

Cosolvent Pretreatment in Cellulosic Biofuel Production: Effect of Tetrahydrofuran-Water on Lignin Structure and Dynamics

Micholas Dean Smith^{a,b}, Barmak Mostofian^a, Xiaolin Cheng^{a,b}, Loukas Petridis^a, Charles M. Cai^{c,d,e}, Charles E. Wyman^{d,e} and Jeremy C. Smith^{a,b*}

^a Center for Molecular Biophysics, University of Tennessee/Oak Ridge National Laboratory, Oak Ridge, TN, 37830 USA

^b Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, M407 Walters Life Sciences, 1414 Cumberland Avenue, Knoxville, TN 37996

^c Center for Environmental Research and Technology (CE-CERT), Bourns College of Engineering, University of California, Riverside, 1084 Columbia Avenue, Riverside USA

^d Department of Chemical and Environmental Engineering, Bourns College of Engineering, University of California, Riverside, 446 Winston Chung Hall, 900 University Ave., Riverside, USA

^e BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, TN 37830 USA

Abstract

The deconstruction of cellulose is an essential step in the production of ethanol from lignocellulosic biomass. However, the presence of lignin hinders this process. Recently, a novel cosolvent based biomass pretreatment method called CELF (Cosolvent Enhanced Lignocellulosic Fractionation) which employs tetrahydrofuran (THF) in a single phase mixture with water, was found to be highly effective at solubilizing and extracting lignin from lignocellulosic biomass and achieving high yields of fermentable sugars. Here, using all-atom molecular-dynamics simulation, we find that THF preferentially solvates lignin, and in doing so, shifts the equilibrium configurational 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 “theta” 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 cellulytic enzymes to cellulose.

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>).

*Corresponding Author

Jeremy C. Smith

Center for Molecular Biophysics

University of Tennessee/Oak Ridge National Laboratory

Oak Ridge, TN, 37830 USA

Phone: 1-865-574-9635/591-4805

Fax: 1-865-576-7651

Email: smithjc@ornl.gov

INTRODUCTION

The production of ethanol for biofuel use in the United State has been hitherto primarily from first generation (corn/food crop) sources¹. However, land-use requirements for growing corn are inefficient, which is detrimental to its long-term use as a primary source for biofuels¹. An important alternative to corn is non-food lignocellulosic biomass (i.e., cellulosic sources). This, however, brings its own challenges, as the technical (and thus economic) cost associated with the chemical and biological deconstruction of this class of feedstock into the basic components needed for fuel production is significantly higher than for first generation crops. As such, there is much interest in finding novel, economically viable, methods to enhance the conversion of lignocellulosic biomass to platform fuel precursors amenable for conversion into renewable liquid fuels²⁻⁵.

Lignocellulosic biomass consists of three primary cell wall components: cellulose, hemicellulose, and lignin, of which hemicellulose and cellulose have been considered the most economically useful, as they are the most amenable to conversion into ethanol⁶. In contrast, lignin plays a natural role of protecting cellulose from chemical and biological breakdown and is thus an agent limiting the economic hydrolysis of cellulose by enzymes to fermentable glucose for ethanol production^{7, 8}. Moreover, besides the protection it provides to cellulose, lignin may also be taken as a potentially viable precursor for the production of non-ethanol biofuels and higher value chemicals if it can be efficiently extracted from biomass^{9, 10}.

Due to its heterogeneous, polymeric, cross-linked structure, lignin is highly resistant to enzymatic degradation, and this, along with its binding to other cell-wall components^{7, 8}, contributes not only to the highly recalcitrant nature of lignocellulosic biomass and the difficulty of efficient enzymatic hydrolysis, but also restricts the ability of lignin to serve as a straightforward feedstock for producing biofuels and bioproducts¹¹⁻¹³. Indeed, lignin degradation acts as the *de facto* limiting factor in the economic production of primary (sugars) and secondary (furfural and 5-hydroxymethylfurfural)

precursors from lignocellulose^{7, 8}. Methods to deal with the challenge posed by this natural polymer range from the production of transgenic crops with reduced lignin content^{5, 14} to the development of chemical (e.g. ionic liquids) and physical (e.g. high-pressure/high-temperature batch processing) pretreatment methods to alter lignin structure and interactions between lignin and other components of lignocellulose^{4, 15}.

Recently, a novel cosolvent based pretreatment method called CELF (Cosolvent Enhanced Lignocellulosic Fractionation) was reported^{9, 16, 17} and exploited^{18, 19}. CELF pretreatment employs tetrahydrofuran (THF) in a single phase mixture with water to augment the deconstruction of biomass. THF is a polar aprotic ether that can serve as a renewable alternative to dioxane, as THF can be produced from the catalytic decarbonylation and hydrogenation of furfural. Further, as THF is in contact with water during the CELF reaction, and never distilled to dryness, the potential hazard of peroxide formation (and associated explosion) is greatly reduced, which allows safe scaling of the CELF process for industrial application.

Functionally, THF has been applied in biomass research to help solubilize Kraft lignin extracted from biomass for analytical gel permeation chromatography (GPC)^{20, 21} as it has high solubility for methylated or acetylated lignins. More recently, CELF exploited THF's solvent and catalytic properties in combination with water at elevated temperature reactions with or without acids, to hydrolyze biomass sugars and promote the extraction and depolymerization of lignin^{16-19, 22}. Because THF is a Lewis base that coordinates with both Lewis acids and strong Brønsted acids, the presence of an acid, even in dilute concentrations, lowers the solution pH and greatly reduces the reaction severity needed to achieve comparable results. The addition of an acid to the CELF pretreatment, however, is not required for the breakdown of biomass (as noted above), but rather accelerates delignification at lower reaction temperatures to prevent THF degradation.

CELF was found to be highly effective at achieving high yields of fermentable sugars as well as their dehydration products furfural and hydroxymethylfurfural (5-HMF) directly from raw maple wood

and raw corn stover. At moderate 1:1 (v/v) THF:H₂O ratios, over 95% of the total sugars were recovered from corn stover using only 2 mg-enzyme/g-glucan enzyme dosages after CELF pretreatment and enzymatic hydrolysis. Fractal kinetics were applied to model the hydrolysis of sugars from CELF pretreated corn stover that correlated lignin removal to the enhanced digestibility of the solids¹⁹. This is because lignin extraction is particularly effective in CELF, removing over 90% of lignin from maple wood, producing a solubilized lignin product that can be precipitated as an ash-free and carbohydrate-free solid product, known as CELF lignin, upon removal and recovery of THF. At higher solvent ratios (3:1 v/v) in combination with metal halide acid catalysts, simultaneous co-production yields of furfural and 5-HMF were 95% and 51% of the theoretical from maple wood, respectively, and 93% of the lignin was solubilized. These yields, along with other characterization work under similar thermochemical pretreatment²³, suggest that the chemical structure of lignin is altered by THF in CELF pretreatment.

In most aqueous-based pretreatments, lignin is not removed entirely from biomass; instead, lignin and pseudo-lignin (material generated by the combination of lignin and hemi-cellulose degradation products^{24, 25}) aggregates onto the cellulose surface, blocking enzymatic access to cellulose and binding unproductively to the enzymes^{12, 13, 26-30}, an undesirable behavior for the production of biofuels. This coalescence of lignin in water can be understood in a general framework of the “quality” of a solvent relative to a polymer³¹⁻³⁴. Three classes of solvent can be considered. In a “bad” solvent, such as water, polymer-polymer interactions are favored, and the polymer collapses to “globular” conformations in which monomers are tightly packed. Furthermore, bad solvent conditions lead to the formation of multi-polymer aggregates that, for lignin, pose a major barrier to cellulose hydrolysis in pretreated biomass. In a hypothetical “theta” solvent, polymer–polymer and polymer–solvent interactions balance exactly, leading to the polymer adopting Gaussian “random-coil” conformations, similar to an ideal chain without excluded volume or intra-chain interactions. Finally, in a “good”

solvent, polymer-solvent interactions are energetically favorable, and the polymer adopts more extended conformations. Two additional important points regarding the solvent-polymer interactions are that 1) polymers do not form aggregates in dilute “theta” and “good” solvents and 2) the quality of the solvent for a given polymer concentration is independent of the polymer’s molecular weight.

As lignin hinders cellulose hydrolysis and has the potential to act as a raw material in its own right, recent efforts^{10, 23, 35-41} have focused on understanding factors determining the three-dimensional structure of lignin. These efforts have demonstrated that softwood lignin undergoes a “hard” to “soft” glass transition between 353K and 373K⁴⁰. Additionally, when aggregated, lignin exhibits a self-similar (fractal) structure over three orders of magnitude in length³⁹, and in aqueous solution at temperatures below the glass transition point, the polymer has a native state corresponding to a “crumpled globule” (defined as a collapsed globular state with a fractal dimension >3 , see SI Fig. 1)⁴⁰. Further, it was recently demonstrated that the stability of this (native) crumpled globule state is maintained by entropic contributions (through the hydrophobic effect) and not enthalpic contributions typical for the stability of globular states in hydrophobic polymers⁴².

In a cosolvent system, such as THF:H₂O in CELF, the question arises as to what structural changes occur in lignin that might facilitate the high degree of extraction from raw cellulosic feedstocks observed experimentally. To address this question, here we apply all-atom molecular dynamics simulations (MD) to examine the structure of lignin in THF:H₂O cosolvent environments similar to those found under CELF pretreatment. We find that, while pure H₂O is a bad solvent for lignin, the THF:H₂O cosolvent is “good”. The simulations used characterize the structural response of lignin through a variety of measures, including scaling law relationships (end-to-end and radius of gyration), solvent accessibility, and local lignin ring-ring orientation. Additionally, we examine the local environment of lignin within THF:H₂O cosolvent systems.

METHODS

To examine the effect of THF:H₂O cosolvent environments on the structure of lignin, all-atom explicit solvent MD simulations of a linear, sixty-one unit, softwood-like, lignin polymer chain were performed, with the lignin polymer made up G (guaiacyl) subunits having β -O4, β -5, 5-5, and α -O4 linkages. The exact composition of this lignin (noted as type L0a in the prior work) is reported elsewhere⁴⁰ and was previously used as a model for softwood-lignin^{39, 40}. Although lignin is a polydisperse material (as noted in other work⁴³), only a single polymer is used to probe the effects of the solvent on the structure of lignin since solvent “quality” effects, in principle, are independent of polymer molecular weight³¹⁻³⁴. The initial lignin conformation used in these simulations was a “crumpled-globule” taken from an equilibrium conformation obtained in a previous simulation of lignin in bulk water⁴⁰. The environmental conditions for our simulations were three different THF:H₂O volume ratios: 0 (corresponding to pure H₂O), 0.43 THF:H₂O, and 0.9 THF:H₂O, with each concentration examined at four different temperatures: 283K, 303K, 378K, and 445K. These temperatures were chosen to sample the range of temperatures under commonly reported CELF pretreatment cases^{9, 16, 17} (with T=283K chosen as an extreme case, in that the low temperature would not be used in an actual CELF pretreatment). The simulated concentrations were chosen to be near those reported by Cai et al⁸ (1:1 THF:H₂O v/v) and one in a mid-range between 1:1 and bulk water. A final note on our simulated system is that no acid is used, which is atypical of CELF pretreatment; however, previous experimental work (note above) on non-acid CELF pretreatment (*i.e.*, THF:water cosolvent system without acid catalyst) has demonstrated that without acid, the method is still effective¹⁹. As such, we focus our attention on the non-acid system to reduce computational costs and note that future studies will examine how the addition of acid effects the structure and dynamics of lignin under CELF conditions.

The simulation protocol was as follows: a short, ten-thousand step energy minimization phase was first performed, followed by five short (1ns) NPT position-restrained simulations at each

temperature with the reference pressure set to 1atm in order to relax the solvent and settle the final box size. From the final coordinates of the NPT simulations, five independent production realizations of length 200ns, in the NVT ensemble, were computed for use in structural characterization and analysis. In both the NPT and NVT simulations, the integration time step was 2fs, with frames saved every 2ps in the NPT and every 20ps in the NVT. The temperature and pressure were controlled with the V-rescale⁴⁰ and Berendsen thermo/barostats⁴⁵, respectively, with the Berendsen thermo/barostats used only in the NPT relaxation phase. To check that the fixed box size had limited influence on the dynamics, the minimum distance between any of the atoms of the lignin molecule to any other atoms of the periodic image of the lignin was computed for the lowest temperature simulations. The minimum distance was found to be (4.155nm), and when compared to the estimated Bejerrum length of the highest THF concentration cosolvent system (3.475nm)⁴⁶, was found to be ~0.7nm larger, which is large enough to prevent box-size effects.

The CHARMM-like lignin force-field⁴⁷ was used along with the recent CHARMM additive ether parameterization for THF^{48, 49} to provide the appropriate interaction potentials for our simulation. As the lignin force-field was derived using a CHARMM-based protocol, the two parameterizations used for the effective force-field are treated as compatible. All simulations used the GROMACS software versions 4.6.7 and 5.0.4⁵⁰⁻⁵³. The migration from version 4.6.7 to version 5 in this study was found necessary due to changing computational resources; however, as the simulations use the same force-fields and simulation conditions, no effect is expected on the results reported below.

The configurational distribution of lignin and its interactions with the cosolvent system were characterized by calculation of the radius of gyration (R_g), solvent accessible surface area (SASA), per-frame fractal dimension, radial distribution functions (RDFs), coordination numbers, spatial densities of solvent around lignin bonds, hydrogen-bonding (hydrogen bonds), self-contacts, lignin ring distance as a function of the number of monomers separating the rings, and the dot products of the normals of lignin rings as a function of ring-ring distance, all calculated using the last 100ns of each simulation.

All properties (excluding the fractal dimension, lignin ring-ring normals dot products, and ring-ring distances) were calculated with internal GROMACS 4.6.7 tools. The fractal dimension, ring-ring dot products, and ring-ring (and thus end-to-end) distance measures were obtained with in-house TCL scripts implemented within the Visual Molecular Dynamics (VMD)⁵⁴ package. For hydrogen bond/self-contacts, the cutoff distance was set at 3.0Å, and the angle cutoff set to 20°. Further, radial distribution functions of lignin and a given cosolvent were computed in two different ways, with one method taking the average of the radial distribution functions (RDFs) between lignin monomers and the solvent component of interest, and the other taking the average RDF between a specific lignin linkage and the solvent. These RDF calculations were further processed into two additional measures, local-solvent occupied volume ratios and coordination-numbers. For the calculation of local-solvent occupied volume ratios, each radial distribution was multiplied by the bulk density and the molecular volume of the solvent of interest (as computed with the 3V software⁵⁵); while, for the coordination numbers, the integral of the product of $4\pi r^2 g(r)$ (where $g(r)$ is the RDF) was computed from 0 to 0.5nm.

Spatial densities of the cosolvent relative to each of the four lignin monomer-monomer linkage types were computed using the `g_sdf` tool from GROMACS 4.6.7 with a bin-width of 0.1 nm. The above-noted densities were constructed by sampling every 5th frame of a given trajectory after centering the coordinates onto one of the four linkages.

The fractal dimension, which is related to the solubility of lignin (as dimensions below 2 indicate the polymer is readily soluble, while above is poorly-soluble), was obtained by fitting a power-law to the dependence of the R_g on the number of monomers separation for every saved frame to obtain the R_g scaling factor ν . This scaling factor was taken to be related to the fractal dimension by the relation: $\nu = (\alpha+2)(5\alpha)^{-1}$ where α is the mass fractal dimension⁵⁶. The relationship between the scaling factor and the mass fractal dimension used here is not the typical one from the Flory polymer theory but is instead a recent scaling law relation developed for hydrophobic chains⁵⁶. Mass-fractal

dimensions closer to $\alpha = 1$ (i.e. below 2) are “good” solvents while fractal dimensions near 3 or above are “bad/poor” solvents. Sample lignin structures at $\alpha = 1.5$ (good solvent), 3 (poor solvent), and 4 (very poor solvent) obtained from the simulations are presented in SI Fig. 1.

RESULTS

To characterize the structure of lignin, calculations of the average solvent accessible surface area (SASA), Radius of Gyration (R_g), and fractal dimension distributions are presented along with ring-ring angle/distance distributions, average distance between lignin rings as a function of monomer separation, and lignin-lignin hydrogen bonds. To characterize THF and H₂O interactions with lignin, lignin-H₂O hydrogen bonds, radial distribution functions of each cosolvent with respect to each lignin monomer, coordination numbers, spatial density functions, and the size scaled lignin-solvent local densities are presented.

Lignin Structural Characterization

Solvent Accessible Surface Area (SASA), Radius of Gyration (R_g), and Mass Fractal Dimension (α).

Global changes in lignin structure are characterized by the equilibrium SASA, R_g , and mass fractal dimension distributions (Figs. 1-3). A cursory view of these figures makes clear that in THF:H₂O solutions, the distributions of these measures are shifted to values corresponding to more open structures, as indicated by fractal dimension values below 3 (below the bad solvent threshold), higher R_g values, and more exposed surface areas (high values of SASA). Focusing on the SASA (Fig. 1), it is clear that there is (virtually) no overlap of the distributions obtained in THF:H₂O and bulk water, with this distinction being greatest for $T > 283\text{K}$. This separation is also found in the R_g and mass fractal distributions (Figs. 2 & 3) for $T > 283\text{K}$ and $T < 445\text{K}$. At the two extreme temperature cases, however, overlaps do form, with the largest overlaps found at $T = 283\text{K}$ with a THF:H₂O v/v ratio of 0.4 and at $T = 445$ with a THF:H₂O v/v ratio of 0.9.

In the case of the fractal dimension, α (Fig. 3), at both THF concentrations and $T > 283\text{K}$, the distributions are shifted to values well below 3, with an average ~ 1.5 ; while in bulk water there is a weak trend with temperature towards a fractal dimension of 3, consistent with a “globular” conformation. Close examination of the error-bars in the apparently bimodal distributions at $T=283\text{K}$ (for all solvent conditions) and $T=303$ in bulk water suggests that the distributions for these temperatures did not completely converge and are likely not actually bimodal. However, despite the lack of convergence, the figures do indicate that at $T < 378\text{K}$ in the case of bulk water, the distribution of α is broad and (due to the smaller error bars at higher values) has its mean at $\alpha = 3$ or higher. Similarly, for the cosolvent systems at $T=283\text{K}$, it can be inferred that the distribution is also broad (in that it has a range from $\alpha = 1$ to $\alpha = 4.5$).

Lignin Ring-Ring/End-to-End Distances and Lignin Ring-Ring Dot Products. Two convenient properties for characterizing the lignin polymer are the distances and dot products between the rings as a function of the number separation between the monomers. Fig. 4 shows 2D histograms of the ring-ring dot products versus ring-ring distances. Comparing the smallest distance bin (the first bin) of each histogram with the remaining distances demonstrates that, regardless of THF concentration and temperature, the dot products are correlated (as shown by the two peaks at dot-product values of 1 and -1) only for values found in the first distance bin (median distance 0.31nm) (see also SI figure 2 for a separate plot of the first two distance bins). A direct calculation of the persistence length by fitting to the end-to-end distribution⁵⁷ (SI figure 3) further corroborates the above ring-ring correlation distance, with values being below the first distance bin median (persistence length values between 0.18nm to 0.22nm) and having negligible temperature dependence. Hence, both the persistence length and the ring-ring correlation distance are on the order of a single monomer-monomer linkage and both of these measures have no significant temperature or concentration dependence. Further, we conclude that lignin is very flexible in all environments tested. Figure 4 also shows that, with increasing temperature,

the maximum sampled ring-ring distances (shown by the change from blue to yellow/red towards the top of the higher temperature histograms) increase, and that the distribution of ring-ring distances broadens; however, this effect is most predominant in THF:H₂O.

A corollary to the ring-ring distance distribution changes observed in Fig. 4 is the direct calculation of mean end-to-end distances as a function of the number of monomer separations, shown in Fig. 5. For both THF concentrations, the average end-to-end distances are substantially greater than those found in bulk water. Additionally, in all THF:H₂O environments (excluding T=283K), the end-to-end scaling profiles are similar to that for a power-law relationship (noted by the straight-line behavior in log-log space, see SI Fig 4). A power-law expression with an exponent of 0.5 is consistent with a polymer in a theta solvent, and, as shown in SI Table 2 at T>283K, for THF:H₂O, the power-law exponent is at or near this theta value. This behavior is in stark contrast to that found in bulk water systems at all temperatures, where the end-to-end distance reaches a plateau beginning at a distance of 2nm, indicating a globular polymer in a bad solvent^{31, 39}. It is also interesting to note that even at T=283K, where at all solvent conditions, lignin approaches a plateau, the plateau in bulk water (regardless of temperature) is found at an end-to-end distance below that in THF:H₂O (~2.5-3nm).

Lignin-Lignin Hydrogen-Bonding and Contacts. Final measures of lignin structure examined are the number of self-contacts and hydrogen-bonds (Fig. 6). As with the previous structural quantities, a clear difference exists between lignin in bulk water and THF:H₂O environments (at all temperatures), with the latter having both fewer hydrogen bonds and contacts. Further, unlike the previous structural properties (radius of gyration, SASA, and fractal dimension), the trends in the number of contacts/hydrogen bonds with temperature are approximately the same in both H₂O and THF:H₂O solvents.

Lignin-Solvent Interactions

Lignin-H₂O Hydrogen Bonding. Lignin, although largely hydrophobic, does have the capacity to form hydrogen bonds with water. However, Fig. 7 shows that, as with the lignin-lignin hydrogen bonds, the

presence of THF reduces the propensity for lignin-H₂O hydrogen bond formation. A further similarity between the lignin-water and lignin-lignin hydrogen bonds is that the decreasing trends with temperature between all of the solvent environments follow one another (e.g., the rate of decrease in number of lignin-H₂O H-bonds with increasing temperature is roughly the same for all solvent conditions).

Lignin-Solvent Radial Distribution Functions. Figure 8 characterizes the distribution of the solvent components about the lignin surface. These distributions demonstrate that THF is more likely to be found near lignin relative to in the bulk environment, while water has a conversely lower density near the polymer. THF thus preferentially solvates lignin. Figure 9 shows that, at all but one temperature and concentration pair, THF occupies more local volume near lignin than water, as seen by the magnitude of the green and red curves being less than for blue and orange. The exception to this trend of greater THF occupancy is for the 0.9 THF:H₂O v/v ratio environment at T=445K, where THF and water occupy approximately the same amount of local volume for distances from lignin between 0.35-0.75nm. As a final observation from these profiles, it is interesting to note that in all but the 0.9 THF:H₂O v/v ratio environment, the trends of THF concentrating near the lignin surface increase with temperature, while the water concentration decreases correspondingly.

Figure 10 gives an additional view of the local solvent makeup near the lignin surface by means of coordination numbers around lignin linkages. These are derived by integration of the distribution functions of THF and H₂O with respect to lignin atoms making a linkage and averaged per linkage type (β -O4, β -5, 5-5, and α -O4). All linkage types have a greater number of water molecules than THF coordinated around them, which is not unexpected due to its smaller size compared to the THF molecule. Both solvent types have a preference to accumulate around the β -5 linkage (red squares in Fig. 10), which consists of two bonds between neighboring monomers giving rise to a ring structure and thus a relatively large area for solvent coordination. Moreover, the coordination of the β -O4 and α -

O4 linkages by H₂O (blue triangles and green diamonds) is higher than that for the 5-5 linkage, and at some temperatures, are slightly higher than that of β -5. Interestingly, when comparing the cosolvent system to the water solvent system, the water coordination numbers in the latter are not only higher, which is expected in that there are no competing THF molecules, but they are also of comparable magnitude among the different linkage types at any temperature. Thus, the addition of THF and its concomitant binding to coordination sites of lignin concentrates water coordination around specific lignin monomer-monomer bonds.

Lignin-Solvent Spatial Densities. Figure 11 provides a visualization of the average spatial densities of the cosolvent system around the four different types of lignin linkage. The results are consistent with the calculated coordination numbers. The figure illustrates the hydration of lignin hydroxyl groups that are somewhat separated from the 5-5 bond, the peak in THF density around the β -5 linkage, and the coordination of sites proximal to the β -O4 and α -O4 linkages by both THF and H₂O. In general, it is found that THF occupies the space around the aryl rings, while water density is increased near the hydroxyl groups. This arrangement, in particular, allows water molecules to access the β -O4 and α -O4 linkages.

DISCUSSION

The present study used large-scale all-atom molecular dynamics simulations to examine the structure of lignin at four different temperatures and THF concentrations. A variety of structural parameters were characterized, including lignin ring-ring and end-to-end distances and angle distributions, classical polymer characteristics such as the mass fractal dimension distributions, and lignin-lignin hydrogen bonds and contacts. Further, the interactions between the solvent and lignin were also examined *via* lignin-H₂O hydrogen bond and lignin-solvent radial distribution function calculations.

From all of the above calculations, it is clear that the conformation of lignin is profoundly altered by the addition of THF as a cosolvent to water compared to a pure water system. To examine

the observed changes in more detail, this discussion of the results is divided into two parts, the first detailing what can be deduced from the properties regarding the changes in the structure of lignin and the second focused on positing a physical mechanism behind these changes.

Lignin is a flexible random coil in THF:H₂O Solutions for $T \geq 303\text{K}$ and swells at $T = 283\text{K}$. Before examining the gamut of structural changes induced by the THF environment, it is beneficial to compare our structural results (Figs. 1-6) with previous work⁴⁰, in which lignin was found to be a crumpled globule under bulk water conditions with $T < 445\text{K}$. A crumpled globule is a densely packed, collapsed structure with a fractal dimension ≥ 3 and with the end-to-end distance as a function of monomer separation obeying a scaling law with a plateau as the chain size increases. In the present simulations in bulk water, we also find that lignin at $T < 445\text{K}$ exists as a crumpled globule (though we note that even at 445K , the crumpled globule state is not entirely absent), as evidenced by the plateau in the end-to-end distances as a function of monomer separation (in Fig. 5) and a mass fractal dimension distribution ranging from 3-5 (Fig. 3).

In comparison, when lignin is simulated in 0.9 THF:H₂O v/v or 0.43 THF:H₂O v/v ($T \geq 303\text{K}$), the end-to-end distances as a function of the number of monomer separation do not plateau (Fig 5), nor does the mass fractal dimension exceed 2.7 (Fig. 3). Indeed the peak of the fractal dimension distribution in the THF cosolvent systems ($T \geq 303\text{K}$) is near 1.77, well below the 2.7 dimension associated with a typical globule state (collapsed polymer), thereby strongly indicating that the lignin chain is a random-coil, *i.e.*, it follows Gaussian statistics⁵². The presence of Gaussian statistics is supported further by the Gaussian-like end-to-end distributions (and their associated fits to Gaussian distributions) presented in SI Fig. 3. Combining the observation that the chain is a random-coil with the short persistence length indicates that in THF:H₂O cosolvent conditions at $T \geq 303\text{K}$, softwood-like lignin exists as a flexible polymer near “theta” solvent conditions. Importantly, when found in this state, lignin should not self-aggregate and would be relatively easily removed during THF:H₂O

cosolvent pretreatment, consistent with experimental results¹⁷.

Regarding the change to lignin in THF:H₂O, it is interesting to consider the extreme low-temperature case ($T=283\text{K}$), in which a variety of structural changes, relative to lignin in bulk water, take place, though not as dramatically. Beginning with the mean end-to-end distance distribution (Fig. 5) and ring-ring distances (Fig. 4), it is clear that the average monomer-monomer distances of the polymer are lengthened (this is also indicated by the decrease in lignin-lignin contacts compared to bulk conditions, see Fig. 6). However, as the distribution of fractal dimension (although widely varying) is centered between 2.5 and 3.5, along with an increase in overlap between the THF:H₂O and bulk R_g distributions, it is unlikely that lignin at this temperature completely adopts random-coil configurations. Combining the above observations makes clear that, although lignin at low temperature ($T=283\text{K}$) is not in a random-coil state, it does swell compared to pure water. This swelling indicates that although temperature plays a role in CELF, the presence alone of THF shifts the equilibrium population of conformations from crumpled-globules to swollen and random-coil states.

THF is the local solvent for lignin and limits lignin-lignin hydrogen bonds/contacts and lignin-H₂O hydrogen bond formation. A benefit of using MD simulations for examining the change in lignin's structure in THF:H₂O environments is that the atomic-level details of the interactions between the cosolvents and lignin are readily resolved. A straightforward calculation clearly shows a decrease in the number of H₂O-lignin hydrogen bonds in the THF:H₂O environments (Fig. 7) compared to pure water. However, this finding may not appear self-evident when considering that lignin in THF:H₂O environments has higher SASA values and is extended in conformation, consistent with more exposed H₂O-lignin hydrogen bond sites. THF is more densely distributed close to lignin than water, and at distances less than 1nm (Figs. 8 & 9), it is clear that THF is the primary solvent that sterically limits the access of water molecules to form hydrogen bonds with lignin. However, the data does indicate that THF is not able to block all of the hydrogen bond sites available, and water molecules therefore still have access to certain locations along the polymer. Indeed, the calculations of coordination numbers

and spatial densities (Figs. 10 & 11) reveal that in the THF systems, water does preferentially occupy sites near hydroxyl groups and to a lesser extent around the α -O4 and β -O4 linkages, and the latter may be important for the efficient hydrolysis of lignin. Among the seven most common linkages (β -O4, β - β , 4-O-5, β -1, 5-5, α -O4, and β -5), the α -O4 linkage and β -O4 linkage tend to cleave most easily during dilute acid pretreatment^{58 59}. For lignin hydrolysis, both protons and water need to have access to, and ideally be pre-positioned for, the cleavage of aryl-ether linkages. The removal of excess water and the preferential arrangement of THF in the immediate vicinity of lignin thus may promote the lignin hydrolytic reaction. This access may be particularly important in CELF⁸ as the existence of protons provided by these water molecules may lead to the hydrolysis necessary to explain the breakdown of lignin into lower molecular weight samples during pretreatment.

Along with the reduction of the number of water-lignin hydrogen bonds, the presence of THF as the primary local solvent implies that in THF:H₂O cosolvent systems, the local environment for lignin is hydrophobic, which limits lignin-lignin interactions. In bulk water environments, the collapsed state of lignin is supported by entropic contributions (i.e., the hydrophobic effect), which encourage lignin-lignin contacts and intrapolymeric hydrogen bonds. Lignin, being predominately solvated by the hydrophobic (THF) medium in the cosolvent systems, reduces these hydrophobic contributions, and as a result reduces the favorability of lignin-lignin interactions (as shown by Fig 6). Consequently, lignin's equilibrium configuration distribution shifts from a crumpled coil to an extended chain. Evidence supporting this shift is found by turning to the unusual decrease in lignin's THF solvation along with its associated increase in hydration, found in the 0.9 THF:H₂O v/v environment at 445K. Comparing the decrease in THF solvation (Figs. 8 & 9) to the modified behavior (shifts) in the distributions of structural metrics (Figs. 1-3) and the increase in lignin-lignin hydrogen bonds and lignin-lignin contacts (Fig 6) to those same characteristics in bulk water (at T=445K), indicates that as THF solvation decreases and water hydration increases, more globular like structures begin to be sampled.

CONCLUSIONS

This study examined the structure of lignin under THF:H₂O cosolvent system, as applied in CELF pretreatment, using all-atom molecular dynamics simulations and provided evidence that under these conditions, lignin adopts extended coil configurations while being preferentially solvated by the THF cosolvent. These findings may be of particular interest to those exploring application of THF:H₂O pretreatment to lignocellulosic biomass. Lignin in a coil conformation will not self-aggregate, and its preferential solvation by THF may allow separation of lignin from cellulose, making lignin more easily removed during pretreatment. This mechanism may account for the reduction in the recalcitrance of lignocellulosic biomass to enzymatic breakdown in the THF:H₂O cosolvent system as the association with cellulose would be disrupted, allowing access of cellulolytic enzymes to the cellulose fibers.

Acknowledgements:

This research was funded by the BioEnergy Science Center, a U.S. Department of Energy (DOE) Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. This research also used the computing resources provided by the U.S. Department of Energy's National Energy Research Scientific Computing Center.

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>).

1. A. D. Sagar and S. Kartha, *Annu Rev Env Resour*, 2007, **32**, 131-167.
2. T. R. Brown and R. C. Brown, *Biofuel Bioprod Bior*, 2013, **7**, 235-245.
3. T. R. Brown, *Bioresource Technol*, 2015, **178**, 166-176.
4. B. Yang and C. E. Wyman, *Biofuel Bioprod Bior*, 2008, **2**, 26-40.
5. 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.
6. Y. Sun and J. Y. Cheng, *Bioresource Technol*, 2002, **83**, 1-11.
7. V. J. H. Sewalt, W. G. Glasser and K. A. Beauchemin, *J Agr Food Chem*, 1997, **45**, 1823-1828.
8. S. Y. Ding, Y. S. Liu, Y. N. Zeng, M. E. Himmel, J. O. Baker and E. A. Bayer, *Science*, 2012, **338**, 1055-1060.
9. C. M. Cai, T. Y. Zhang, R. Kumar and C. E. Wyman, *J Chem Technol Biot*, 2014, **89**, 2-10.
10. A. J. Ragauskas, G. T. Beckham, M. J. Biddy, R. Chandra, F. Chen, M. F. Davis, B. H. Davison, R. A. Dixon, P. Gilna, M. Keller, P. Langan, A. K. Naskar, J. N. Saddler, T. J. Tschaplinski, G. A. Tuskan and C. E. Wyman, *Science*, 2014, **344**, 709.
11. H. Jorgensen, J. B. Kristensen and C. Felby, *Biofuel Bioprod Bior*, 2007, **1**, 119-134.
12. L. Kumar, V. Arantes, R. Chandra and J. Saddler, *Bioresource Technol*, 2012, **103**, 201-208.
13. H. J. Li, Y. Q. Pu, R. Kumar, A. J. Ragauskas and C. E. Wyman, *Biotechnol Bioeng*, 2014, **111**, 485-492.
14. A. Furtado, J. S. Lupoi, N. V. Hoang, A. Healey, S. Singh, B. A. Simmons and R. J. Henry, *Plant Biotechnol J*, 2014, **12**, 1246-1258.
15. Y. Q. Pu, N. Jiang and A. J. Ragauskas, *J Wood Chem Technol*, 2007, **27**, 23-33.
16. C. M. Cai, N. Nagane, R. Kumar and C. E. Wyman, *Green Chem*, 2014, **16**, 3819-3829.
17. C. M. Cai, T. Y. Zhang, R. Kumar and C. E. Wyman, *Green Chem*, 2013, **15**, 3140-3145.
18. Z. C. Jiang, T. He, J. M. Li and C. W. Hu, *Green Chem*, 2014, **16**, 4257-4265.
19. T. Y. Nguyen, C. M. Cai, R. Kumar and C. E. Wyman, *Chemsuschem*, 2015, **8**, 1716-1725.
20. A. Abe, K. Dušek, S. Kobayashi and S. Błażewicz, *Biopolymers : lignin, proteins, bioactive nanocomposites*, Springer, Berlin, 2010.
21. E. Sjöström and R. Alén, *Analytical methods in wood chemistry, pulping, and papermaking*, Springer, Berlin ; New York, 1999.
22. J. X. Long, Q. Zhang, T. J. Wang, X. H. Zhang, Y. Xu and L. L. Ma, *Bioresource Technol*, 2014, **154**, 10-17.
23. Y. Q. Pu, F. Hu, F. Huang, B. H. Davison and A. J. Ragauskas, *Biotechnol Biofuels*, 2013, **6**.
24. P. Sannigrahi, D. H. Kim, S. Jung and A. Ragauskas, *Energ Environ Sci*, 2011, **4**, 1306-1310.
25. J. Jakobsons, B. Hortling, P. Erins and J. Sundquist, *Holzforschung*, 1995, **49**, 51-59.

26. B. S. Donohoe, S. R. Decker, M. P. Tucker, M. E. Himmel and T. B. Vinzant, *Biotechnol Bioeng*, 2008, **101**, 913-925.
27. T. Eriksson, J. Borjesson and F. Tjerneld, *Enzyme Microb Tech*, 2002, **31**, 353-364.
28. A. Berlin, M. Balakshin, N. Gilkes, J. Kadla, V. Maximenko, S. Kubo and J. Saddler, *J Biotechnol*, 2006, **125**, 198-209.
29. D. W. Sammond, J. M. Yarbrough, E. Mansfield, Y. J. Bomble, S. E. Hobdey, S. R. Decker, L. E. Taylor, M. G. Resch, J. J. Bozell, M. E. Himmel, T. B. Vinzant and M. F. Crowley, *J Biol Chem*, 2014, **289**, 20960-20969.
30. J. L. Rahikainen, J. D. Evans, S. Mikander, A. Kalliola, T. Puranen, T. Tamminen, K. Marjamaa and K. Kruus, *Enzyme Microb Tech*, 2013, **53**, 315-321.
31. A. Y. Grosberg and D. V. Kuznetsov, *Macromolecules*, 1992, **25**, 1996-2003.
32. P. J. Flory, *J Chem Phys*, 1949, **17**, 303-310.
33. P.-G. d. Gennes, *Scaling concepts in polymer physics*, Cornell University Press, Ithaca, N.Y., 1979.
34. B. Y. Ha and D. Thirumalai, *Phys Rev A*, 1992, **46**, R3012-R3015.
35. U. Vainio, N. Maximova, B. Hortling, J. Laine, P. Stenius, L. K. Simola, J. Gravitis and R. Serimaa, *Langmuir*, 2004, **20**, 9736-9744.
36. S. E. Harton, S. V. Pingali, G. A. Nunnery, D. A. Baker, S. H. Walker, D. C. Muddiman, T. Koga, T. G. Rials, V. S. Urban and P. Langan, *Acs Macro Lett*, 2012, **1**, 568-573.
37. Y. H. Deng, Y. Q. Guo, Y. Qian, H. M. Lou and X. Q. Qiu, *Bioresources*, 2014, **9**, 6304-6315.
38. R. L. Silveira, S. R. Stoyanov, S. Gusarov, M. S. Skaf and A. Kovalenko, *J Phys Chem Lett*, 2015, **6**, 206-211.
39. L. Petridis, S. V. Pingali, V. Urban, W. T. Heller, H. M. O'Neill, M. Foston, A. Ragauskas and J. C. Smith, *Phys Rev E*, 2011, **84**.
40. L. Petridis, R. Schulz and J. C. Smith, *J Am Chem Soc*, 2011, **133**, 20277-20287.
41. G. Cheng, S. Datta, Z. L. Liu, C. Wang, J. K. Murton, P. A. Brown, M. S. Jablin, M. Dubey, J. Majewski, C. E. Halbert, J. F. Browning, A. R. Esker, B. J. Watson, H. T. Zhang, S. W. Hutcheson, D. L. Huber, K. L. Sale, B. A. Simmons and M. S. Kent, *Langmuir*, 2012, **28**, 8348-8358.
42. M. V. Athawale, G. Goel, T. Ghosh, T. M. Truskett and S. Garde, *P Natl Acad Sci USA*, 2007, **104**, 733-738.
43. A. Tolbert, H. Akinosho, R. Khunsupat, A. K. Naskar and A. J. Ragauskas, *Biofuel Bioprod Bior*, 2014, **8**, 836-856.
44. G. Bussi, D. Donadio and M. Parrinello, *J Chem Phys*, 2007, **126**.
45. H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola and J. R. Haak, *J Chem Phys*, 1984, **81**, 3684-3690.
46. A. C. Kumbharkhane, S. N. Helambe, M. P. Lokhande, S. Doraiswamy and S. C. Mehrotra, *Pramana-J Phys*, 1996, **46**, 91-98.
47. L. Petridis and J. C. Smith, *J Comput Chem*, 2009, **30**, 457-467.
48. I. Vorobyov, V. M. Anisimov, S. Greene, R. M. Venable, A. Moser, R. W. Pastor and A. D. MacKerell, *J Chem Theory Comput*, 2007, **3**, 1120-1133.
49. H. Lee, R. M. Venable, A. D. MacKerell and R. W. Pastor, *Biophys J*, 2008, **95**, 1590-1599.
50. H. Bekker, H. J. C. Berendsen, E. J. Dijkstra, S. Achterop, R. Vondrumen, D. Vanderspoel, A. Sijbers, H. Keegstra, B. Reitsma and M. K. R. Renardus, *Physics Computing '92*, 1993, 252-256.
51. B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J Chem Theory Comput*, 2008, **4**, 435-447.
52. S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess and E. Lindahl, *Bioinformatics*, 2013, **29**, 845-854.

- 53. D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J Comput Chem*, 2005, **26**, 1701-1718.
- 54. W. Humphrey, A. Dalke and K. Schulten, *J Mol Graph Model*, 1996, **14**, 33-38.
- 55. N. R. Voss and M. Gerstein, *Nucleic Acids Res*, 2010, **38**, W555-W562.
- 56. L. Hong and J. Z. Lei, *J Polym Sci Pol Phys*, 2009, **47**, 207-214.
- 57. M. Rubinstein, R. H. Colby and Knovel (Firm), Oxford University Press,, Oxford ; New York2003, p. 1 online resource (460 p.).
- 58. S. Kim, S. C. Chmely, M. R. Nimos, Y. J. Bomble, T. D. Foust, R. S. Paton and G. T. Beckham, *J Phys Chem Lett*, 2011, **2**, 2846-2852.
- 59. W. Qin, L. N. Wu, Z. M. Zheng, C. Q. Dong and Y. P. Yang, *Molecules*, 2014, **19**, 21335-21349.



Figures)

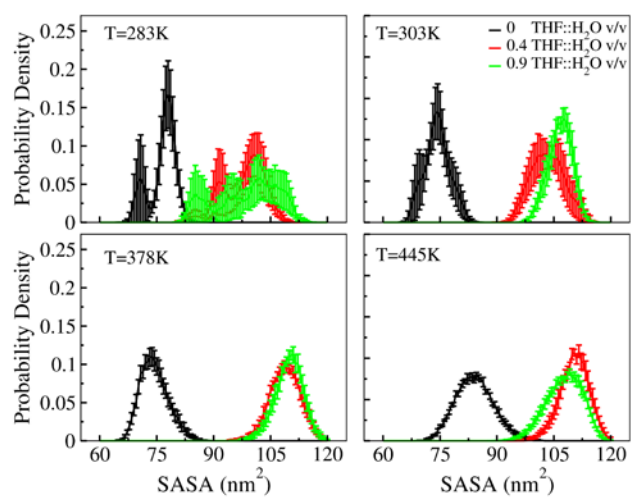


Figure 1) Average solvent-accessible surface area (SASA) distributions. Error-bars are standard error of the mean of each histogram bin.

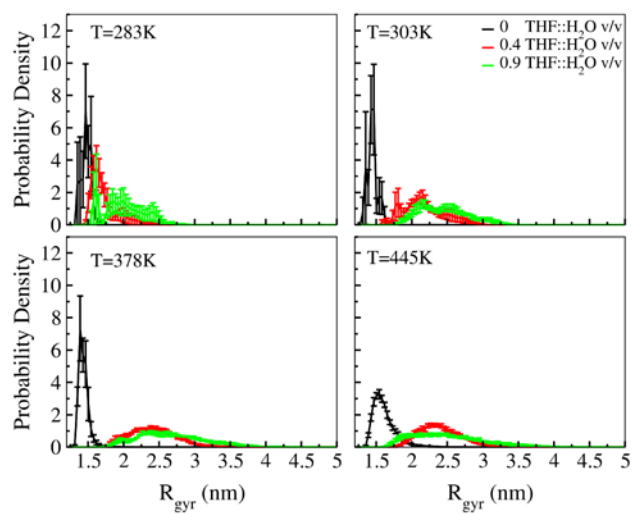


Figure 2) Average radius of gyration (R_{g}) distributions of lignin. Error-bars are standard error of the mean of each histogram bin.

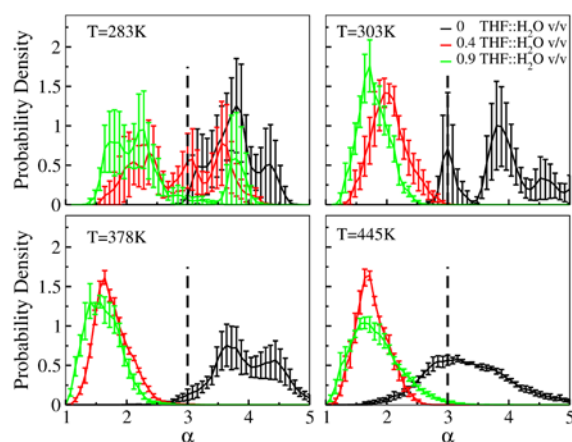


Figure 3) Average mass fractal dimension (α) distributions from fractal dimension calculations on each frame of the last 100ns of each trajectory. Low values of α indicate lignin's conformations being more coil-like. The dashed line indicates boundary between coil states (left) and globular states (right). Error-bars are standard error of the mean of each histogram bin.

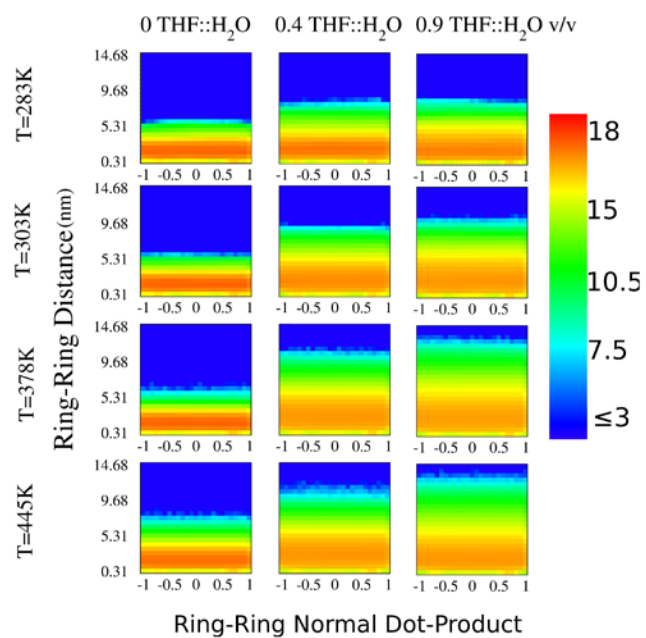


Figure 4) 2D Histogram of the scalar (dot) product of the lignin ring-normals as a function of ring-ring distance. Values of 1(-1) indicate alignment(anti-alignment) of monomer rings. Color scale is logarithm of the fraction of frames in each bin. Y-Axis labels correspond to median distance values (nm) for each bin.

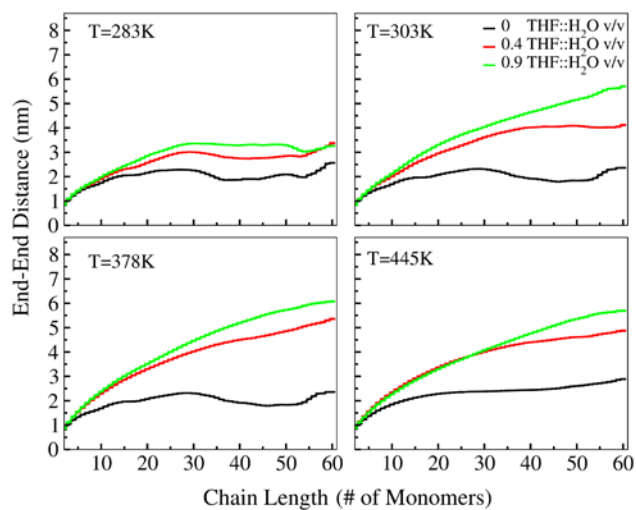


Figure 5) Lignin End-End (Ring) Distances versus chain length (in number of monomers). Error bars are standard errors of the mean and are typically of the same size as the width of each line. The same plot projected onto logarithmic axis is provided as SI Fig. 4.

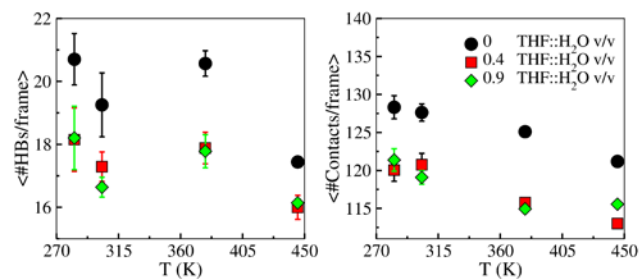


Figure 6) (Left) Average number of lignin-lignin hydrogen-bonds per frame (one frame corresponds to 20ps). (Right) Average number of lignin-lignin contacts per frame. Error bars are standard error of the mean.

Figure 7) Average number of lignin-water HBs per frame (one frame corresponds to 20ps). Error bars are equal to or smaller than the size of each shape and are of the standard error of the mean.

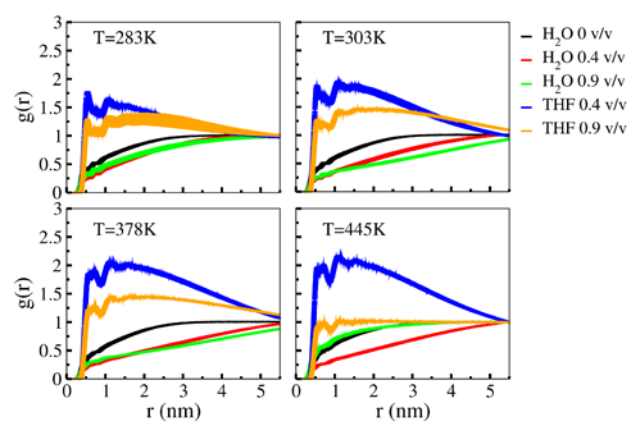


Figure 8) Mean Lignin-Solvent radial distribution functions. Colors indicate the cosolvent of interest. Error-bars are standard error of the mean and are of the order of the width of the lines.

Figure 9) Solvent occupied local volume ratios. The radius of the local volume near the lignin is taken to be a sphere with a radius centered at 0.35nm and ranging up to ~1.1nm. Error-bars are standard error of the mean and are the width of the line.

Figure 10) Solvent-lignin (bond) coordination numbers. Top row sub-figures correspond to H₂O coordination while the bottom row correspond to THF coordination. Error-bars are standard-error of the mean and are at most the size of the data-points.

Figure 11) Spatial densities of solvent centered about lignin-linkages (type noted in upper left corner). The orange contours describe the location where THF ring atoms occur at more than three times the bulk concentration and the blue contours correspond to locations of water oxygen with at least 1.5 times the bulk concentration.