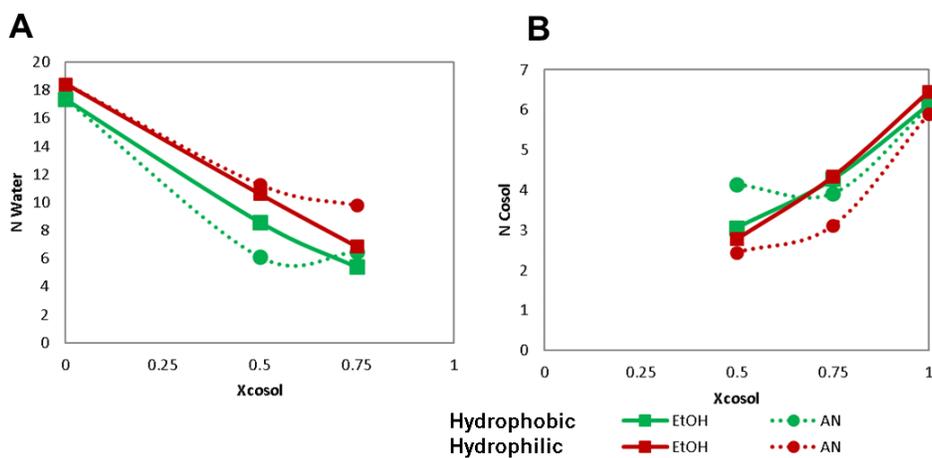


Figure S6-6. $O_{\text{lig}}-O_{\text{water}}$ and $O_{\text{lig}}-O_{\text{ACN}}$ in water-ACN mixtures (wt%)

6. SI Chapter 7. First-Principles Simulations of Two-Dimensional Electronic Spectroscopy of lignin model monomers and their β -O-4 linked dimer

7.1 Vertical excitation energies of lignin monomers

Table S 6-1. Vertical GS \rightarrow SN Excitation Energies (eV) and Transition Dipole Moments Magnitudes (au) in the anisole, guaiacyl, creosol, and homoveratrole above the first ionization potential. Anisole calculations were obtained at the CAS (7,8), while the rest of the monomers were obtained at the CAS(8,10) level. ^aDominant configurations (referred to orbitals in Figure X) and corresponding CI coefficients. States nomenclature refers to the ideal D6h symmetry of benzene (Table X). Experimental transition energies are reported, when available, in parentheses.

State <i>N_w</i>	Anisole						Guaiacol (GUA)						Creosol (CRE)						Homoveratrole (HVER)										
	TDM			TDM			TDM			TDM			TDM			TDM			TDM			TDM							
	E _E	Trans.	Coeff	GS	Lb	Label	E _E	Trans.	Coeff	GS	Lb	E _E	Trans.	Coeff	GS	Lb	E _E	Trans.	Coeff	GS	Lb	E _E	Trans.	Coeff	GS	Lb	Label		
9.31	2,3->5,4	-0.4	0.41	0.24			8.44	0a->5	-0.69	0.41	0.05	8.48	0a->5	0.65	0.58	0.28	8.11	0a->5	-0.82	0.64	0.21							7	
	3->5	-0.33						1->4	-0.42				3->4	0.29					-0.28										
	3->5,4	-0.32						3->6	-0.32				2p,3->5,4	-0.27					1->4	0.46	0.52	0.51						8	
	2,3->5,4	-0.42						3->4	-0.45	0.62	0.3	8.52	3->5,4	0.35	0.42	0.26			3->4	0.38									
	2,3->5	-0.32						2,3->5,4	-0.25				1->4	0.35					3->5,4	0.34									
	2,3->5	0.27						0a->4	0.24				3->6	-0.29					3->5,4	-0.25									
8.63	0->5	0.59	0.83	0.86			8.82	0a->5	0.5	0.6	0.56	8.95	0a->5	0.49	0.56	0.47	8.42	1->4	-0.52	0.66	0.63							9	
	1->5	-0.58						1->4	-0.44				3->4	-0.37					0a->4	0.39									
								3->5,4	-0.36				1->4	0.36					3->5,4	0.32									
								3->6	-0.23				2p->5	-0.28					3->6	-0.31									
8.67	1->4	0.58	0.18	0.39			8.98	0a->4	0.76	0.07	0.22	9.05	0a->4	0.52	0.36	0.42	8.46	0a->4	0.5	0.1	0.49							10	
	0->4	-0.48						1->5	0.31				2p,3->5	0.33					1->5	0.45									
	3->6	0.31						0b->5	0.35	0.05	0.97		2p,3->5	-0.28					2->6	0.37									
								0b->5	-0.33				0b->5	0.23					0b->5	0.3									
								1->5	0.33				1->5	0.44	0.1	1.07	8.62	0b->5	-0.42	0.24	0.88								
								2,3->5,4	-0.32				2p,3->5,4	0.3					0a->4	0.4									
9.99	0->5	-0.52	0.14	0.2			9.45	0b->5	0.35	0.05	0.97	9.15	1->5	0.24					0a->4	0.4									
	2->6	0.5						0a->4	-0.33				0b->5	0.24					0b->5	-0.28									
	1->5	-0.28						1->5	0.33				2p,3->5,4	0.3					1->5	-0.28									
	2->5,4	0.26						2,3->5,4	-0.32				0a->4	0.71	0.37	0.61	9.19	2,3->5,4	-0.27										
													0a->4	0.23					2,3->5,4	0.39									
													2p,3->5	-0.23					3->5	-0.52	0.07	0.25							
10.14	3->5,4	-0.55	0.26	0.36			9.6	2,3->5	-0.42	0.49	0.61	9.56	3->5	0.26					2p,3->5	0.26									
	2,3->5,4	0.38						2->5	0.26				2p,3->5	0.26					2p,3->5	0.26									
	2,3->5	0.33						1->4	0.25				2p,3->5	0.26					2p,3->5	0.26									
	3->5	0.45	0.1	0.31				3->5	-0.38	0.18	0.41		2p,3->5	0.26					2p,3->5	0.26									
	2,3->5,4	0.36						2->5	-0.35				0b->4	-0.49	0.22	0.51	10.31	0b->4	-0.49	0.22	0.51								
	3->4	0.28						1->5	0.34				0b->5	0.45					3->6	0.35									
	0->4	0.58	0.17	0.38				0b->4	0.67	0.15	0.34		2p,3->5	0.31					3->6	-0.29									
	3->6	0.37						2,3->5	-0.24				0b->5	0.45	0.05	0.21	9.98	0b->5	0.57	0.22	0.03								
	2->4	0.27						0b->5	0.6	0.11	0.26		0b->4	0.41					2->6	-0.48									
								2->6	-0.39				0b->4	-0.31					2->6	-0.48									
								2->4	-0.35				2p,3->5	0.31					2p,3->5	0.34									
								3->4	0.56	0.11	0.48		3->4	-0.48	0.05	0.48	10.6	3->4	-0.48	0.05	0.48								
								2->5	-0.41				2p,3->5,4	-0.42					2,3->5,4	0.43									
								2,3->5,4	-0.39				2p,3->5,4	-0.37					2,3->5,4	0.42									
10.98	2->5	0.62	0.06	0.47			10.65	3->4	0.56	0.11	0.48	10.29	0b->5	0.45	0.05	0.21	10.35	3->4	-0.57	0.13	0.47								
	3->4	-0.38						2->5	-0.41				0b->4	0.41					2,3->5,4	0.43									
	2,3->5,4	0.28						2,3->5,4	-0.39				2p,3->5,4	-0.37					2,3->5,4	0.42									

Table S 6-2. Vertical GS→SN Excitation Energies (eV) and Transition Dipole Moments Magnitudes (au) in the anisole, guaiacyl, creosol, and homoveratrole at the CASSCF and RASSCF level. States nomenclature refers to the ideal D6h symmetry of benzene (Table X).

State	GUA			CRE			HVER		
	CAS(10,8)	RAS(4,5 0,0 4,3)	Δ	CAS(10,8)	RAS(4,5 0,0 4,3)	Δ	CAS(10,8)	RAS(4,5 0,0 4,3)	Δ
¹ B _{2u}	4.58	4.59	-0.01	4.5	4.5	0	4.51	4.51	0
¹ B _{1u}	5.86	5.86	0	5.83	5.81	0.02	5.65	5.64	0.01
¹ E _{1u}	6.69	6.62	0.07	6.51	6.45	0.06	6.42	6.36	0.06
¹ E _{1u}	6.7	6.65	0.05	6.52	6.46	0.06	6.49	6.42	0.07
¹ E _{2g}	7.16	7.16	0	7.09	7.07	0.02	6.95	6.95	0
¹ E _{2g}	7.22	7.19	0.03	7.09	7.08	0.01	7.02	6.99	0.03
N _w	8.44	8.44	0	8.48	8.48	0	8.11	8.09	0.02
1D	9	9.03	-0.03	8.52	8.56	-0.04	8.39	8.41	-0.02
21E _{2g}	8.82	8.86	-0.04	8.95	8.99	-0.04	8.42	8.41	0.01
21E _{2g}	8.98	8.98	0	9.05	9.06	-0.01	8.46	8.46	0
1O	9.45	9.42	0.03	9.15	9.14	0.01	8.62	8.64	-0.02
1O				9.19	9.16	0.03			
2D	9.6	9.62	-0.02	9.56	9.58	-0.02	9.49	9.5	-0.01
3D	9.63	9.7	-0.07				10.07	10.08	-0.01
2O	10.36	10.38	-0.02	10.31	10.32	-0.01	9.87	9.87	0
2O	10.34	10.35	-0.01	10.29	10.29	0	9.98	9.99	-0.01
4D	10.65	10.67	-0.02	10.6	10.61	-0.01	10.35	10.35	0

7.2 Vertical excitation energies of the unstacked lignin dimer

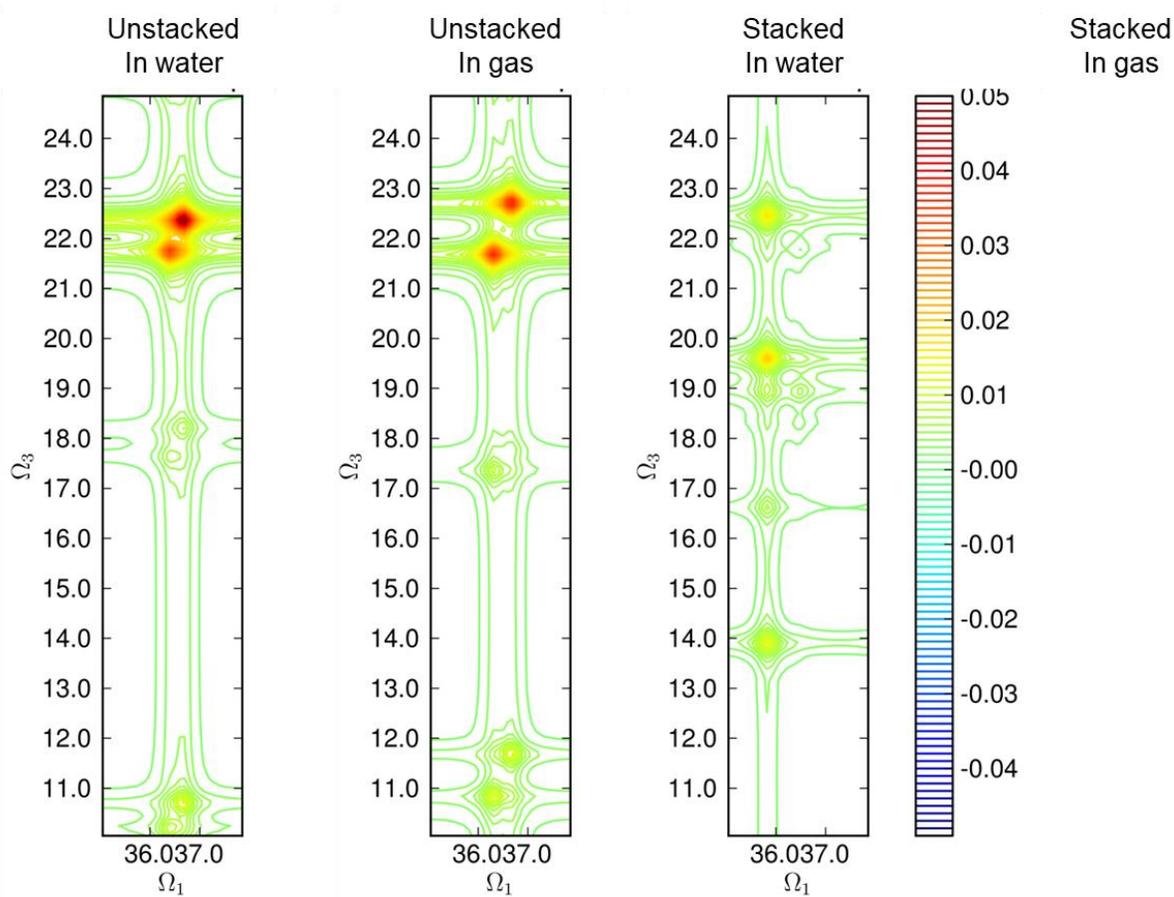
Table S 6-3. Vertical excitation energies of the unstacked lignin dimer

SA10-ss-PT2-CAS(20,16)								SA-50-ssPT2-RAS(4,10 0,0 4,6)							
Lb*→Sn		Trans.	Label	Lb*→Sn		Lb→Sn		Lb*→Sn		Trans.	Label	Lb*→Sn		Lb→Sn	
E(eV)	TDM			E (cm-1)	TDM	E (cm-1)	TDM	E(eV)	TDM			E (cm-1)	TDM	E (cm-1)	TDM
0			GS	-36.38	0.21	-36.70	0.25	0			GS	-36.29	0.23	-36.70	0.27
4.51	0.21	3*→5*	1*, Lb*	0.00		-0.32	0.00	4.5	0.23	3*→5n*	1*, Lb*	0.00		-0.32	0.00
		2*→4*								2*→4*					
		3→5*								3*→4*					
		2*→4													
4.55	0.25	3→5	1, Lb	0.32	0.00	0.00		4.55	0.27	3→5n	1, Lb	0.32	0.00	0.00	
		2→4								2→4					
		3*→5								3→4					
		2→4*								2→5n					
5.77	0.71	3→4*	2*?	10.16	0.18		0.03	5.47	0.72	3*→4*	2*	7.82	0.15	7.50	0.02
		3*→4*								2*→5n*					
		1*→4*								3*→5n*					
		3*→4								1*→4*					
5.87	0.49	3→4	2	10.97	0.03	10.65	0.17	5.61	0.50	3→4	2(1)	8.95	0.02	8.63	0.15
		3*→4								3→5n					
		2→5								2→5n					
		3*→4*								2→4					
		1→4								1→4					
								5.99	2.41	2*→4*	3*	12.02	0.08	11.70	0.03
										3*→5n*					
										2→4					
6.7	0.35	2*→5*	4*	17.66	0.11	17.34	0.03	6.13	1.01	2*→5n*	4*	13.15	0.16	12.82	0.03
		3→4*								3*→4*					
		3*→4*								2→5n					

SA10-ss-PT2-CAS(20,16)								SA-50-ssPT2-RAS(4,10 0,0 4,6)							
Lb*→Sn		Trans.	Label	Lb*→Sn		Lb→Sn		Lb*→Sn		Trans.	Label	Lb*→Sn		Lb→Sn	
E(eV)	TDM			E (cm-1)	TDM	E (cm-1)	TDM	E(eV)	TDM			E (cm-1)	TDM		
6.81	0.33	1*->4*	4	18.47	0.02	18.23	0.11	6.19	2.38	3->4		13.63	0.14	13.31	0.05
		2->5								2*->5n*	4*(4)				
		3->4								2->5n					
		1->4								3->4					
										3*->4*					
								6.22	1.76	2->4					
										3->5n*	ct1.1	13.87	0.02	13.55	0.11
										2->5n					
								6.36	1.39	2->4					
										3->5n*	3->5n*	14.92	0.03	14.60	0.12
										2->4					
										3->5n					
										2->5n					
7.21	0.53	1*->5*	6*	21.78	0.22	21.45	0.01	7.02	0.60	1*->5n*	6*	20.33	0.22	19.92	0.01
		2*->6*								2*->5n*					
		2*->4*,5*								2*->6*					
										3*->4*					
										3*->5n*					
								7.03	0.42	1*->4*	5*?new?	20.33	0.19	20.00	0.04
										2*->5n*					
										2*,3*->4*					
										3*->4*,5n*					
										3*->6*					
								7.08	0.19	3->4*	ct2.1	20.73	0.03	20.41	0.02
								7.1	0.37	1->4	5?new?	20.97	0.01	20.57	0.16
										2->5n					
										3->6					
										2,3->4					
										3->5n					
										3->4					
								7.23	0.25	0as*->5n*	1Ss*.1	22.02	0.22	21.62	0.02
										Obs*->4*					
										1*->4*					
										Obs*->5n*					
										3*->6*					
7.32	0.64	1->5	6	22.66	0.01	22.42	0.23	7.23	0.73	1->5n	6	22.02	0.03	21.70	0.24
		2->5								2->4,5n					
		2->6								2->6					
										1->4					
										2->5n					
										3->5n					
										2->4					
7.49	0.02	3,3*->4*,4	?	24.04	0.00	23.71	0.00	7.48	0.02	3*,3->4*,4	2*+2	24.04	0.00	23.71	0.01
		3->4*								3*,3->4*,5n					
										2*,3->5n*,4					
										2,3*->4*,5n					
								7.3	0.10	3*->5 n	ct1.3	22.58	0.03	22.26	0.05
										3*->4					
								7.52	0.79	0as*->5n*	1Ss*.2	24.36	0.49	24.04	0.04
										Obs*->4*					
										1*->4*					
								7.6	0.15	2->5n*	ct4	24.92	0.07	24.60	0.09
								7.62	0.39	Obs*->5n*	3Ss*	25.16	0.77	24.84	0.01
										Obs*->4*					
										1*->5n*					
										0as*->4*					
								7.63	0.12	3->4*	ct2(ct1)	25.16	0.02	24.84	0.02
										3*,3->5n*,5n					
										3*,3->5n*,4					
										2*,3->4*,5n					
								7.64	0.28	0as*->4*	2Ss*.1	25.33	0.13	25.00	0.08
								7.66	0.05	2->4*	ct3.1	25.41	0.09	25.08	0.12
						dark		8.15		2,3*->4*,5n	ID?				
										2*,3->5n*,4					
										2,3*->4*,4					
										2*,3->5n*,5n					
						dark		8.15		2*2->5n*,5n	4*+4				
										3*,3->4*,4					
										2*2->5n*,4					

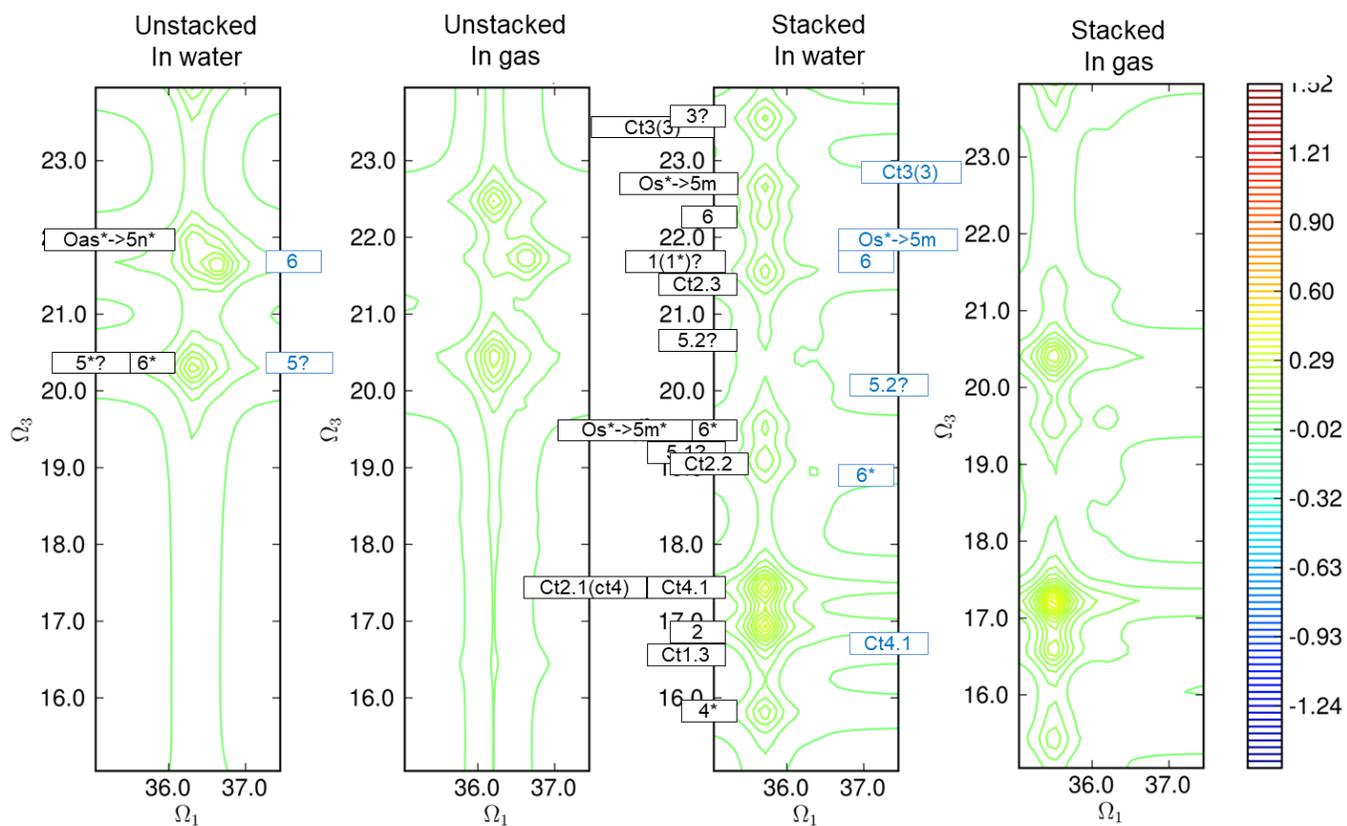
								7.92	0.22	no		m.3	28.15	0.06	27.50	0.02
								8.28	0.11	2d*,3d-->5m,4m		1D?	31.05	0.25	30.41	0.11
										2d,3d*-->5m*,4d						
								8.42	0.21	no		m.4	32.18	0.09	31.54	0.02
								8.5	0.21	GS		GS	32.91	0.14	32.26	0.11
								8.76	0.09	l->5m*		C16.1	34.92	0.16	34.28	0.15
								8.79	0.29	0b*->5m*		3S*(6*)	35.17	0.38	34.52	0.34
										l*->5m*						
								8.96	0.57	no		m.5	36.54	0.12	35.89	0.12
								9.03	0.10	l->5m		6	37.10	0.15	36.46	0.12
								9.2	0.04	no		m.5	38.47	0.13	37.83	0.07
								9.4	0.08	no		m.6	40.09	0.11	39.44	0.04
								9.43	0.07	no		m.7	40.41	0.14	39.68	0.04
								10.25	0.10	no		m.8	47.02	0.05	46.30	0.03

7.4 Comparison of 2D spectra of lignin dimer in gas phase and in water



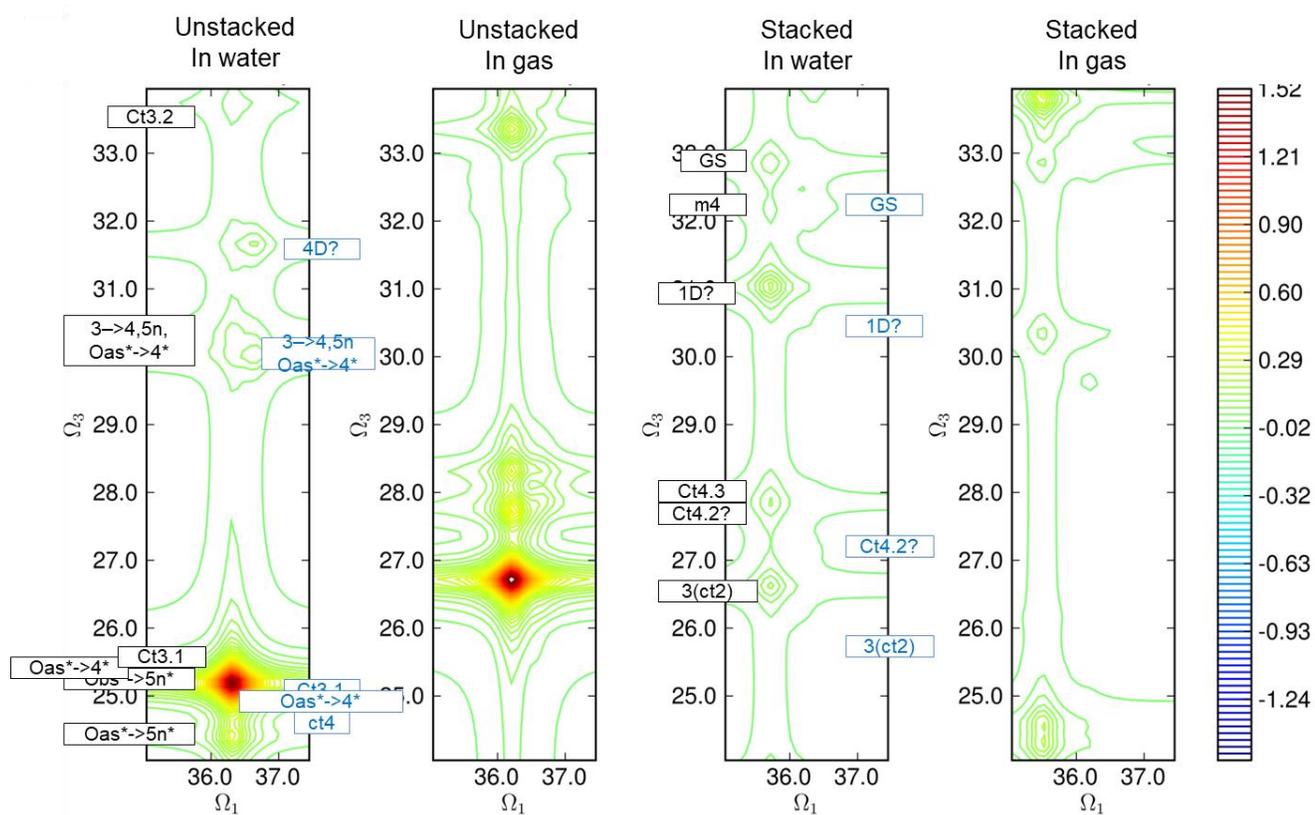
SA10-CAS (20,16)

Figure S 6-1. 2D spectra of lignin dimer in gas phase and water at the SA10-CAS(20,16) level



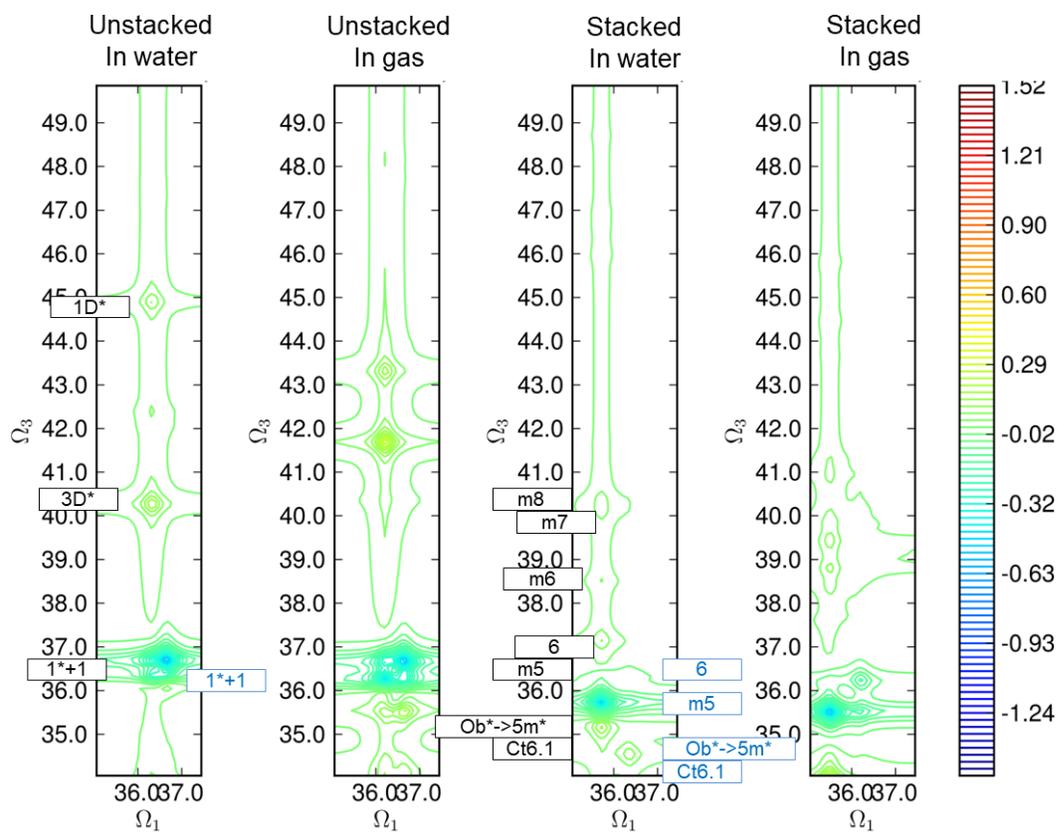
SA50-RAS (4,10|0,0|4,6)

Figure S 6-2. 2-color (15000-24000 cm^{-1}) 2D spectra of lignin dimer in gas phase and water at the SA50-RAS(4,10/0,0/4,6) level



SA50-RAS (4,10|0,0|4,6)

Figure S 6-3. 2-color (24000-34000 cm^{-1}) 2D spectra of lignin dimer in gas phase and water at the SA50-RAS(4,10/0,0/4,6) level



SA50-RAS (4,10|0,0|4,6)

Figure S 6-4. 1-color 2D spectra of lignin dimer in gas phase and water at the SA50-RAS(4,10|0,0|4,6) level

7.5 2D spectra of HVER monomer and monomer G' from the unstacked lignin dimer in gas phase

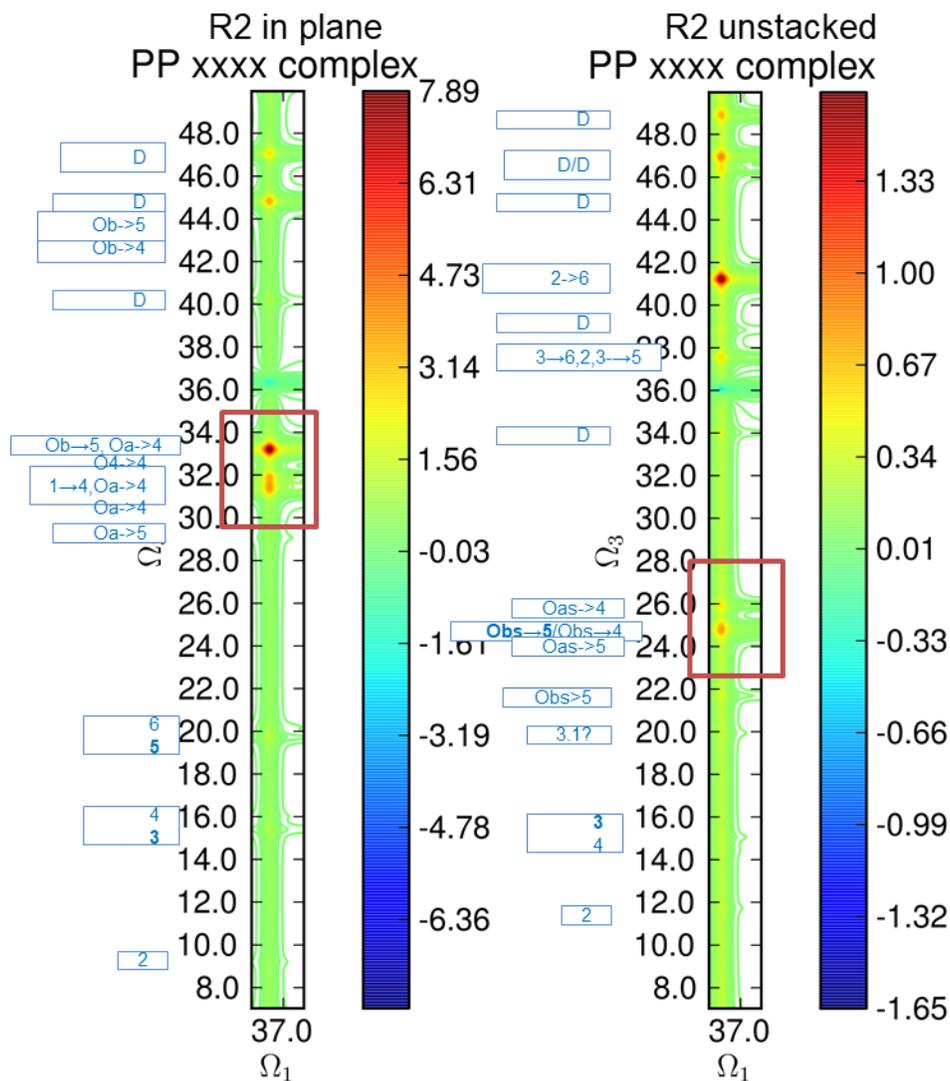
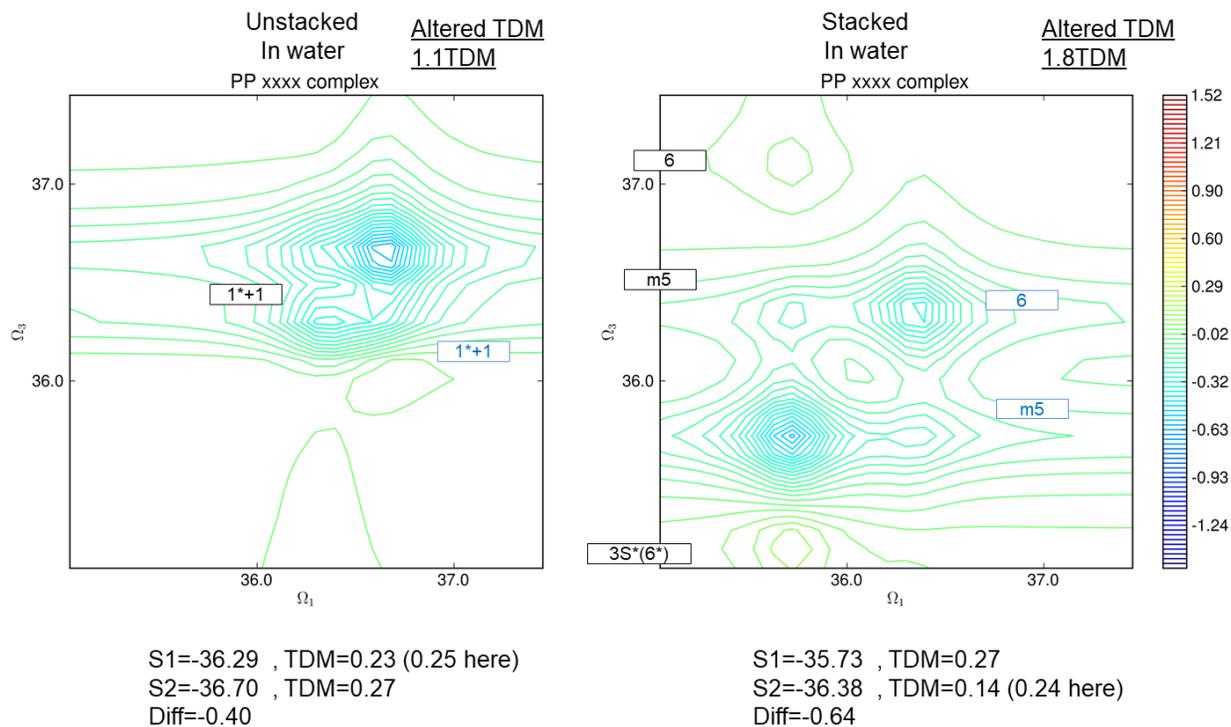


Figure S 6-5. 2-color 2D spectra of HVER monomer and monomer G' from the unstacked lignin dimer in gas phase

7.6 TDM Modified 1-color 2D spectra of unstacked and stacked lignin dimer in solution



SA50-RAS (4,10|0,0|4,6)

Figure S 6-6. TDM Modified 1-color 2D spectra of unstacked and stacked lignin dimer in solution

References

1. Pires, R.M., et al., *Viscosity and Density of Water + Ethyl Acetate + Ethanol Mixtures at 298.15 and 318.15 K and Atmospheric Pressure*. Journal of Chemical & Engineering Data, 2007. **52**(4): p. 1240-1245.
2. Cunningham, G.P., G.A. Vidulich, and R.L. Kay, *Several properties of acetonitrile-water, acetonitrile-methanol, and ethylene carbonate-water systems*. Journal of Chemical & Engineering Data, 1967. **12**(3): p. 336-337.

3. Hurle, R.L., A.J. Easteal, and L.A. Woolf, *Self-diffusion in monohydric alcohols under pressure. Methanol, methanol (2H) and ethanol*. Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases, 1985. **81**(3): p. 769-779.
4. Pang, F.-M., et al., *Densities and viscosities of aqueous solutions of 1-propanol and 2-propanol at temperatures from 293.15 K to 333.15 K*. Journal of Molecular Liquids, 2007. **136**(1): p. 71-78.
5. Grande, M.d.C., H.L. Bianchi, and C.M. Marschoff, *On the density of pure acetonitrile*. Anales de la Asociación Química Argentina, 2004. **92**: p. 109-114.
6. Lide, D.R., *CRC handbook of chemistry and physics*. 12J204, 2004.
7. Tanaka, K., *Self-diffusion coefficients of water in pure water and in aqueous solutions of several electrolytes with 18O and 2H as tracers*. Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases, 1978. **74**(0): p. 1879-1881.
8. Shaker-Gaafar, N., et al., *p,T-Dependence of Self-diffusion in Liquid Ethanol and the Propanols*. Berichte der Bunsengesellschaft für physikalische Chemie, 1993. **97**(6): p. 805-811.
9. Hurle, R.L. and L.A. Woolf, *Self-diffusion in liquid acetonitrile under pressure*. Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases, 1982. **78**(7): p. 2233-2238.

Abstract

Extraction of cellulose and lignin from plant biomass remains a major issue for enabling more economic and green production of lignocellulosic renewable fuels and byproducts. Although the use of multicomponent solvents has provided remarkable results in wood fractionation processes most of the currently used methods rely on empirically elaborated protocols. Understanding the physicochemical mechanisms of biomass breakdown and its interactions with solvent medium during fractionation will lead to more efficient use of biomass. This defined the focus in this thesis work on a systematic and detailed description of the interactions between lignocellulose components with binary water-organic mixtures of ethanol and acetonitrile. Our results and their analysis were obtained predominantly with molecular dynamics (MD) simulations, and supported by additional studies of quantum chemical (Density Functional Theory) and mixed quantum mechanical (QM) and classical MD scheme (QM/MM). With these tools we first established a non-linear behavior of the mixed solvent structures, thermodynamic properties and dynamic hardness, as a measure for their global reactivity. The analysis of the average numbers of HBs with the liquid composition shows that alcohol molecules tend to substitute water molecules, allowing compensating for the loss of H-bonds in the water solvent domains. The role of organic component in water solvent mixtures on the conformational changes induced in the main wood components (cellulose, lignin and hemicellulose) is highlighted and their dependence on distinct solvent compositions is unveiled for each organic solvent component and its content in water. This dependence is explained by preferential solute-solvent interatomic interactions as a function of solvent compositions. Subsequently, the evolution of interaction forces in lignin-cellulose and lignin-xylan complexes are also found to have solvent-dependent profiles. All this supports the general conclusion about specific solvent actions on lignocellulose compounds being the driving factors in the observed macroscopic non-linear behavior in wood swelling in mixed water-organic solvents.

Key-words: binary water-organic solvents; lignocelluloses; molecular dynamics, QM/MM

Résumé

L'extraction de la cellulose et de la lignine à partir de la biomasse végétale reste un enjeu majeur pour une production plus économique et verte de carburants et de sous-produits lignocellulosiques renouvelables. Bien que l'utilisation de mélanges de solvants ait donné des résultats remarquables dans les processus de fractionnement du bois, la plupart des méthodes actuellement utilisées reposent sur des protocoles empiriques. Comprendre les mécanismes physico-chimiques de la décomposition de la biomasse et ses interactions avec les solvants lors du fractionnement permettra d'améliorer l'efficacité de ces méthodes. Dans cette perspective, cette thèse porte sur la description systématique et détaillée des interactions entre les composantes de la lignocellulose et des mélanges binaires eau-solvant organique (éthanol ou acétonitrile). Nos résultats et leur analyse ont été obtenus principalement avec des simulations de dynamique moléculaire (MD) et étayés par des études supplémentaires basées sur la chimie quantique (théorie de la densité fonctionnelle) et la mécanique quantique (QM) et le schéma mixte mécanique quantique/mécanique classique (QM/MM). Avec ces outils, nous avons d'abord établi un comportement non linéaire des structures avec des solvants mixtes, de leurs propriétés thermodynamiques et de la dureté dynamique, en relation avec leur réactivité globale. L'analyse du nombre moyen de liaisons H avec la composition du liquide montre que les molécules d'alcool ont tendance à se substituer aux molécules d'eau, ce qui permet de compenser la perte de liaisons H dans les domaines solvatés par l'eau. Le rôle des groupements organiques dans les mélanges de solvants sur les changements de conformation induits dans les principaux composantes du bois (cellulose, lignine et hémicellulose) est mis en évidence et leur dépendance vis-à-vis de la composition en solvants distincts est dévoilée pour chaque solvant organique et sa concentration dans l'eau. Cette dépendance est expliquée par les interactions interatomiques préférentielles soluté-solvant en fonction des compositions en solvants. Par la suite, l'évolution des forces d'interaction dans les complexes de lignine-cellulose et de lignine-xylane s'avère également présenter des profils dépendant des solvants. Tout ceci conforte la conclusion générale selon laquelle les actions spécifiques des mélanges de solvants à base d'eau et de composés organiques sur les composantes de la lignocellulose sont des facteurs déterminants dans le comportement non linéaire macroscopique observé dans le gonflement du bois

Mots-clefs : mélanges binaires eau-solvant organique; lignocellulose; dynamique moléculaire, QM / MM

Abstract

L'estrazione di cellulosa e lignina dalla biomassa vegetale rimane una grande sfida per consentire una produzione più economica ed ecocompatibile di carburanti e sottoprodotti lignocellulosici rinnovabili. Sebbene l'uso di solventi multicomponente abbia prodotto risultati notevoli nei processi di frazionamento del legno, la maggior parte dei metodi attualmente utilizzati si basa su protocolli sviluppati empiricamente. Comprendere i meccanismi fisico-chimici della decomposizione della biomassa e le sue interazioni con il solvente durante il frazionamento consentirà un uso più efficiente della biomassa. Questa la motivazione del lavoro di tesi su una descrizione sistematica e dettagliata delle interazioni tra i componenti della lignocellulosa con miscele binarie acqua-organico con etanolo e acetonitrile. I nostri risultati e le nostre analisi sono stati ottenuti principalmente con simulazioni di dinamica molecolare (MD) e supportati da ulteriori studi sulla chimica quantistica (teoria del funzionale della densità), sulla meccanica quantistica (QM) e sullo schema misto meccanica quantistica / meccanica classica (QM / MM). Con questi strumenti, abbiamo prima definito un comportamento non lineare delle strutture dei solventi misti, le loro proprietà termodinamiche e la durezza dinamica, che misura la loro reattività complessiva. L'analisi del numero medio di legami idrogeno con la composizione mostra che le molecole di alcol tendono a sostituire le molecole di acqua, il che rende possibile compensare la perdita di legami H nelle frazioni acqua del solvente. Viene evidenziato il ruolo dei componenti organici delle miscele di solventi acquosi sui cambiamenti conformazionali indotti nei principali componenti del legno (cellulosa, lignina ed emicellulosa) e la loro dipendenza dalla composizione di diversi solventi misti viene chiarito per ciascun componente di solvente organico e per il suo contenuto in acqua. Questa dipendenza è spiegata dalle interazioni interatomiche preferenziali soluto-solvente in funzione delle composizioni del solvente. Anche l'evoluzione delle forze di interazione nei complessi lignina-cellulosa e lignina-xilano sembra mostrare profili dipendenti dal solvente. Tutto ciò supporta la conclusione generale che le azioni specifiche di solventi misti a base di acqua e organici su composti di lignocellulosa sono determinanti nel comportamento macroscopico non lineare osservato nel gonfiamento del legno.

Parole chiave: solventi binari idro-organici; lignocellulosica; dinamica molecolare, QM / MM