Structure, flexibility and hydration properties of lignin dimers studied with Molecular Dynamics simulations

Wood Chemistry

Klara Hackenstrass, Merima Hasani and Malin Wohlert*

Structure, flexibility and hydration properties of lignin dimers studied with Molecular Dynamics simulations

https://doi.org/10.1515/hf-2023-0054
Received May 26, 2023; accepted December 19, 2023;
published online January 5, 2024

Abstract: Lignin is an abundant polymer found in wood and grasses, but due to its heterogeneity and complex macromolecular structure it has been less utilized than cellulose. While the building blocks are known, the way they are linked is less understood. Here, Molecular Dynamics simulations were used to systematically characterize seven linkages found in native lignin. Their influence on lignin and water structure, and their interactions were analyzed. The study is limited to guaiacyl (G-G) dimers connected by the following common softwood linkages; 5-5′, 4-O-5′, α-O-4′, β-1′, β-5′, β-O-4′ and β-β′. The simulations show that the linkage has a significant effect on conformational preference and lignin-water interaction. Especially, the behavior of the β-O-4′ shows unique properties, both in terms of conformational freedom and interaction with water. Within the β-O-4′ dimer, π–π stacking between the aromatic rings is possible. The molecule has two distinct common conformations, one compressed and one extended. These preferences also lead to a different effect of β-O-4′ dimer on the surrounding water, where water is found close to the linkage itself but expelled from the aromatic rings to a larger extent than the other linkages. These findings are important for lignin solubility as well as its depolymerization mechanisms.

Keywords: hydration; lignin; molecular dynamics

***Corresponding author: Malin Wohlert**, Department of Materials Science and Engineering, Uppsala University, SE-75103, Uppsala, Sweden, E-mail: malin.wohlert@angstrom.uu.se. https://orcid.org/0000-0002-1620-6631

Klara Hackenstrass, Department of Materials Science and Engineering, Uppsala University, SE-75103, Uppsala, Sweden. https://orcid.org/0000-0002-5380-2184

Merima Hasani, Forest Products and Chemical Engineering, Chalmers University of Technology, SE-41296, Gothenburg, Sweden; and Wallenberg Wood Science Center, Chalmers University of Technology, SE-41296, Gothenburg, Sweden

1 Introduction

The world's biomass is largely comprised of lignin, which next to cellulose is the most abundant biopolymer (Boerjan et al. 2003; Ralph et al. 2004). Its name refers to a heterogeneous family of aromatic polymers that play an important role in the biological function of vascular plants, allowing for water to be transported over long distances without the plant losing its structural stability (Kenrick and Crane 1997). Despite its abundance, lignin is traditionally considered a byproduct of several large scale industrial processes such as the commonly applied kraft process in pulp production (Gandini 2008). Within the kraft process, lignin is fragmented, extracted from wood cells and subsequently combusted as part of the process' energy cycle, leaving an energy surplus at the mill (Gierer 1980; Jönsson et al. 2013). Exploiting the potential of lignin beyond energy production is desirable, it has been estimated that 6–9 million tons of kraft lignin per year could be extracted without altering the energy balance within the pulp mills (Berlin and Balakshin 2014). Therefore, there is a growing interest in developing approaches to valorize lignin and use it as a feedstock for sustainable material or chemical production (Ragauskas et al. 2014; Rinaldi et al. 2016), and to that end, better understanding of its fundamental structure-property relations is needed.

A possible explanation for the industrial hesitance to use lignin lays in its heterogeneity. All lignins are mainly comprised of the three monomers guaiacyl (G), syringyl (S) and *p*-hydroxyphenyl (H or P) connected by ether or carbon-carbon bonds. The prevalence of each monomer and linkage is dependent on the plant specie, age and the tissue within one plant (Dellon et al. 2017; Nawawi et al. 2016; Timell 1983). Moreover, processing of lignin, particularly extraction from the wood tissue through rather harsh degradive chemistry, significantly changes its native structure adding even further to its structural diversity (Karlsson et al. 2020).

Characterizing lignin has also proven difficult, ranging from the heterogeneity within a single plant, to problems extracting lignin in its native structure and selecting

representative samples. One way to determine the molecular weight distribution of a polymer is size-exclusion chromatography. The molecular weight distribution obtained with this method however depends on the used column and solvent. The same applies when comparing between different experimental methods (Crestini et al. 2011; Lawoko et al. 2021). Techniques such as 2D-NMR are increasingly used to characterize lignin samples, adding information about monomer and linkage prevalence to existing studies (Capanema et al. 2004). Despite all dissents, it is generally agreed upon that softwood lignin consists mostly of G monomers (95 %) and the most frequent interunit linkage is β-O-4′ (50 %), though percentages vary (Capanema et al. 2004; Chang and Jiang 2020; Erickson et al. 1973). Experimentally most lignin samples are insoluble in water, which has given lignin the attribute of being hydrophobic. Chemically monolignols are substituted phenolic compounds with different degrees of methoxylation, combining aromatic hydrophobic groups as well as multiple hydrophilic hydroxy groups decorating the alkyl side chains and phenolic substituents (He et al. 2016; Henriksson 2017; Petridis and Smith 2016).

The growing understanding of lignin structure and the increasing interest to use lignin beyond combustion has lead to scientific interest in exploiting its potential fields of application (Ciesielski et al. 2020; Rinaldi et al. 2016). Consequently, this also leads to an increasing need to understand behavior of lignin during processing, where reactivity, solubility, diffusion and sorption properties are of critical importance (Mattsson et al. 2017).

Molecular Dynamics (MD) computational simulations are commonly used to study biopolymers and their interaction (Ciesielski et al. 2020). Since any MD simulation requires the chemical structure of all included materials, different approaches have been used to build representative lignin models. When aiming for a representative structure, one way is to use repeating building blocks, that recreate the ratios of prevalence expected in natural lignin, reducing effects due to changing linkages between lignin oligomers with different degrees of polymerization (Petridis and Smith 2016). A more recent approach as introduced by Yanez et al. (2016) is adopting a stochastic approach to vary the linkages and monomers when building a lignin polymer (Dellon et al. 2017; Yanez et al. 2016).

This way of generating lignin polymers together with the creation of a force field specifically designed for the monomers and linkages found in lignin has allowed studies of the interaction between lignin and other wood components such as cellulose and hemicellulose or the influence of the solvent (Jahan et al. 2022; Vermaas et al. 2019a). It has been shown that the interplay between the amphiphilic lignin and different solvents results in different configurations of the lignin depending on the hydrophilicity of the solvent (Jahan et al. 2022).

A second approach to handle the heterogeneity of lignin is to look at smaller segments and characterize the effects a specific linkage or monomer has on the behavior of lignin. Examples using this approach are MD or density functional theory (DFT) simulations in vacuum accompanied with experimental NMR or UV/IR studies (Besombes et al. 2003; Dean et al. 2014). The aforementioned MD study on the interaction of lignin with other wood components also included some results of lignin dimers at a cellulose surface, showing differences obtained with GG dimers depending on the linkages (Vermaas et al. 2019a). Another example is the work of Aguilera-Segura et al. (2019), combining cell wall sorption experiments of ethanol-water mixtures at different ratios and MD studies that include a simulation of a lignin dimer (G-β-O-4′-G) in the same ethanol-water mixtures. Correlating conformational behavior of lignin with its structure, linkage type and frequency in particular, is of high importance when elucidating or predicting mass transfer behavior of lignin during processing. Especially interesting is improving understanding for mass transfer properties as a function of structural features during delignification processes, where lignin linkages are partly broken aiming at dissolution and extraction. Yet, while the chemical reactions cleaving the linkages are relatively fast, diffusion of the formed lignin fragments through the wood tissue is comparably slow and thus, the rate determining step, probably decisively affected by the conformational behavior (Dang et al. 2016; Mattsson et al. 2017; Santos et al. 2011). A deeper understanding of lignin mass transfer (and its conformational component, closely related to the structural features) would open up for an improved process design allowing for a higher resource efficiency and a structural control when it comes to lignin building blocks (Baptista et al. 2006, 2008). Moreover, conformational behavior of lignin affects other types of processing steps such as precipitation, filtration and washing, essential unit operations of the main technology for recovery of kraft lignin for material applications (LignoBoost) (Öhman et al. 2007; Zhu and Theliander 2015), as well as behavior during membrane filtration used for both dewatering and fractionation of technical lignins (Aminzadeh et al. 2018).

In the present study, MD is used to systematically investigate the impact different naturally occurring structural variations have on the conformational and dynamic behavior of lignin as well as its effect on the surrounding water. This is important as it is a significant step towards more complex systems applied in large scale processing such as alkaline aqueous systems extensively applied in the above mentioned kraft pulping or aqueous systems with

varied pH used for recovery of kraft lignin (LignoBoost). Using smaller lignin model structures allows to link observed behavior as an effect of linkage and extrapolate how behaviors seen in larger lignin models correlate to single structural motives. Providing this information will help interpreting information gained by simulations with more realistic and representative lignin molecules and bridge information gained from simulations using varying native lignin structures.

2 Materials and methods

2.1 Simulated systems

In this study the results of Molecular Dynamics simulations of seven different single lignin dimers in water are presented. While native lignins may consist of different monolignol units (e.g., G, H, S), this study is focusing on guaiacyl (G) units, see Figure 1, being the dominant monolignol in softwood (Erickson et al. 1973). To create the dimers in this study, two monolignols are connected by one of seven different linkages, their

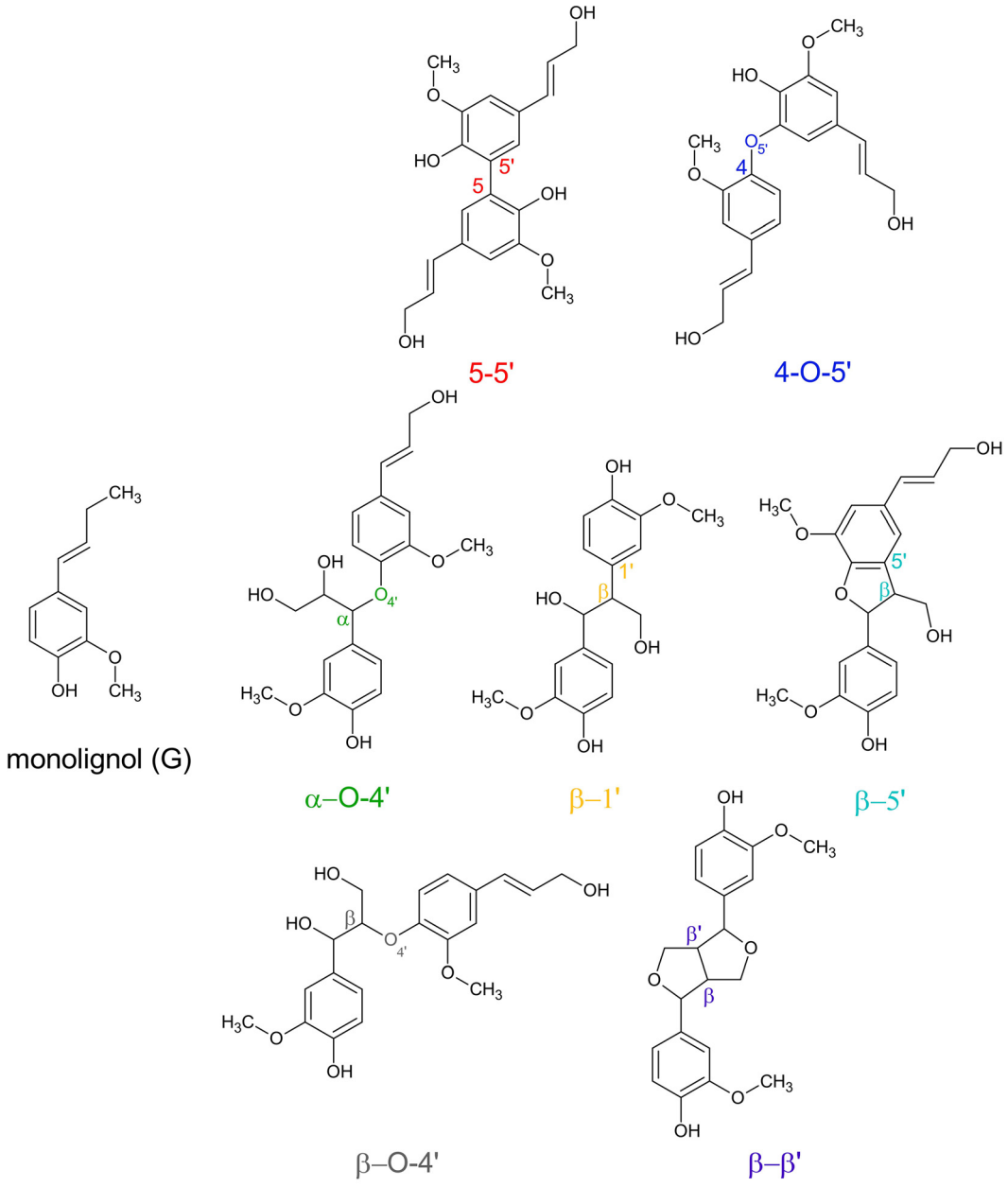

Figure 1: Chemical structure of the monolignol (G) and seven linkages connecting two monolignols. The atom labeling is according to commonly used nomenclature in lignin chemistry using both Greek letters and numbers for the end group and aromatic rings respectively. The linkages are sorted by increasing number of atoms in between the ring carbons (5-5′: 0; 4-O-5′: 1; *α*-O-4′, *β*-1′ and *β*-5′: 2; *β*-O-4′: 3 and *β*-*β*′: 4).

chemical structure can also be seen in Figure 1. The configuration is *R* for the *α*-carbon and *S* for the *β*-carbon whenever applicable in the molecule. The effect on linkage variation is expected to be dominating over the variation in stereochemistry. Previous MD simulations on small lignin molecules showed that property dependence, like binding energy to cellulose, of different stereochemistries is negligible compared to variation among different linkages (Vermaas et al. 2019a).

2.2 Simulation protocol

A single dimer was placed in a box of dimensions $4 \times 4 \times 4$ nm with periodic boundary conditions, filled with water molecules at 1000 kg/m^3 (the number of water molecules varied between 2146 and 2150). The simulation was equilibrated by a succession of NVT and NPT simulations, followed by a 200 ns NPT production run using the leap-frog algorithm with a time step of 2 fs Temperature was kept at 300 K using the v-rescale thermostat (Bussi et al. 2007) (τ_T = 0.1 ps), pressure was maintained with the Parinello-Rahman barostat (Parrinello and Rahman 1981) (τ_P = 2 ps). Direct Lennard-Jones and Coulomb interactions were truncated at 1.4 nm and Particle-Mesh Ewald method was used for calculating long-range electrostatic interactions (Darden et al. 1993; Essmann et al. 1995). Due to long-range van der Waals interaction, dispersion corrections for energy and pressure were applied. All bonds to hydrogen atoms in the lignin and water molecules were constrained with P-LINCS (Hess 2008). The coordinates were saved every 2 ps The simulations were performed using GROMACS 2021; Abraham et al. (2015) with the CHARMM force fields for lignin (Vermaas et al. 2019b) and TIP3P water (Jorgensen et al. 1983) that were transformed with the TopoTools plugin (Vermaas et al. 2016) for VMD (Humphrey et al. 1996).

2.3 Analysis

Simulations were analyzed using GROMACS functions (Abraham et al. 2015). VMD 1.9.4a55 (Humphrey et al. 1996) was used for visualization of the simulation, matplotlit (Hunter 2007), seaborn (Waskom 2021) and numpy (van der Walt et al. 2011) for visualization of data. All distributions and average values of properties are obtained by using the recorded values over the entire 200 ns production run.

In the analysis of the orientation and distance between the aromatic rings, the center of mass was calculated using the position of the six aromatic C atoms, excluding any substituents including H atoms. The definition of how the dimer is grouped into smaller segments (that are averaged over) used for the radial distribution functions can be found in the SI.

3 Results and discussion

In the following section, detailed analysis of lignin conformation, energetical analysis and characterization of water structure at the lignin-water interface is presented.

3.1 Lignin structure

This section contains results from analysis of how the aromatic rings are positioned with respect to each other as well as the global molecular property of radius of gyration.

3.1.1 Angle and distance between aromatic rings

All lignin dimers in this study consist of two aromatic rings with a linkage in between, decorated with functional groups. Since the linkages are formed between different atoms in the monomer, the dimer structure varies. Structural features may be recognized simply by considering some basic characteristic properties such as.

- The number of atoms separating the rings.
- The angle Θ between the normal vectors of ring planes.
- The distance between the mass centers of the rings.

The number of atoms separating the rings is visualized in Figure 1. It should be noted that the smallest linkage is the 5-5′ linkage, directly connecting two ring carbons and the largest is the *β*-*β*′, having cyclic substructures between the aromatic rings that will affect its molecular flexibility. The *β*-O-4′ linkage is the largest linear linkage, allowing for many conformational options. Interestingly, in softwood, the most common linkages are the *β*-O-4′ (44–50 %) and 5-5′ (10–23 %) (Chang and Jiang 2020; Dellon et al. 2017; Erickson et al. 1973) which are clearly different in their structural nature and the way they contribute to the total flexibility of the polymer.

As a characteristic measure, the distribution of the cosine of the angle Θ between the planes of the aromatic rings is presented for all linkages in Figure 2. The direction of the normal vector is arbitrary, therefore absolute values were chosen. Values close to 0 indicate an orthogonal orientation whereas a co-planar orientation results in values close to 1. It should be noted, that either value can be the result of multiple different conformations of the dimer.

Figure 2A shows that angular distribution strongly depends on the linkage type. The shortest linkage (5-5′) has a preferred conformation of the rings at $|\cos(\Theta)| = 0.52$, corresponding to an angle between the normals that is either 58.4 or 121.6°, i.e., somewhere between orthogonal and co-planar orientation. The dimers with the 4-O-5′ and *α*-O-4′ linkages show a predominantly orthogonal orientation of the aromatic rings, where the larger *α*-O-4′ linkage allows for more variations than the smaller 4-O-5′. The *β*-1′ and *β*-O-4′ linkages both show a peak close to 1, i.e., co-planar preference. However, the larger *β*-O-4′ linkage also allows for a range of orientations, even orthogonal. The *β*-5′ is the only linkage showing two distinct peaks at both orthogonal and co-planar orientation, the orthogonal being more pronounced. The other cyclic linkage, the *β*-*β*′ dimer, appears to have no preferred angle, although there is an excess towards more orthogonal orientations.

As a complement to the angular distribution, the distance between the mass centers (COM distance) of the aromatic rings is also presented in Figure 2B. The two

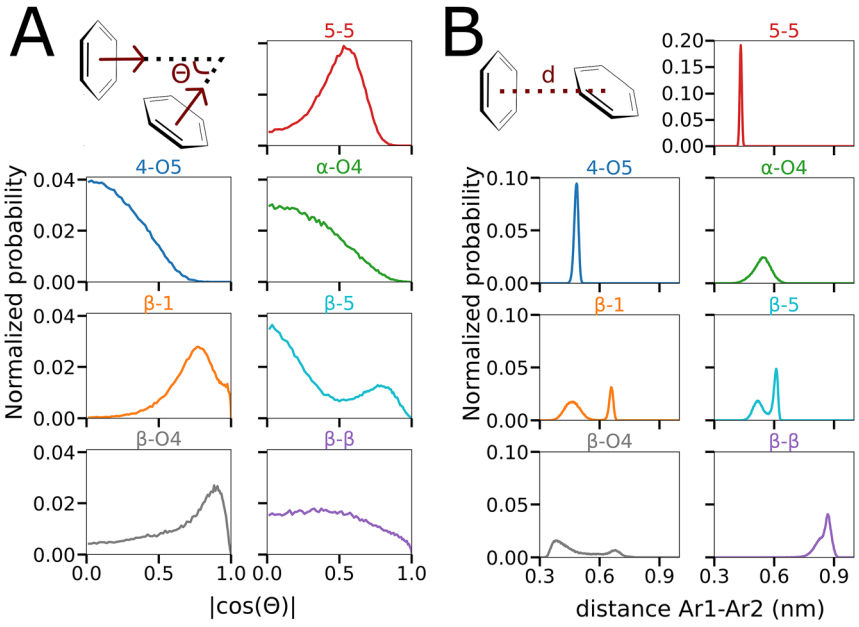

Figure 2: Probability distributions of angle and distance between aromatic ring planes. A: schematic representation of the measured angle Θ between the normal vectors of the planes, followed by the distribution of the absolute cosine of the angles for seven dimers with varying linkages. B: schematic representation of the distance d between the centers of mass of the aromatic rings (Ar1 and Ar2) and the measured distribution of the distances for the seven dimers. Note: due to the sharp peak in the distribution of the distance for the dilignol with the 5-5′ linkage, a different scale of the *y*-axis is used.

linkages that prefer orthogonal orientation (4-O-5′ and *α*-O-4′ linkages) have fairly short COM distances, whereas the co-planar linkages (*β*-1′ and *β*-O-4′) show distributions around one shorter and one larger value.

It is of interest to compare these results with experimental wide angle X-ray scattering (WAXS) data that is available for lignin materials. It has been shown by WAXS that *π*–*π* interactions between aromatic rings exist in lignin-based thermoset materials, and it was concluded that both sandwich and T-shaped interactions occur to different extent (Jawerth et al. 2020; Ribca et al. 2021). However, from experiments it cannot be distinguished whether these structural features occur inter- or intramolecularly. The distance between sandwich stacked lignin aromatic rings in the experimental study was found to be between 0.44 and 0.45 nm. In MD simulations using classical force-fields, quantum mechanics phenomena such as *π*–*π* stacking are implicit effects from the applied non-bonded interaction parameters, and may be detected from geometric criteria (Sherrill et al. 2009). The simulation results indicate that a co-planar conformation and a ring–ring distance within this range is in fact fulfilled by G-G dimers with *β*-O-4′ linkage. From these geometric criteria it can be concluded that there is a possibility that a high amount of *β*-O-4′ linkages will increase intramolecular *π*–*π* sandwich structural features in lignin based material. The dimer with the *β*-1′ linkage however has two major conformations, where each fulfills either the distance or the angle criteria, see Figure S1 in the Supplementary material.

3.1.2 Radius of gyration R_g

The R_g of the different dimers is presented in Figure 3 and it can be noted that all distributions have at least one peak between 0.4 and 0.5 nm and only a few have peaks where R_g < 0.4 nm. Interestingly, the two dimers that show co-planar orientation, the *β*-O-4′ and *β*-1′ linkages, are those with maxima at R_g < 0.4 nm. Both these linkages also prefer their conformations where R_g < 0.4 nm; *β*-1′ with 72 % and *β*-O-4′ with 62 % occupation. The *β*-5′ linkage with a cyclic structure also shows two distinct conformation populations, both with R_g > 0.4 nm. By evaluating time series of R_g (Figure S2 in the Supplementary material), it is apparent that the compact structures are stable between less than 1 ns (*β*-5′) and a few ns (*β*-O-4′) whereas the (*β*-1′) linkage remains in one conformation longer, in the order of 10 ns.

It is important to clarify that structural data obtained in this study represents single lignin dimers in water. Water is known to be a poor lignin solvent, a fact that is easily understood by considering its biological function to enable water to be transported over long distances including both transport through the lumen and the nanopores where lignin deposition was recently shown by 3D electron tomography (Fernando et al. 2023). On the other hand, pure water is indeed able to dissolve lignin that is comprised of G units and only *β*-O-4′ linkages (He et al. 2016), indicating an intricate relationship between lignin and water. The conformation of any polymer in dilute solution is highly dependent on solute-solvent interactions, and one particular parameter that is useful in

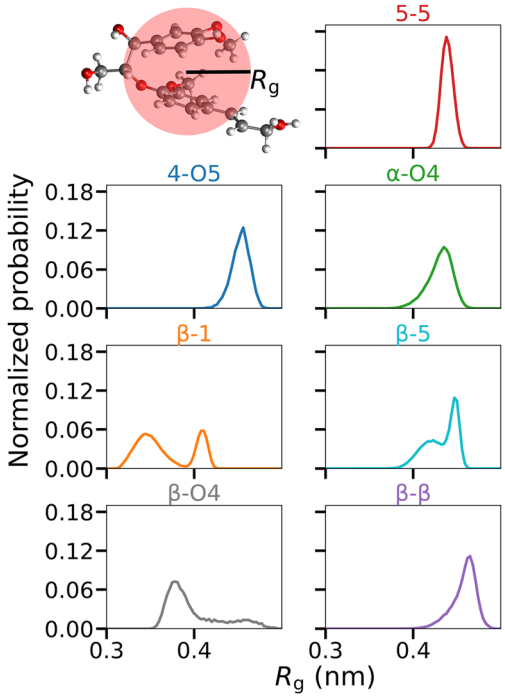

Figure 3: Graphical illustration of the radius of gyration (R_g) and R_g distributions for all seven dimers.

this context is the parameter shown here, radius of gyration, R_g. Radius of gyration provides us with information of the solute conformation, but is also a measure of solvent quality; poor, good, or Θ solvents according to Flory's theory of polymer solvation (Flory 1953).

The radius of gyration of larger softwood and hardwood lignin models (degree of polymerization DP = 61) have recently been examined by Jahan and coworkers (Jahan et al. 2022), proposing a scaling behavior for lignin with R_g as a function of monomer units. Comparison of these R_g values to the presented results gives some valuable insight. The dimer's R_g values of around 0.4–0.5 nm are in agreement with the published scaling behavior (number of monomers N = 2) of both softwood ($R_g \approx 0.5$ nm) and hardwood lignin ($R_g \approx 0.4$ nm) in pure water. Since their softwood polymer model consists of only G monomers with different linkages, where 50 % were β-O-4′, whereas the hardwood model has both G and S and also higher content of β-O-4′ (60 %) the presented results propose that it could be the higher β-O-4′ content that gives rise to the more compact hardwood lignin structure. Although the impact of having both G and S monomers cannot be determined from the presented data, it can be seen that some linkages are expected to have a larger influence on the varying radius of gyration observed at different temperatures and solvent conditions due to their observed flexibility.

3.2 Energetical analysis

Conformational preferences are governed by the free energy landscape, i.e., enthalpy or entropy, and differences in enthalpy between the different systems in this study is due to potential energy differences. Hence, characterizing the potential energy contributions and their relative magnitude is of interest and convenient since they are readily available from MD simulations. Analysis of hydrogen bonds, detected by geometric criteria, is followed by a complete picture of total potential energy contributions.

3.2.1 Hydrogen bonds

Each lignin dimer in the present study consists of two aromatic rings but also functional groups such as hydroxy and methoxy groups, some of the dimers also comprise ether oxygens. The presence of these enables hydrogen bonds (HB) to be formed, either intramolecularly within the lignin molecule or intermolecularly to the surrounding water.

The hydrogen bond concept was introduced a century ago (Gibb 2020) and has since then been essential for description of structure-property relationships in biological matter (Jeffrey 1991). Its importance for lignin structure is however less explored. Lignin is generally considered a hydrophobic polymer, although it is much less hydrophobic than biomolecules such as waxes and suberin due to the presence of the aforementioned groups (Henriksson 2017; Kunst and Samuels 2003; Pollard et al. 2008). However, hydrogen bonds in the context of lignin are mostly discussed from the perspective that it contributes to interaction between lignin and other cell wall components such as cellulose (Jiang et al. 2016), not specifically the role of HB:s for lignin macromolecular structure.

Given the heterogeneity of lignin polymers, it is likely that its hydrophobicity/hydrophilicity varies depending on structure and it is therefore of interest to investigate the effect different linkages have on HB ability of a lignin dimer. The results are presented in Table 1. The dimer with least HB options is the β-β′ dimer, also reflected in its average number of HB, where it is unable to form intramolecular HB:s and also has the lowest number of HB:s to water. At the other extreme, with in total 11 donors/acceptors, are the two dimers with ether bonds; α-O-4′ and β-O-4′. While α-O-4′ uses its HB ability to form intramolecular HB to a slightly larger extent, they both are prone to form 25 % more HB to water as compared to β-β′. In between the extremes the remaining four linkages are found and it is noted that those involving β-carbons (except β-O-4') rarely form intramolecular HB:s.

Table 1: HB ability (nr. of HB donors + acceptors) and time averages (with statistical errors) of intra- and intermolecular hydrogen bonds.

Type	*β*-*β*′	*β*-5′	4-O-5′	*β*-1′	5-5′	*α*-O-4′	*β*-O-4′
Donors + acceptors	8	9	9	10	10	11	11
Intra – average	0.00	0.00	0.18	0.08	0.20	0.47	0.23
Intra – error	0.001	0.007	0.001	0.001	0.001	0.002	0.001
Inter – average	7.59	8.77	8.16	9.50	8.72	9.59	9.91
Inter – error	0.005	0.005	0.005	0.005	0.006	0.005	0.005

3.2.2 Potential energy

Potential energy contributions were analyzed. The total potential energy of each system can be divided into contributions from bonded (bond, angular and dihedral potentials) and non-bonded (Coulomb and Lennard-Jones (LJ) potentials) energy terms respectively.

$$\begin{aligned} E_{\text{tot}} &= E_{\text{bonded}} + E_{\text{nonbonded}} \\ &= E_{\text{bonded}} + E^{\text{Coul}}_{\text{nonbonded}} + E^{LJ}_{\text{nonbonded}} \end{aligned} \quad (1)$$

All potential energy terms are provided in Table S1. The largest absolute contribution to the total energy of the system comes from water–water interaction, however these values are similar for all systems. The most interesting is to compare separate contributions due to interactions that are intramolecular within lignin (LIG-LIG) or intermolecular lignin-water interactions (LIG-SOL). Time averages of non-bonded contributions are therefore presented in Figure 4 (time series of both bonded and non-bonded contributions are provided in Figure S3). It can be seen that these contributions are either non-advantageous (positive) and advantageous (negative). The advantageous contributions are due to lignin-water interaction and the short-ranged (SR) intramolecular lignin interactions. The special type of intramolecular interaction between atoms that are also forming a dihedral (Coul-14 and LJ-14) are positive in all cases. It is clear that the differences in energetics between the systems are due to intramolecular Coulomb interactions within the lignin dimers (Coul-14:LIG-LIG and Coul-SR:LIG-LIG).

The differences in intramolecular HB:s (Table 1) is partly but not completely reflected in the intramolecular Coul-SR:LIG-LIG interactions. The *β*-O-4′ and *α*-O-4′ dimers are able to form intramolecular HB:s relatively often, and both have high intramolecular Coulomb energies. Also 5-5′ and 4-O-5′ have similar frequency of intramolecular HB and their corresponding Coulomb energies agree. The *β*-1′ occasionally forms intramolecular HB:s, also reflected by its Coulomb energy, and the *β*-5′ dimer with zero intramolecular HB:s has the lowest contribution from intramolecular Coulomb interactions, in fact the intermolecular LJ interactions for this dimer are more important. On the contrary, the *β*-*β*′ dimer, also with zero intramolecular HB, shows a relatively high intramolecular Coulomb energy. This could be attributed to the interaction between the ether oxygen and hydroxy endgroups, that are not covered in the applied geometric criteria for recognizing HB:s, this was confirmed when a wider definition (cutoff angle <50° and cutoff distance <0.5 nm) of the interaction was used. The magnitude of energetic penalty from Coul-14-terms seems to correlate with the magnitude of the advantageous Coul-SR interactions.

Altogether, the energetic picture that emerges is that the systems have similar potential energy contributions from all terms except the lignin-lignin intramolecular Coulomb interactions.

3.3 Water interfacial structure

Water affects lignin structure, but how does lignin affect structure of the surrounding water? Here, structural data concerning water structure in the vicinity of the lignin dimers is presented. Given the complex and combined hydrophobic and hydrophilic interactions between lignin and water, such properties are of interest.

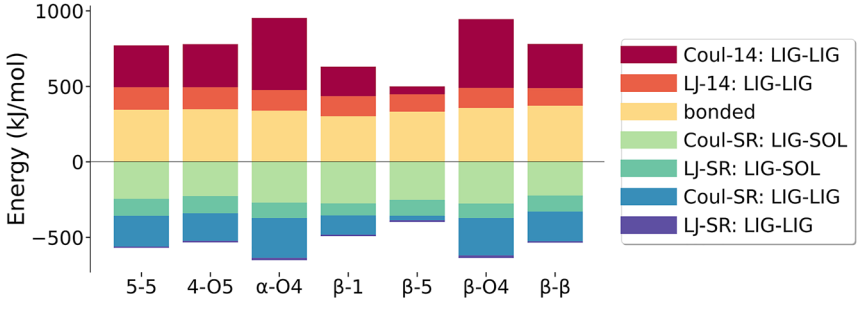

Figure 4: Potential energy contributions. Bonded term is for the entire system due to interactions between atoms that either in different molecules (intermolecular) or within the same molecule (intramolecular). Values are time averages, statistical error is small and can be found in Table S1. Time series of both bonded and non-bonded energy terms can be found in Figure S3.

3.3.1 Solvent accessible surface area (SASA)

The solvent accessible surface area (SASA), i.e., the surface characterized around the lignin dimer by probing its accessibility of a solvent sphere with van der Waals radius of 0.14 nm, representing water, is shown in Figure 5. Water has, for most of the linkage types (5-5′, 4-O-5′, *α*-O4′, *β*-5′), similar accessibility to the dimer surface. The SASA of these molecules are all similarly shaped normal distributions with maxima between 6 and 7 nm^2/molecule. The remaining three dimers are different. The *β*-O-4′ has a broad peak shifted toward a somewhat smaller area compared to the others, most likely an effect of its flexibility and presence of two different major conformations as was shown by both R_g and angle between planes (Figures 2 and 3). The *β*-1′ also showed co-planar conformations in Figure 2 and has a smaller area. The *β*–*β*′ dimer is not co-planar and is large in terms of R_g (Figure 3), but its accessibility for water is negatively affected by its cyclic linkage structure, this dimer also showed least interaction with water in terms of its number of HB (Table 1).

The SASA values for all dimers are significantly larger than an envelope surface of a sphere with radii set to R_g; for instance a sphere with radius R_g = 0.45 nm has an envelope surface of 2.5 nm^2. Although only dimers were included in this study, which have limited possibilities to modify its surface area, R_g and SASA of softwood and hardwood lignin polymers of DP 61 presented by Jahan et al. 2022 showed similar relation where their R_g of polymers in pure water corresponded to spheres with area only about 37 % of the SASA. Since SASA of a solute in water is connected to solvent entropy, a key property that in turn is related to hydrophobic effect (Caro et al. 2017), it is concluded that there seems to be a delicate balance between hydrophobicity and hydrophilicity governing the behavior of lignin in water.

3.3.2 Solute-solvent radial distribution functions (RDF)

To elucidate water structure in the vicinity of the lignin dimers, and enhance where it is most likely to find water, the radial distribution function (RDF) of water is analyzed with respect to different sites of the lignin molecules. The main purpose is to compare local water environment around the different dimers. To that end the following sites of lignin was considered: aromatic rings, ring hydroxy, ring methoxy and linkages, their respective definitions are depicted in Figure S4 in Supplementary material.

The resulting RDF:s are shown in Figure 6. It can be noted that water is similarly structured around the aromatic rings, regardless of linkage. However, the RDF corresponding to the

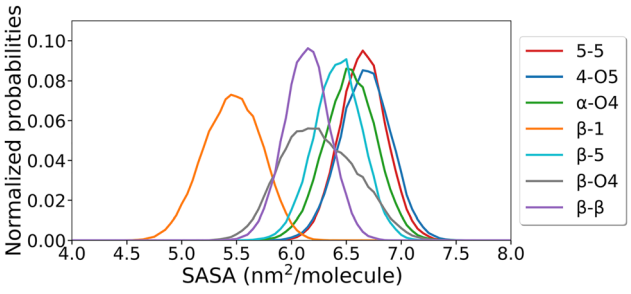

Figure 5: Solvent accessible surface area (SASA) distributions of the different dimers.

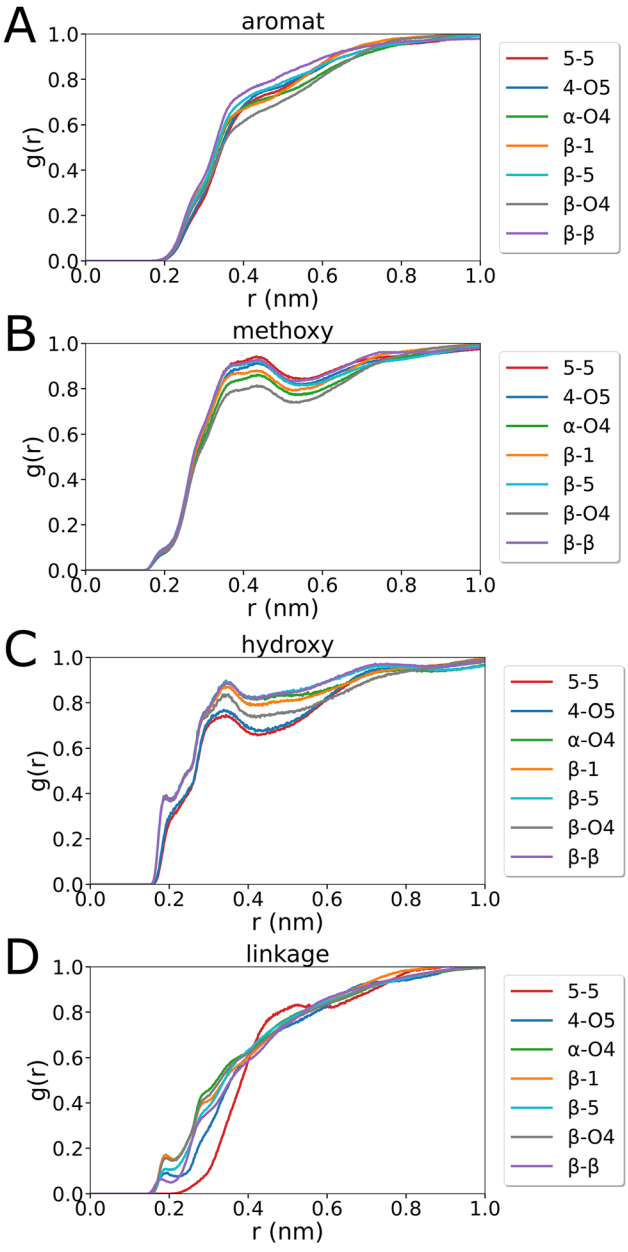

Figure 6: Radial distribution functions of water with respect to different functional groups within the lignin dimers. The specification of the functional groups is found in Supplementary material, Figure S4.

β-O-4′ dimer stands out, significantly lower than the other dimers between 0.35 and 0.50 nm. The same appears to be the case around the methoxy groups, β-O-4′ has again a less pronounced second hydration shell. Both of these features could be explained by the frequently found folded conformation of the β-O-4′ dimer. The depletion of water in the second solvation shell of the hydrophobic parts of β-O-4′ lignin in combination with its ability to alternate between a compact and an extended conformation are effects that may contribute to the experimentally observed solubility of G lignin linked by only β-O-4′ (He et al. 2016).

Solvation of the water attracting ring hydroxy groups show two distinct patterns depending on the location of the hydroxy group within the molecule, i.e., if it is close to the linkage (5-5′, 4-O-5′) or far from it (all the rest). Also here, the β-O-4′ RDF is lower at the second hydration shell compared to the other dimers with hydroxy groups far from the linkage.

Water structure around the actual linkages is obviously affected by their chemical differences, and especially the 5-5′ linkage stands out, being only a covalent bond between two ring carbons. All the other linkages have water in the vicinity to some extent, even β-1′, despite being a carbon-carbon linkage. Its flexibility (it frequently occupies co-planar conformation) and the proximity to linkage hydroxys render this possible. The β-O-4′ and β-1′ linkages indicate the most hydrated environments.

4 Conclusions

By Molecular Dynamics (MD) simulations of single lignin dimers with varying linkages in water as model systems, this study provides a systematic investigation of the role of linkage on structural and energetical properties of lignin as well as surrounding water.

This study shows that the linkage does significantly affect both conformational preferences of lignin, water structure around lignin, and lignin-water interaction. The most striking result is the behavior of the β-O4′ linked dimer, which shows apparently different properties compared to the other linkages, both in terms of conformational freedom and its interaction with water. Within the β-O4′ dimer, direct interaction between the rings is possible due to a folded conformation (π–π stacking) and the molecule shows two common conformations, one compressed and one extended. The β-O4′ dimer also has a unique effect on the surrounding water, which is important for lignin solubility. This is especially interesting considering it is the most prevalent linkage in softwood and hardwood lignin and predominantly consumed upon common delignification reactions, where lignin molecules are de-polymerized. In future studies, it would be of interest to investigate the effect of varying β-O4′ content in longer lignin polymers, and its consequences for solubility in different solvents including aqueous systems with varying pH and ionic strength found in large scale lignin processing.

Acknowledgment: The authors would like to acknowledge colleagues within the research school of Resource Smart Processes (https://www.resurssmarta.se/) for support and discussion.
Research ethics: Not applicable.
Author contributions: The authors have accepted responsibility for the entire content of this manuscript and approved its submission.
Competing interests: The authors state no conflict of interest.
Research funding: The authors would like to acknowledge Vinnova (project number: 2021-02086) for funding and the National Academic Infrastructure for Supercomputing in Sweden (NAISS, former SNIC) for the provided computational resources under projects SNIC 2022/22-1163 and 2022/22-989.
Data availability: The raw data can be obtained on request from the corresponding author.

References

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., and Lindahl, E. (2015). Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2: 19–25.

Aguilera-Segura, S.M., Bossu, J., Corn, S., Trens, P., Mineva, T., Le Moigne, N., and Di Renzo, F. (2019). Synergistic sorption of mixed solvents in wood cell walls: experimental and theoretical approach. *Macromol. Symp.* 386: 1900022.

Aminzadeh, S., Lauberts, M., Dobele, G., Ponomarenko, J., Mattsson, T., Lindström, M.E., and Sevastyanova, O. (2018). Membrane filtration of kraft lignin: structural characteristics and antioxidant activity of the low-molecular-weight fraction. *Ind. Crops Prod.* 112: 200–209.

Baptista, C., Robert, D., and Duarte, A.P. (2006). Effect of pulping conditions on lignin structure from maritime pine kraft pulps. *Chem. Eng. J.* 121: 153–158.

Baptista, C., Robert, D., and Duarte, A.P. (2008). Relationship between lignin structure and delignification degree in pinus pinaster kraft pulps. *Bioresour. Technol.* 99: 2349–2356.

Berlin, A. and Balakshin, M. (2014). Industrial lignins: analysis, properties, and applications (chapter 18). In: Bioenergy Research. Elsevier B.V., Amsterdam, pp. 315–336.

Besombes, S., Robert, D., Utille, J.-P., Taravel, F.R., and Mazeau, K. (2003). 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. *Holzforschung* 57: 266–274.

Boerjan, W., Ralph, J., and Baucher, M. (2003). Lignin biosynthesis. *Annu. Rev. Plant Biol.* 54: 519–546.

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126: 014101.

Capanema, E.A., Balakshin, M.Y., and Kadla, J.F. (2004). A comprehensive approach for quantitative lignin characterization by nmr spectroscopy. *J. Agric. Food Chem.* 52: 1850–1860.

Caro, J.A., Harpole, K.W., Kasinath, V., Lim, J., Granja, J., Valentine, K.G., Sharp, K.A., and Wand, A.J. (2017). Entropy in molecular recognition by proteins. *Proc. Natl. Acad. Sci. U.S.A.* 114: 6563–6568.

Chang, H.-M. and Jiang, X. (2020). Biphenyl structure and its impact on the macromolecular structure of lignin: a critical review. *J. Wood Chem. Technol.* 40: 81–90.

Ciesielski, P.N., Pecha, M.B., Lattanzi, A.M., Bharadwaj, V.S., Crowley, M.F., Bu, L., Vermaas, J.V., Steirer, K.X., and Crowley, M.F. (2020). Advances in multiscale modeling of lignocellulosic biomass. *ACS Sustain. Chem. Eng.* 8: 3512–3531.

Crestini, C., Melone, F., Sette, M., and Saladino, R. (2011). Milled wood lignin: a linear oligomer. *Biomacromolecules* 12: 3928–3935.

Dang, B.T., Brelid, H., and Theliander, H. (2016). The impact of ionic strength on the molecular weight distribution (mwd) of lignin dissolved during softwood kraft cooking in a flow-through reactor. *Holzforschung* 70: 495–501.

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh ewald: an n ·log(n) method for ewald sums in large systems. *J. Chem. Phys.* 98: 10089–10092.

Dean, J.C., Walsh, P.S., Biswas, B., Ramachandran, P.V., and Zwier, T.S. (2014). Single-conformation UV and IR spectroscopy of model g-type lignin dilignols: the β–o–4 and β-β linkages. *Chem. Sci.* 5: 1940.

Dellon, L.D., Yanez, A.J., Li, W., Mabon, R., and Broadbelt, L.J. (2017). Computational generation of lignin libraries from diverse biomass sources. *Energy Fuels* 31: 8263–8274.

Erickson, M., Larsson, S., Miksche, G.E., Wiehager, A.-C., Lindgren, B.O., and Swahn, C.-G. (1973). Gaschromatographische analyse von ligninoxydationsprodukten. VIII. Zur struktur des lignins der fichte. *Acta Chem. Scand.* 27: 903–914.

Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L.G. (1995). A smooth particle mesh ewald method. *J. Chem. Phys.* 103: 8577–8593.

Fernando, D., Kowalczyk, M., Guindos, P., Auer, M., and Daniel, G. (2023). Electron tomography unravels new insights into fiber cell wall nanostructure; exploring 3d macromolecular biopolymeric nano-architecture of spruce fiber secondary walls. *Sci. Rep.* 13: 2350.

Flory, P.J. (1953). Principles of polymer chemistry. Cornell University Press, Ithaca, NY.

Gandini, A. (2008). Polymers from renewable resources: a challenge for the future of macromolecular materials. *Macromolecules* 41: 9491–9504.

Gibb, B.C. (2020). The centenary (maybe) of the hydrogen bond. *Nat. Chem.* 12: 665–667.

Gierer, J. (1980). Chemical aspects of kraft pulping. *Wood Sci. Technol.* 14: 241–266.

He, T., Jiang, Z., Wu, P., Yi, J., Li, J., and Hu, C. (2016). Fractionation for further conversion: from raw corn stover to lactic acid. *Sci. Rep.* 6: 38623.

Henriksson, G. (2017). What are the biological functions of lignin and its complexation with carbohydrates? *Nord. Pulp Pap. Res. J.* 32: 527–541.

Hess, B. (2008). P-lincs: a parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* 4: 116–122.

Humphrey, W., Dalke, A., and Schulten, K. (1996). Vmd: visual molecular dynamics. *J. Mol. Graphics* 14: 33–38.

Hunter, J.D. (2007). Matplotlib: a 2d graphics environment. *Comput. Sci. Eng.* 9: 90–95.

Jahan, N., Huda, M.M., Tran, Q.X., and Rai, N. (2022). Effect of solvent quality on structure and dynamics of lignin in solution. *J. Phys. Chem. B* 126: 5752–5764.

Jawerth, M.E., Brett, C.J., Terrier, C., Larsson, P.T., Lawoko, M., Roth, S.V., Lundmark, S., and Johansson, M. (2020). Mechanical and morphological properties of lignin-based thermosets. *ACS Appl. Polym. Mater.* 2: 668–676.

Jeffrey, G.A. (1991). *Hydrogen bonding in biological structures*, Study ed. Springer Berlin Heidelberg, Berlin, Heidelberg.

Jiang, Z., Zhang, H., He, T., Lv, X., Yi, J., Li, J., and Hu, C. (2016). Understanding the cleavage of inter- and intramolecular linkages in corncob residue for utilization of lignin to produce monophenols. *Green Chem*. 18: 4109–4115.

Jönsson, J., Pettersson, K., Berntsson, T., and Harvey, S. (2013). Comparison of options for utilization of a potential steam surplus at kraft pulp mills – economic performance and CO_2 emissions. *Int. J. Energy Res.* 37: 1017–1035.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79: 926–935.

Karlsson, M., Giummarella, N., Lindén, P.A., and Lawoko, M. (2020). Toward a consolidated lignin biorefinery: preserving the lignin structure through additive-free protection strategies. *ChemSusChem* 13: 4666–4677.

Kenrick, P. and Crane, P.R. (1997). The origin and early evolution of plants on land. *Nature* 389: 33–39.

Kunst, L. and Samuels, A.L. (2003). Biosynthesis and secretion of plant cuticular wax. *Prog. Lipid Res.* 42: 51–80.

Lawoko, M., Berglund, L., and Johansson, M. (2021). Lignin as a renewable substrate for polymers: from molecular understanding and isolation to targeted applications. *ACS Sustain. Chem. Eng.* 9: 5481–5485.

Mattsson, C., Hasani, M., Dang, B., Mayzel, M., and Theliander, H. (2017). About structural changes of lignin during kraft cooking and the kinetics of delignification. *Holzforschung* 71: 545–553.

Nawawi, D.S., Syafii, W., Akiyama, T., and Matsumoto, Y. (2016). Characteristics of guaiacyl-syringyl lignin in reaction wood in the gymnosperm gnetum gnemon l. *Holzforschung* 70: 593–602.

Öhman, F., Wallmo, H., and Theliander, H. (2007). Precipitation and filtration of lignin from black liquor of different origin. *Nord. Pulp Pap. Res. J.* 22: 188–193.

Parrinello, M. and Rahman, A. (1981). Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* 52: 7182–7190.

Petridis, L. and Smith, J.C. (2016). Conformations of low-molecular-weight lignin polymers in water. *ChemSusChem* 9: 289–295.

Pollard, M., Beisson, F., Li, Y., and Ohlrogge, J.B. (2008). Building lipid barriers: biosynthesis of cutin and suberin. *Trends Plant Sci*. 13: 236–246.

Ragauskas, A.J., Beckham, G.T., Biddy, M.J., Chandra, R., Chen, F., Davis, M.F., Davison, B.H., Dixon, R.A., Gilna, P., Keller, M., et al. (2014). Lignin valorization: improving lignin processing in the biorefinery. *Science* 344: 1246843.

Ralph, J., Lundquist, K., Brunow, G., Lu, F., Kim, H., Schatz, P.F., Marita, J.M., Hatfield, R.D., Ralph, S.A., Christensen, J.H., et al. (2004). Lignins: natural polymers from oxidative coupling of 4-hydroxyphenyl-propanoids. *Phytochem. Rev.* 3: 29–60.

Ribca, I., Jawerth, M.E., Brett, C.J., Lawoko, M., Schwartzkopf, M., Chumakov, A., Roth, S.V., and Johansson, M. (2021). Exploring the effects of different cross-linkers on lignin-based thermoset properties and morphologies. *ACS Sustain. Chem. Eng.* 9: 1692–1702.

Rinaldi, R., Jastrzebski, R., Clough, M.T., Ralph, J., Kennema, M., Bruijnincx, P.C.A., and Weckhuysen, B.M. (2016). Paving the way for lignin

valorisation: recent advances in bioengineering, biorefining and catalysis. *Angew. Chem., Int. Ed.* 55: 8164–8215.

Santos, R., Capanema, E., Balakshin, M., Chang, H.-m., and Jameel, H. (2011). Effect of hardwoods characteristics on kraft pulping process: emphasis on lignin structure. *BioResources* 6: 3623–3637.

Sherrill, C.D., Sumpter, B.G., Sinnokrot, M.O., Marshall, M.S., Hohenstein, E.G., Walker, R.C., and Gould, I.R. (2009). Assessment of standard force field models against high-quality ab initio potential curves for prototypes of π–π, ch/π, and sh/π interactions. *J. Comput. Chem.* 30: 2187–2193.

Timell, T.E. (1983). Origin and evolution of compression wood. *Holzforschung* 37: 1–10.

van der Walt, S., Colbert, S.C., and Varoquaux, G. (2011). The numpy array: a structure for efficient numerical *computation*. *Comput. Sci. Eng.* 13: 22–30.

Vermaas, J.V., Hardy, D.J., Stone, J.E., Tajkhorshid, E., and Kohlmeyer, A. (2016). Topogromacs: automated topology conversion from charmm to gromacs within vmd. *J. Chem. Inf. Model.* 56: 1112–1116.

Vermaas, J.V., Crowley, M.F., and Beckham, G.T. (2019a). A quantitative molecular atlas for interactions between lignin and cellulose. *ACS Sustain. Chem. Eng.* 7: 19570–19583.

Vermaas, J.V., Petridis, L., Ralph, J., Crowley, M.F., and Beckham, G.T. (2019b). Systematic parameterization of lignin for the charmm force field. *Green Chem*. 21: 109–122.

Waskom, M. (2021). Seaborn: statistical data visualization. *J. Open Source Softw.* 6: 3021.

Yanez, A.J., Li, W., Mabon, R., and Broadbelt, L.J. (2016). A stochastic method to generate libraries of structural representations of lignin. *Energy Fuels* 30: 5835–5845.

Zhu, W. and Theliander, H. (2015). Precipitation of lignin from softwood black liquor: an investigation of the equilibrium and molecular properties of lignin. *BioResources* 10: 1696–1714.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/hf-2023-0054).