Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Polymer Principles Behind Solubilizing Lignin With Organic Cosolvents for Bioenergy†

Derya Vural,[a,b,c] Jeremy C. Smith,[a,b] and Loukas Petridis*[a,b]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Lignin solubilization is key to a viable biorefinery because its removal leads to facile deconstruction of biomass and because the isolated lignin can serve to derive precursors of novel high-value materials. The mixing of organic solvents with water has been shown to improve biomass fractionation and lignin conversion reactions. However, generally-applicable solubilization strategies are lacking because of the remarkable variability of lignin across plant feedstocks. Here, to obtain a predictive understanding of lignin solvation, we perform molecular dynamics simulations of model lignin polymers in two mixtures of water with polar aprotic solvents: tetrahydrofuran (THF):water and γ-valerolactone (GVL):water. The model lignins include *H*-, *G*- and *S*-only homopolymers and lignins with an *S:G* 1:1 ratio. We find that a well-established theory of self-avoiding polymers in a "good" solvent describes accurately the physical conformations of all types of lignin in both solvents. As the degree of methoxy substitution increases in the homopolymers, the distributions of the lignin radius of gyration and the Flory exponent ν, which describes the lignin-solvent interactions, do not change in THF:water, while ν shifts to slightly higher values in GVL:water. We attribute this increase to the interaction between the methyl group of GVL with the lignin methoxy groups. We also find that the reduction in the lignin radius of gyration due to branching is accurately described by the Zimm-Stockmayer theory for both THF:water and GVL:water. The above findings validate the applicability of polymer physics concepts to lignin and suggest that GVL:water may have the most favorable interaction with *S*-lignin, whereas the interactions of THF:water with lignin are independent of lignin monomeric content.

1 Introduction

Realizing the full potential of the utilization of lignocellulosic biomass for low-carbon energy and material supply requires understanding how to efficiently break down and fractionate its carbohydrates and lignin.[1–3] Solubilizing lignin during biomass deconstruction is therefore key for the production of biofuels and lignin-based bioproducts.[4] Not only does lignin impede enzymatic biomass degradation,[5–7] but the development of value-added products from lignin is considered as critical for the economic sustainability of a biorefinery.[8] After lignin removal, hemicelluloses and cellulose can be converted into liquid transportation biofuels,[9–11] while the solubilized lignin can serve as a valuable precursor of carbon fibers, colloids and thermoplastics.[12–16]

Various polar aprotic solvents have been employed to solubilize lignocellulose,[17] such as sulfolane, dimethyl sulfoxide, 1,4-dioxide, tetrahydrofuran (THF) and γ-valerolactone (GVL).[18–22] In particular, the use of THF has been shown to reduce biomass recalcitrance to deconstruction and to increase lignin fractionation.[23–27] THF:water exhibits efficiency for the solubilization of the intermediates and products, and shifting reaction equilibrium, thereby significantly improving the yield and selectivity of target products.[28] Mixing THF and water enables multiple processes that synergistically improve biomass fractionation by acting on all three major biomass components. THF-water is a good solvent for lignin,[29,30] leading to the dissociation of lignin from itself and from cellulose.[2] Further, the addition of THF to water significantly enhances cellulose deconstruction and lignocellulose solubilization[31] and also slows the rate of xylan hydrolysis, allowing an economically desirable "single-pot" conversion of both xylan and cellulose to fuel precursors.[32]

GVL has also been used for the production of chemicals from both the carbohydrates and the lignin of biomass.[33–36] Similar to

[a] *Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, Knoxville, TN 37996, USA. E-mail: petridisl@ornl.gov*
[b] *UT/ORNL Center for Molecular Biophysics, Oak Ridge National Laboratory, P.O. Box 2008, Tennessee 37831.*
[c] *Department of Physics, Giresun University, Giresun, 28200, Turkey*

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 00.0000/00000000.

‡ Additional footnotes to the title and authors can be included *e.g.* 'Present address:' or 'These authors contributed equally to this work' as above using the symbols: ‡, §, and ¶. Please place the appropriate symbol next to the author's name and include a `\footnotetext` entry in the the correct place in the list.

Table 1 The chemical composition of lignin-like molecules. The primary sequence of each molecules is shown in Tables S-2 to S-7s

Molecule	Units	Linkages
H-lignin	*H* (100%)	$\beta-O-4$ (100%)
G-lignin	*G* (100%)	$\beta-O-4$ (100%)
G-lignin (*1-branch*)	*G*(100%)	$\beta-O-4$ (94%) $5-5$ (6%)
S/G-lignin (*alternating*)	*G* (50%) *S* (50%)	$\beta-O-4$ (100%)
S/G-lignin (*random*)	*G* (50%) *S* (50%)	$\beta-O-4$ (88%) $\beta-5$ (12%)
S-lignin	*S* (100%)	$\beta-O-4$ (100%)

THF, GVL also acts on all major biomass components. Fractionation of biomass by GVL isolates the solid cellulose fraction, dissolves hemicellulose and solubilizes lignin. Further, GVL improves the reaction rates and product selectivity of the acid-catalyzed conversion of xylose into furfural.[37] The highly-pure lignin fraction can be depolymerized to its monomeric units for further upgrading.[34]

Lignin is made primarily from the polymerization of three subunits, p-hydroxyl phenol (*H*), guaiacyl (*G*) and syringyl (*S*), that differ in the degree of methoxylation of their aromatic ring. Further, there are suggestions that lignin architecture may not be linear as it may contain branch points: monomer units that are covalently linked to three other units. However, both the exact type of branch points and the existence of branches are disputed.[38,39] Lignin's chemical structure is irregular and heterogenous, varying between different plant species and even across layers of the same cell wall.[40] Perhaps because of this heterogeneity, a simple description of the fundamental physical chemistry of the interaction of lignin with aprotic organic pretreatment solvents is lacking. Establishing such principles should provide significant conceptual impetus to the important field of bioenergy and lead to rational improvement of fractionation conditions.

Here, we examine whether established concepts in polymer theory are applicable to lignin. One way to quantify the interactions between lignin and a solvent is via the concept of the quality of the solvent developed by Flory[30,41], which connects the relative strength of the intra-polymer and polymer-solvent interactions to the configuration the polymer adopts. In a "poor" solvent intra-polymer interactions are favored and the polymer adopts collapsed configurations. In a "theta" solvent the interactions are exactly balanced, leading to a random coil configuration in which the distribution of the radius of gyration of the configuration of a polymer is a Gaussian. In a "good" solvent the polymer-solvent interactions are favored and a self-avoiding configuration, consistent with excluded volume interactions, is found. We also test whether the Zimm-Stockmayer theory, which describes the conformations of branched polymers in a "theta" solvent and predicts their radius of gyration as a function of the length of the branches,[42] is valid for lignin in aprotic solvents.

We investigate by molecular dynamics (MD) simulations the solvation of lignin polymers, with different *S/G/H* compositions, in three solvents: water, THF:water and GVL:water. We find that the size and shape of lignin is similar in THF:water and GVL:water, apart from *S*-lignin in GVL:water which adopts more elongated structures than the other systems. Similarly, THF:water and GVL:water are found to be "good" solvents for all the lignins. The lignin solvent interactions, quantified by the Flory exponent, become slightly more favorable with increasing lignin methoxy substitution in GVL:water, but do not vary in THF:water.

2 Methods

2.1 Lignin Models

Six lignin polymers were constructed. Three linear homopolymers were built, whose units all connected with $\beta-O-4$ linkages, of

(i) *H*-only units

(ii) *G*-only units

(iii) *S*-only units

(iv) A branched homopolymer of *G*-only units was connected with $\beta-O-4$ linkages and a 5-5 linkage at a "Y" branch point.

Two linear polymers were built that contain *S/G* units in 1/1 ratio:

(v) one with alternating *S* and *G* units connected with $\beta-O-4$ linkages called here "*S*&*G*-lignin (alternating)"

(vi) the other with *S* and *G* units randomly distributed along the lignin sequence, and connected with a linkage distribution representative of *Populus trichocarpa* with 88% $\beta-O-4$ and 12% $\beta-5$, called here "*S*&*G*-lignin (random)".

All lignin polymers comprised 18 units, with a molecular weight of $\sim$ 4 kDa. The unit and linkage compositions can be found in Table 1, and the primary sequence of each molecules is shown in Tables S-6 to S-11. Each of the six models was solvated in three solvents: water, THF:water 1:1 v/v, and GVL:water 1 : 1 v/v. The three homopolymers (*H*-only, *G*-only and *S*-only) were also simulated in pure THF and pure GVL.

2.2 MD Simulations

Molecular simulations are commonly employed to study polymer solubilization. citechen2012,Mananghaya MD simulations were performed using GROMACS 5.2[43] at 297 K. We employed the CHARMM force field for lignin[44], THF and GVL[45,46], and the TIP3P[47] model for water. The Van der Waals interactions were truncated at 11 Å, and the electrostatic interactions represented using the Particle Mesh Ewald method[48] with a real-space cutoff of 11 nm. All bonds including hydrogen atoms were constrained

Table 2 The time-averaged radius of gyration (R_g) of lignin molecules in five solvents (water, THF:water, GVL:water, pure THF and pure GVL)

Molecules	water	THF:water	GVL:water	pure THF	pure GVL
H-lignin	9.56 ± 0.26	14.61 ± 0.75	13.83 ± 0.49	12.89 ± 0.79	17.91 ± 0.42
G-lignin	10.31 ± 0.39	14.45 ± 0.33	14.56 ± 0.52	13.71 ± 0.61	18.84 ± 0.40
G-lignin (*1-branch*)	10.05 ± 0.32	12.87 ± 0.51	12.27 ± 0.56		
S/G-lignin (*alternating*)	10.81 ± 0.37	15.22 ± 0.35	14.83 ± 0.54		
S/G-lignin (*random*)	10.43 ± 0.74	14.79 ± 0.57	13.99 ± 0.60		
Averaged *S*-lignin *G*-lignin	10.75 ± 0.45	14.28 ± 0.70	14.87 ± 0.77		
S-lignin	11.19 ± 0.50	14.11 ± 1.08	15.18 ± 1.01	13.51 ± 1.25	19.54 ± 0.72

with a linear constraints solver algorithm (LINCS)[49]. The cutoff distance for nearest neighbors was 11 Å and the neighbor list was updated every 20 steps. The Nose-Hoover algorithm[50,51] with a coupling time of 2 ps and the Parrinello-Rahman algorithm[52] with a coupling time of 1 ps were used for the temperature and pressure coupling, respectively. Typical delignification reactions employ 160 °C for 15 min for THF:water, but 120 °C for 120 min for GVL:water[22]. Here we are interested in the solubility of lignin at room temperature, after delignification reactions of biomass have taken place and the lignin has been cooled down for processing.

The energy of the system was first minimized using 50,000 steepest-descent steps. The system was then equilibrated in the NPT ensemble at 1 bar for 5 ns. To improve statistics, three independent 500 ns production simulations were performed of each model, starting with a different velocity distribution. The coordinates were saved every 1 ps. The integration step was 2 fs. All calculations were performed on the Edison supercomputer at NERSC.

2.3 Analysis of MD Simulations

Analysis was performed using the last 350 ns of each simulation employing GROMACS 5.2[43] and Virtual Molecular Dynamics (VMD)[53]. The radius of gyration (R_g) and the asphericity (Δ)

$$\Delta = \langle \frac{(L_1 - L_2)^2 + (L_1 - L_3)^2 + (L_2 - L_3)^2}{2(L_1 + L_2 + L_3)^2} \rangle \qquad (1)$$

where L_1, L_2 and L_2 are the eigenvalues of the radius of gyration tensor of the lignin molecule and $\langle \cdot \rangle$ represents a time average, were computed by using the *gmx rdf* and *gmx polystat* GROMACS tools. The $\beta - O - 4$ dihedral angle was taken between the C_α, C_β atoms of first monomer and O_4', C_4' atoms of second monomer, and computed by using the *gmx angle* GROMACS tool. Error estimates were calculated using block averaging, by using the *gmx analyze* GROMACS tool.

3 Results

3.1 Conformations of lignin molecules

MD simulations were performed of six lignin molecules (Table 1) with different *S/G/H* ratios in three solvents environments (water, THF:water and GVL:water). The R_g for all lignin molecules is larger in THF:water and GVL:water, $R_g \approx 15$ Å, than in water, $R_g \approx 10$ Å (Table 2, Figures 1 and S-1). Representative snapshots from the simulations indicate that lignin molecules adopt collapsed structures in water, whereas their structure becomes coil-like in the presence of THF and GVL (Figure 2).[30,32,54,55]

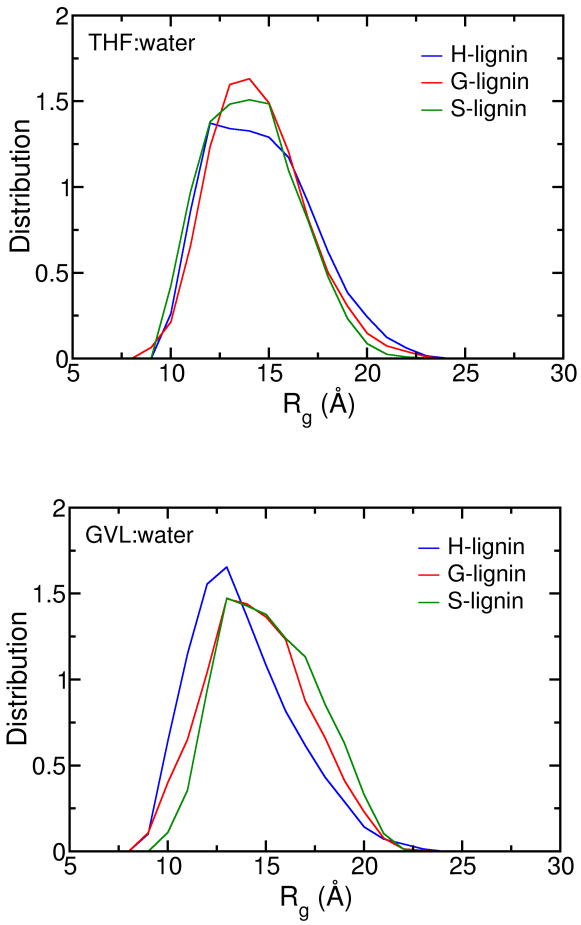

Fig. 1 The probability distribution of R_g for the three homopolymers in the two solutions.

Although the R_g of all lignin molecules are statistically similar in the two co-solvents, there is an interesting trend (Figure 1). As the degree of methoxy substitution increases in the homopolymers (*H*-lignin has the lowest and *S*-lignin the highest), the distribution of R_g remains almost the same in THF:water, while it shifts to slightly higher R_g values in GVL:water. Overall, the most extended conformations are found for *S*-lignin in GVL:water. The size of lignin molecules is found to have an additive dependence on monomer composition: the R_g of lignins with 1:1 *S:G* ratios

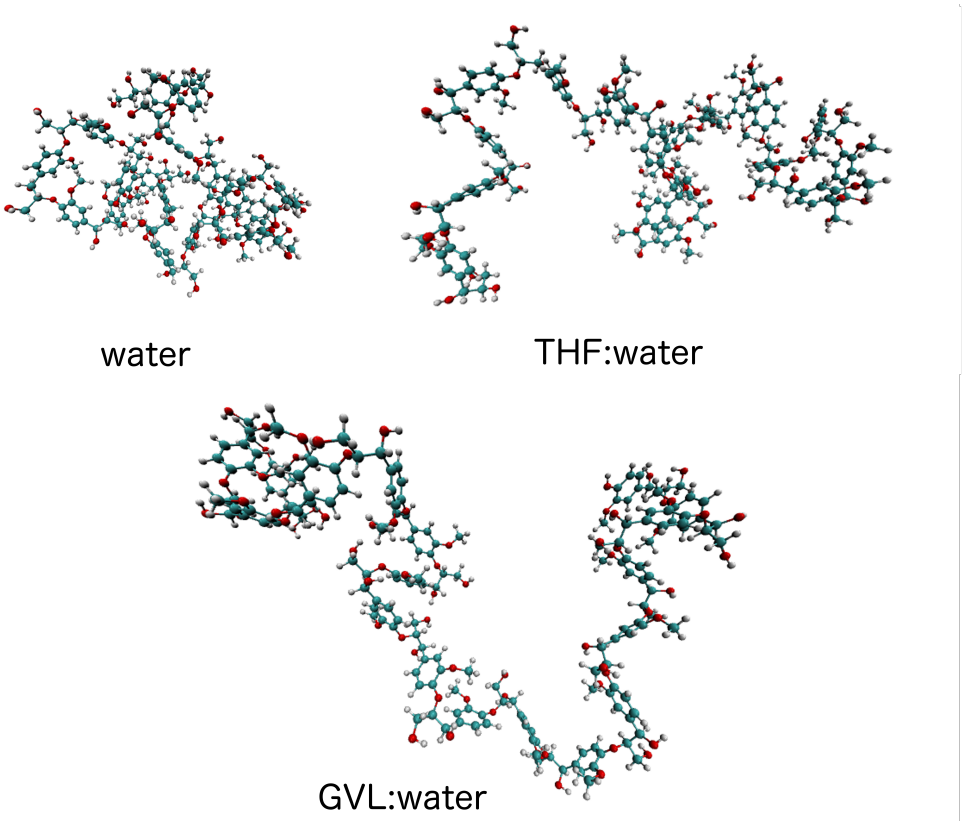

Fig. 2 *G*-lignin in water, THF:water and GVL:water, which has the R_g in Figure 1.

is similar to the average of the R_g of *G*-lignin and *S*-lignin homopolymers (Table 2). We also find that *S/G*-lignin (*alternating*) and *S/G*-lignin (*random*) have similar R_g. Thus the R_g of *S/G*-lignin does not depend on the precise sequence of its monomer units, in agreement with a previous report on vanilla lignin.[56]

The R_g of lignin in pure GVL is considerably larger than in pure THF (Table 2). Addition of water reduces slightly, not-statistically significant, the Rg of lignin in pure GVL, but increases the Rg in pure THF. Therefore, assuming the Rg is a measure of the quality of a solvent, the presence of water does not alter significantly the quality of THF, whereas water makes GVL a "worse" solvent for lignin. This may explain why addition of water reduces the solubility of enzymatic hydrolysis lignin in GVL.[57]

While branching does not change the size of the *G*-lignin in water, the R_g of the branched *G*-lignin is smaller than its linear counterpart in the organic solvents. The theory of Zimm and Stockmayer (ZS) for polymers in theta solvents[42] predicts the ratio of the R_g^2 of a branched molecule to that of its linear counterpart to be

$$g_{ZS} = \frac{R^2_{g,branched}}{R^2_{g,linear}} = \sum_{\mu=1}^{3} \left(\frac{3N_\mu^2}{N^2} - \frac{2N_\mu^3}{N^3}\right) \quad (2)$$

where N_μ is the number of monomers in one of the branches μ, and $N = N_1 + N_2 + N_3$ is the total number of monomers. Here, the branch lengths are $N_\mu = 5, 6, 7$ and $N = 18$, thus using Eq. 2 we obtain a theoretical prediction of $g_{ZS} = 0.78$. The ratio of the R_g calculated from the simulation trajectories yields $g_{ZS} = 0.79 \pm 0.07$ for THF:water, and $g_{ZS} = 0.71 \pm 0.08$ for GVL. Thus, the ZS theory is applicable to lignin in both THF:water and GVL:water co-solvent solutions.

The shape of the lignin polymer is described here by the asphericity parameter, Δ, which varies between $\Delta = 0$, corresponding to a spherical shape, and $\Delta = 1$, for a rod-like shape. Similar trends to those in R_g are observed, consistent with more spherical conformations having smaller R_g (Figure 3): the lignins are most spherical in water whereas the largest asphericity is found for *S*-lignin in GVL, while the average asphericity of *G*- and *S*-lignins matches the *S:G* 1:1 lignin. Further, addition of a branch point makes the lignin more spherical. Although, the structure of lignin polymer varies with the solvent, the distribution of the $\theta = C_\alpha - C_\beta - O_4 - C_4$ dihedral angles in the $\beta - O - 4$ linkages is almost independent of the solvent (Figure S-2). The distribution becomes more broadened with increasing the number of methyl groups (Table S-1).

To test which polymer model is applicable to lignin, the distribution of the end-to end distance, $P(r)$ was calculated for each lignin molecule and fitted by the following function[58,59]

$$P(r) = Lr^{\frac{1}{1-\nu}} \exp(-Kr^{\frac{1}{1-\nu}}) \quad (3)$$

where K and L are constants and ν is the Flory parameter. For an ideal chain in a "theta" solvent, $\nu = 1/2$ and $P(r)$ is a Gaussian function. For a self-avoiding chain in a "good" solvent $\nu > 1/2$. We fitted Eq. 3 to the $P(r)$ calculated from the MD trajectories and obtained values of ν indicative of "good" solvent conditions (Figure 4 and Table 3). Further the trend in ν matches that of the

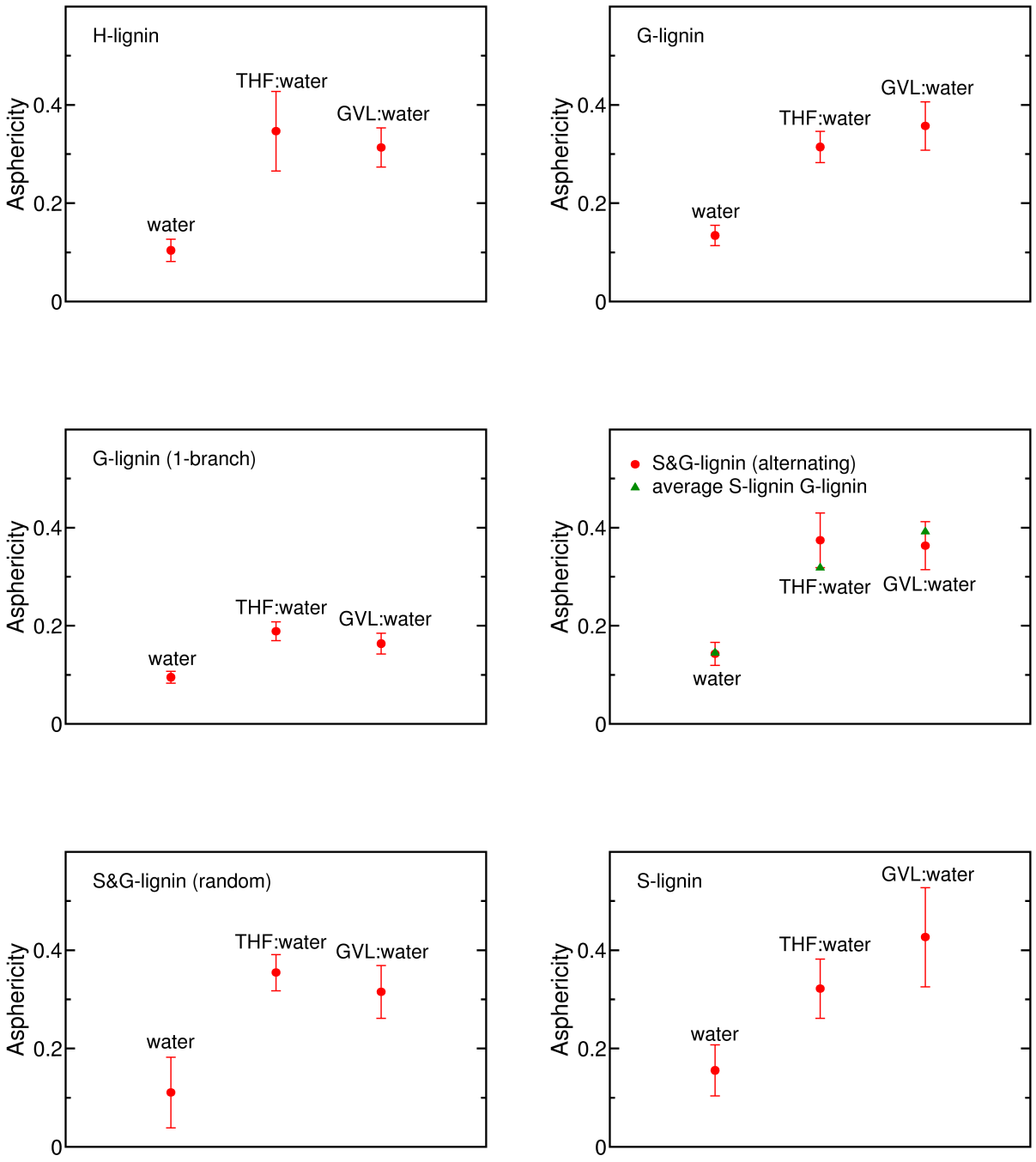

Fig. 3 The asphericity (Δ) of six lignin molecules in three solvents; water, THF:water and GVL:water. The asphericity for each lignin molecules is represented by red dots with error bars, and green dots represent the averaged of the asphericity of *G*- and *S*-lignins. The larger error bars for THF:water and GVL:water is observed due to the cross over from spherical to aspherical form.

R_g: For THF:water, all lignin homopolymers are found to have $\nu = 0.61$, while for GVL:water ν increases with the degree of methoxylation of lignin. We note that constraining $\nu = 1/2$ to obtain a Gaussian $P(r)$ does not provide a good fit to the data (Figure 4).

The conformation of a polymer can also be characterized by the dependence of the radius of gyration, R_n of a segment between monomers i and j, on the number of monomers $n = |i - j|$ that make up the segment (Figure 5). Renormalization-group arguments predict that a polymer in a "good" solvent R_n follows an

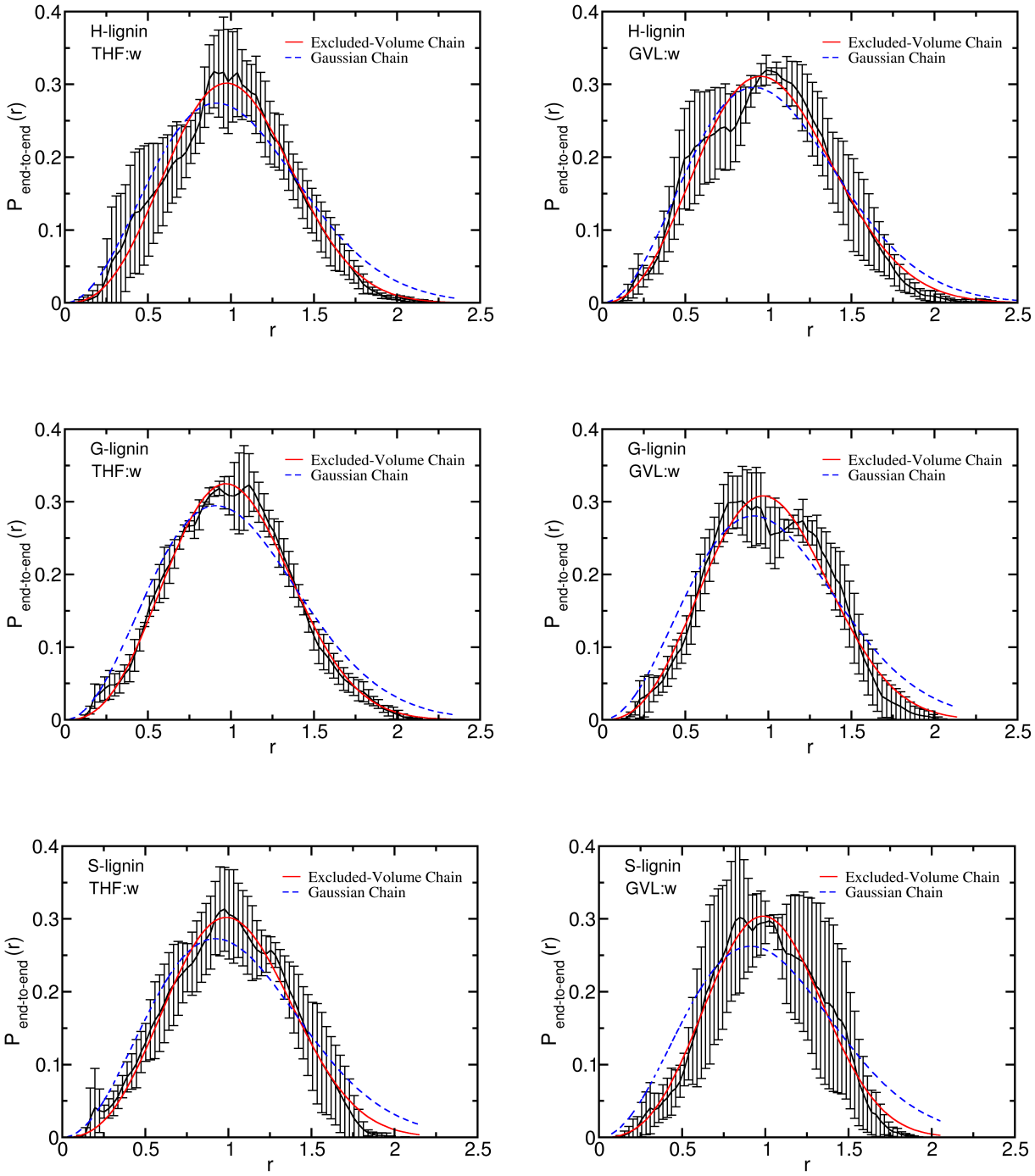

Fig. 4 The fit of the end-to-end distribution function $P(r)$ in Eq. 3 to the MD data (black). ν is a free parameter in the solid red line, whereas $\nu = 1/2$ in the dotted blue line and $P(r)$ is a Gaussian.

Table 3 Values of the Flory parameter ν obtained by fitting Eq. 3 to Figure 4, and Eq. 4 to Figure 5.

THF:water	Eq. 3	Eq. 4	**GVL:water**	Eq. 3	Eq. 4
H-lignin	0.61	0.61 ± 0.01	**H-lignin**	0.56	0.59 ± 0.01
G-lignin	0.61	0.60 ± 0.01	**G-lignin**	0.60	0.61 ± 0.01
S-lignin	0.61	0.60 ± 0.01	**S-lignin**	0.64	0.65 ± 0.01

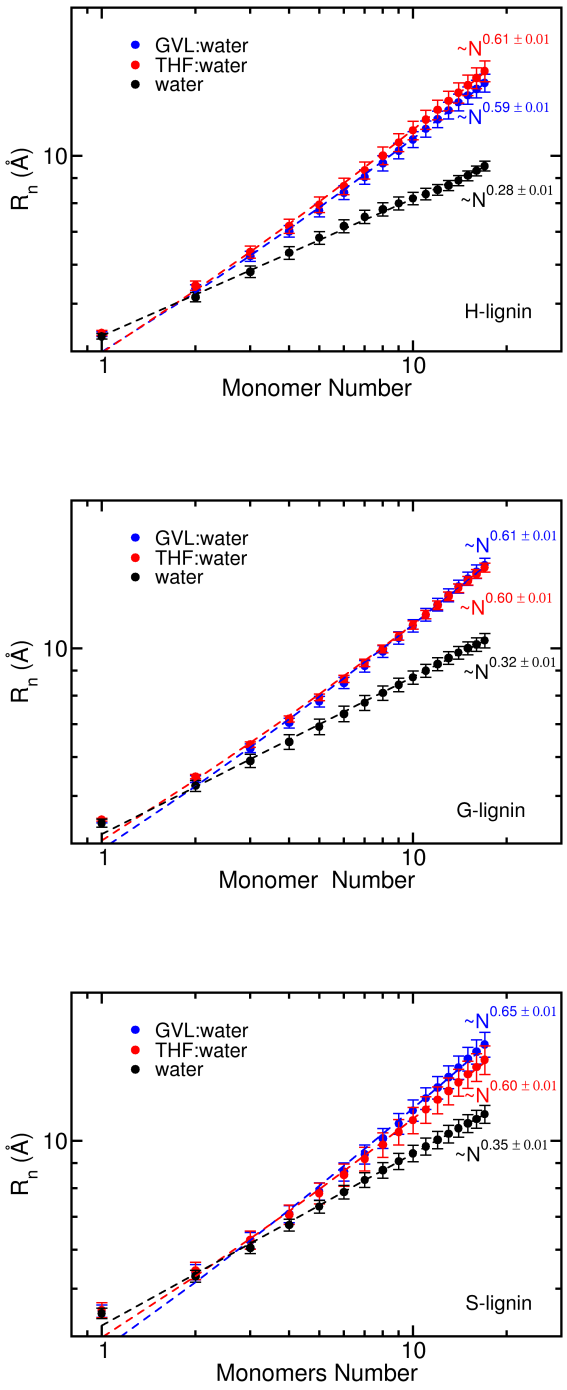

Fig. 5 The radius of gyration of a segment of a lignin molecule vs. the number of monomers in that segment.We note that R_g in Figure 1 corresponds to $R_n(n=18)$.

expansion in powers of n[58] :

$$R_n = An^{\nu}(1+\frac{B}{n^{\beta}}+\cdots)^{1/2}. \quad (4)$$

The so-called critical exponents ν and β are universal, i.e. do not depend on the specifics of the polymer, and are predicted to be $\beta = 0.56$ and $\nu = 0.588$ for an excluded volume chain. We fitted Eq. 4 to the simulation data, using A, B and ν as free parameters, while keeping $\beta = 0.56$ constant (Figure 5). In the case of water, we took $B = 0$. In water $\nu \simeq 1/3$, a value indicative of "bad" solvent conditions[30]. In THF:water and GVL:water, all lignins have $\nu \sim 0.6$, which corresponds to a "good" solvent conditions. The interaction energies of lignin with THF:water and GVL:water are similar (Table S5). This finding is consistent with both cosolvents having a similar ν, as the Flory exponent is determined by the solvent-polymer effective interactions. A good agreement is found between the ν values obtained from the two independent fits to Eqs. 3 and 4, thus strongly suggesting lignin in aprotic solvents behaves as a a self-avoiding polymer in "good" solvent conditions.

3.2 Interactions between lignin and the solvent

The solvation of the lignins was characterized by calculating the radial distribution function $g(r)$ of the organic solvents (THF and GVL) around the lignin molecules (Figure 6). In general, the more well-defined the first peak in $g(r)$ is, the more structured the solvation shell. Similarly, the higher the first peak intensity is, the larger the local density of the solvent. We therefore observe that THF has a more structured solvation shell than GVL (Figures 6a, b), and that increasing the degree of lignin methoxylation decreases the first solvation shell density for both solvents.

A difference between the two solvents studied here is the presence of a methyl group in GVL. The methyl group of GVL is found to have high probability of being near the methyl groups of *S*- and *G*-lignin (Figure 7). This preferential interaction between the methyl groups of GVL and lignin may explain why the structural properties of lignin depend on the degree of its methoxylation only when it is solvated in GVL, but not when solvated in THF, which does not have a methyl group.

3.3 Discussion

The development of effective techniques for converting lignin to valuable products is key for sustainable biorefinery schemes[60]. This requires developing separation technologies that obtain high purity lignin as well as downstream processes that upgrade lignin to novel chemicals and materials[61]. The use of organic solvents, such as THF and GVL, may be critical in both steps. These solvents are employed to fractionate biomass and during the catalytic conversion of lignin. However, to fully realize the potential to valorize lignin, we must overcome the challenges posed by its the chemical complexity. Lignin is a heteropolymer that lacks a well-defined primary structure and whose average chemical composition may vary between cells of the same plant[62]. One approach to overcoming this challenge is to provide physico-chemical principles that are broadly applicable to lignin.

The results presented here demonstrate that, despite its heterogeneous chemical structure, well-established polymer theories do describe lignin solvation in polar aprotic solvents. Specifically, lignin can be accurately modeled by Flory's description of a self-avoiding polymer, and aprotic cosolvents are "good" solvents for lignin. Previously, we have reported THF:water to be a "theta"

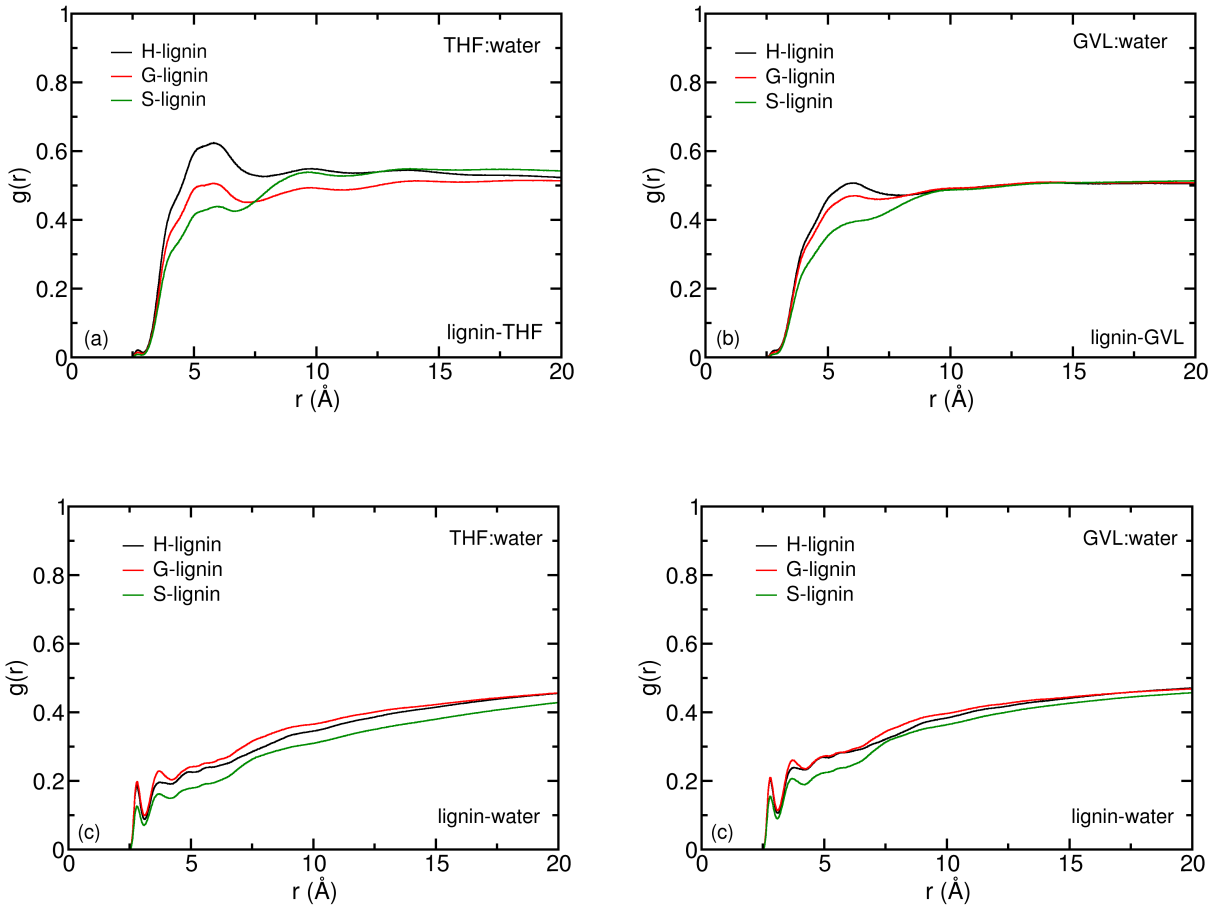

Fig. 6 The radial distribution function of non-hydrogen atoms of water and co-solvents (THF and GVL) around non-hydrogen atoms of lignin. See also Figure S-4. Distributions are normalized at large distances to their partial volumes.

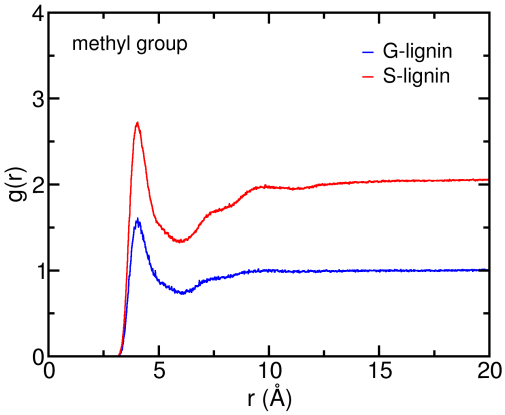

Fig. 7 The radial distribution function of methyl group of co-solvent, GVL, around methyl groups of lignin

solvent for lignin[29,30], a conclusion that was based on fitting R_n in Figure 5 assuming an infinitely long polymer chain, *i.e.* assuming $B = 0$ in Eq. 4. The approach we employ here may be better suited for oligomeric lignin. In addition to the lignin conformational properties studied here, it has been shown that the lignin glass transition dynamics can also be described by concepts developed for chemically homogeneous polymers[63,64]. We thus suggest that despite the chemical complexity and variability of lignin, the physical properties of lignin follow theoretical predictions representative of simple polymers.

Lignin in THF:water adopts similar conformations independent of the degree of methoxylation of its subunits. In contrast, in GVL:water, a solvent which contains a methyl group, increasing the lignin methoxylation leads to more extended conformations. This is due to the interaction of GVL methyl groups with lignin methoxyl groups. In general, the more extended the configuration of a polymer in solution, the more soluble the polymer will be in that solvent. Therefore, our results suggest that THF:water will have similar performance in solubilizing lignin, independent of its chemical composition, and that GVL:water might be best employed to solubilize lignin rich in *S* units, such as lignin isolated from Poplar.

Although the conformations of lignin are similar in the two solvent systems, the distribution of the solvents around lignin differs. THF has a more structured solvation shell, as evidenced by

the sharper features of the radial distribution function. The less ordered GVL solvation shell indicates entropy may play a bigger role in solvation of lignin by GVL than by THF. Given that GVL is bulkier than THF and that S units are bulkier than H units, there is a trend towards the more bulky the lignin units or the solvent, the less ordered is the lignin solvation shell.

In this work, we have examined the physical properties of lignin. However, the heterogeneity of lignin poses challenges in its chemical upgrading that have not been considered here. For example, depolymerization of lignin lacks specificity, resulting in diverse products that require expensive separation[65]. Further, the *S/G* ratio of the monomers produced by catalytic depolymerization of lignin show only a weak correlation with the *S/G* ratio of the biomass feedstock[66]. Moreover, the heterogeneity of lignin chemistry is thought to be one of the main reasons lignin-based carbon fiber is currently impractical for commercial uses[67]. Local solvent effects have been shown to be important for the chemical conversion of biomass feedstocks[68–72]. It would be helpful to the bioenergy community if future studies investigate the interdependence between physical effects at play in lignin dissolution. We also note that the recalcitrance of untreated biomass to deconstruction is influenced by many factors, some of which are not described here, such as cellulose degree of polymerization, cellulose accessibility and total lignin content.[73,74].

The intermolecular interactions in lignin make it difficult to solubilize in common solvents unless it has been acetylated.[75] However, ionic liquids can efficiently dissolve lignin, making them ideal solvents for the processing of biomass. The cations play an important role, stabilizing lignin's aromatic rings and hydroxyl groups[76], although hydrogen-bonds between anions and the lignin have also been shown to be important for lignin solubilization.[77]

4 Conclusions

We performed MD simulations of lignin models with different *S/G/H* ratios in two polar aprotic solvents: THF:water and GVL:water with 1 : 1 volume ratio. Overall, we find that lignin conformations can be described by the well-established theory of a self-avoiding polymer in a "good" solvent. The size of lignin homopolymers is largely independent of the *S/G/H* ratios in THF:water, but increases with the degree of lignin methoxyl substitution in GVL:water. This behavior is reflected in the Flory exponent, which is $\nu \simeq 0.60$ in THF:water for all lignins, but varies slightly in GVL: $\nu = 0.58, 0.61, 0.65$ for *H*-lignin, *G*-lignin and *S*-lignin, respectively. The methoxyl groups seem to have an "additive" effect on the structure of lignin, in the sense that the properties of a lignin with a *S:G* 1:1 ratio lie in between those of *S*-lignin and *G*-lignin. Assuming the solubility of isolated lignin is primarily determined by its interaction with the solvent, our findings suggest that use of GVL:water may be a judicious choice to enhance solubilization of isolated lignin rich in S-units, whereas in THF:water the lignin solubility is independent of its composition.

Conflicts of interest

"There are no conflicts to declare".

Acknowledgements

This research was supported by the Genomic Science Program, Office of Biological and Environmental Research, U.S. Department of Energy (DOE), under Contract FWP ERKP752. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for DOE under Contract DE-AC05-00OR22725.

- Supporting of Information: More analysis data supporting the results presenting in this paper and the detailed information about the structure of lignin models.

Notes and references

1 T. R. Brown, *Bioresour. Technol.*, 2015, **178**, 166–176.

2 A. S. Patri, M. Barmak, Y. Pu, N. Ciaffone, M. Soliman, M. D. Smith, R. Kumar, X. Cheng, C. E. Wyman, L. Tetard, A. J. Ragauskas, J. C. Smith, L. Petridis and C. M. Cai, *J. Am. Chem. Soc.*, 2019, **141**, 12545–12557.

3 A. Grossman and W. Vermerris, *Curr. Opin. Biotechno.*, 2019, **56**, 112–120.

4 P. P. Thoresen, L. Matsakas, U. Rova and P. Christakopoulos, *Bioresource Technology*, 2020, **306**, 123189.

5 M. McCann and N. Carpita, *J. Exp. Bot.*, 2015, **66**, 4109–4118.

6 J. V. Vermaas, L. Petridis, X. Qi, B. Schulz, R. Lindner and J. C. t. Smith, *Biotechnol. Biofuels*, 2015, **8**, 1–16.

7 H. Li, Y. Pu, R. Kumar, A. J. Ragauskas and C. E. Wyman, *Biotechnol. Biofuels*, 2014, **111(3)**, 485–92.

8 R. van Rijn, I. U. Nieves, K. T. Shanmugam, L. O. Ingram and W. Vermerris, *Bioenerg. Res.*, 2018, **11**, 414–425.

9 B. Yang and C. E. Wyman, *Biofuels, Bioprod. Biorefin.*, 2008, **2**, 26–40.

10 M. E. Himmel, S. Y. Ding, D. K. Johnson, W. S. Adney, M. R. Nimlos, J. W. Brady and T. D. Foust, *Science*, 2007, **315**, 804–807.

11 C. Somerville, H. Youngs, C. Taylor, S. C. Davis and S. P. Long, *Science*, 2010, **329**, 790–792.

12 D. A. Baker, N. C. Gallego and F. S. Baker, *J. Appl. Polym. Sci.*, 2012, **124**, 227–234.

13 S. Kubo and J. F. Kadla, *Macromolecules*, 2004, **37**, 6904–6911.

14 S. L. Hilburg, A. N. Elder, H. Chung, R. L. Ferebee, M. R. Bockstaller and N. R. Washburn, *Polymer*, 2014, **55**, 995 – 1003.

15 M. H. Sipponen, M. Smyth, T. Leskinen, L.-S. Johansson and M. Osterberg, *Green Chem.*, 2017, **19**, 5831–5840.

16 E. Ten and W. Vermerris, *J. Appl. Polym. Sci.*, 2015, **132**,.

17 M. A. Mellmer, C. Sanpitakseree, B. Demir, K. Ma, W. A. Elliott, P. Bai, R. L. Johnson, T. W. Walker, B. H. Shanks, R. M. Rioux, M. Neurock and J. A. Dumesic, *Nature Comm.*, 2019, **10**, 1132.

18 B. R. Caes and R. T. Raines, *Chem. Sus. Chem.*, 2011, **4**, 353–356.

19 J.-M. Robinson, *US patent*, 2012, **8**, 263,792 B2.

20 V. Nikolakis, S. H. Mushrif, B. Herbert, K. S. Booksh and D. G. Vlachos, *J. Phys. Chem. B*, 2012, **116**, 11274–11283.

21 L.-K. Ren, L.-F. Zhu, T. Qi, J.-Q. Tang, H.-Q. Yang and C.-W. Hu, *ACS Catal.*, 2012, **7**, 2199–2212.

22 X. MEng, S. Bhagia, Y. Wang, Y. Zhou, Y. Pu, J. R. Dunlap, L. Shuai, A. J. Ragauskas and C. G. Yoo, *Industrial Crops and Products*, 2020, **146**, 112144.

23 T. Y. Nguyen, C. M. Cai, R. Kumar and C. E. Wyman, *Chem. Sus. Chem.*, 2015, **8**, 1716–1725.

24 Z. C. Jiang, T. He, J. M. Li and C. W. Hu, *Green Chem.*, 2014, **16**, 4257–4265.

25 C. M. Cai, T. Y. Zhang, R. Kumar and C. E. Wyman, *Green Chem.*, 2013, **15**, 3140–3145.

26 C. M. Cai, N. Nagane, R. Kumar and C. E. Wyman, *Green Chem.*, 2014, **16**, 3819–3829.

27 C. M. Cai, T. Y. Zhang, R. Kumar and C. E. Wyman, *J. Chem. Technol. Biotechnol.*, 2014, **89**, 2–10.

28 J. Li, W. Zhang, S. Xu and C. Hu, *Frontiers in Chemistry*, 2020, **8**, 70.

29 M. D. Smith, B. Mostofian, X. Cheng, L. Petridis, C. M. Cai, C. E. Wyman and J. C. Smith, *Green Chem.*, 2016, **18**, 1268–1277.

30 L. Petridis and J. C. Smith, *Nat. Rev. Chem.*, 2018, **2**, 382–389.

31 B. Mostofian, C. M. Cai, M. D. Smith, L. Petridis, X. Cheng, C. E. Wyman and J. C. Smith, *J. Am. Chem. Soc.*, 2016, **138**, 10869–10878.

32 M. D. Smith, C. M. Cai, X. Cheng, L. Petridis and J. C. Smith, *Green Chem.*, 2018, **20**, 1612–1620.

33 J. S. Luterbacher, J. M. Rand, D. M. Alonso, J. Han, J. T. Youngquist, C. T. Maravelias, B. F. Pfleger and J. A. Dumesic, *Science*, 2014, **343**, 277–280.

34 J. S. Luterbacher, A. Azarpira, A. H. Motagamwala, F. Lu, J. Ralph and J. A. Dumesic, *Energ. Environ. Sci.*, 2015, **8**, 2657–2663.

35 D. M. Alonso, S. G. Wettstein, M. A. Mellmer, I. Gurbuz, E., J. T. Youngquist, C. T. Maravelias, B. F. Pfleger and J. A. Dumesic, *Energy Environ. Sci.*, 2013, **6**, 76–80.

36 D. M. t. Alonso, *Sci. Adv.*, 2017, **3**, e1603301.

37 M. A. t. Mellmer, *Angew. Chem. Int. Edit*, 2014, **53**, 11872–11875.

38 J. Ralph, C. Lapierre and W. Boerjan, *Curr. Opin. Biotechnol.*, 2019, **56**, 240–249.

39 C. Crestini, F. Melone, M. Sette and R. Saladino, *Curr. Opin. Biotechnol.*, 2011, **12**, 3928–3935.

40 Y. Mottiar, R. Vanholme, W. Boerjan, J. Ralph and S. D. Mansfield, *Curr. Opin. Biotechnol.*, 2016, **37**, 190–200.

41 P. J. Flory, *"Biopolymers : Statistical mechanics of chain molecules*", Newyork: Wiley, 1969.

42 B. H. Zimm and W. H. Stockmayer, *J. Chem. Phys.*, 1949, **17**, 1301–1314.

43 M. J. Abraham, T. Murtola, R. Schulz, S. Pall, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1-2**, 19–25.

44 L. Petridis and J. C. Smith, *J. Comp. Chem.*, 2009, **30**, 457–467.

45 I. Vorobyov, M. Anisimov, S. Greene, R. M. Venable, A. Moser, R. W. Pastor and A. D. MacKerell, *J. Chem. Theory Comput.*, 2007, **3**, 1120–1133.

46 H. Lee, R. M. Venable, A. D. MacKerell and R. W. Pastor, *Biophys. J.*, 2008, **95**, 1590–1599.

47 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Pastor, *J. Chem. Phys.*, 1983, **79**, 926.

48 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577.

49 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comp. Chem.*, 1997, **18**, 1463.

50 G. J. Martyna, D. J. Tobias and M. L. Klein, *J. Chem. Phys.*, 1994, **101**, 4177–4189.

51 W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695.

52 M. Parrinello and A. Rahman, *J. App. Phys.*, 1981, **52**, 7182.

53 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.

54 L. Petridis, R. Schulz and J. C. Smith, *J. Am. Chem. Soc.*, 2011, **133**, 20277–20287.

55 L. Petridis and J. C. Smith, *Chem. Sus. Chem.*, 2016, **9**, 289–295.

56 T. B. Rawal, M. Zahran, B. Dhital, O. Akbilgic and L. Petridis, *Biochim. Biophys. Acta*, 2020, **1864**, 129547.

57 Z. Xue, X. Zhao, R.-c. Sun and T. Mu, *ACS Sustainable Chemistry & Engineering*, 2016, **4**, 3864–3870.

58 J. S. Valleau, *J. Chem. Phys.*, 1996, **104**, 3071.

59 C. Domb, J. Gillis and G. Wilmers, *Proceedings of the Physical Society*, 1965, **85**, 625–645.

60 H. Wang, Y. Pu, A. Ragauskas and B. Yang, *Biores. Tech.*, 2019, **271**, 449–461.

61 R. Rinaldi, R. Jastrzebski, M. T. Clough, J. Ralph, M. Kennema, P. C. A. Bruijnincx and B. M. Weckhuysen, *Agewandte Chemie*, 2016, **55**, 8164–8215.

62 G. A. Tuskan, W. Muchero, T. A. Tschaplinski and A. J. Ragauskas, *Curr. Opin. in Biotech.*, 2019, **56**, 250–257.

63 D. Vural, C. Gainaru, H. O'Neill, Y. Pu, M. D. Micholas Dean Smith, J. M. Parks, S. V. Pingali, E. Mamontov, B. H. Davison, A. P. Sokolov, A. J. Ragauskas, J. C. Smith and L. Petridis, *Green Chemistry*, 2018, **20**, 1602–1611.

64 D. Vural, J. C. Smith and L. Petridis, *Phys. Chem. Chem. Phys.*, 2018, **20**, 20504–20512.

65 Z. Xu, P. Lei, R. Zhai, Z. Wen and M. Jin, *Biotechnology for Biofuels*, 2019, **12**, 32.

66 E. M. Anderson, M. L. Stone, R. Katahira, M. Reed, W. Muchero, K. J. Ramirez, G. T. Beckham and Y. Román-Leshkov, *Nature Comm.*, 2019, **10**, 2033.

67 J. S. Yuan, Q. Li and A. J. Ragauskas, *Tappi Journal*, 2017, **16**, 107.

68 T. W. Walker, A. K. Chew, H. X. Li, B. Demir, Z. C. Zhang, G. W. Huber, R. C. Van Lehn and J. A. Dumesic, *Energy Environ. Sci.*, 2018, **11**, 617–628.

69 V. Vasudevan and S. H. Mushrif, *RSC Adv*, 2015, **5**, 20756–20763.

70 J. J. Varghese and S. H. Mushrif, *React. Chem. Eng.*, 2019, **4**, 165–206.
71 X. Wang and R. Rinaldi, *Chem. Sus. Chem.*, 2012, **5**, 1455–66.
72 L. Shuai, M. Questell-Santiago and J. S. Luterbacher, *Green Chem.*, 2016, **18**, 937.
73 X. Z. t. Meng, *Chem. Sus. Chem.*, 2017, **10**, 139–150.
74 C. G. Yoo, A. Dumitrache, W. Muchero, J. Natzke, H. Akinosho, M. Li, R. M. Sykes, S. D. Brown, B. Davison, G. A. Tuskan, Y. Pu and A. J. Ragauskas, *ACS Sustainable Chem. Eng.*, 2018, **6**, 2162–2168.
75 W. Zhao, L.-P. Xiao, G. Song, R.-C. Sun, L. He, S. Singh, B. A. Simmons and G. Cheng, *Green Chemistry*, 2017, **19**, 3272–3281.
76 J. Zubeltzu, E. Formoso and E. Rezebal, *Journal of Molecular Liquids*, 2020, **303**, 112588.
77 Y. Zhang, H. He, K. Dong, M. Fan and S. Zhang, *RSC Adv.*, 2017, **7**, 12670–12681.