SUPPORTING INFORMATION

Deconstruction of biomass enabled by local de-mixing of cosolvents at cellulose and lignin surfaces

Sai Venkatesh Pingali, Micholas Dean Smith#, Shih-Hsien Liu#, Takat B. Rawal#, Yunqiao Pu, Riddhi Shah, Barbara Evans, Volker Urban, Brian H. Davison, Charles Cai, Arthur J. Ragauskas, Hugh M. O'Neill, Jeremy C. Smith, Loukas Petridis (petridisl@ornl.gov)

Supplementary Methods

Poplar lignin isolation and acetylation

Poplar wood used in this study was provided by Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States. The poplar wood was debarked, chipped and Wiley milled through a 40-mesh screen (Thomas Scientific, Swedesboro, NJ). Lignin was isolated from the poplar biomass according to a literature method.(53-55) In brief, the Wiley milled poplar wood powders were extracted with ethanol: toluene (1:2, v/v) for 24 h in a Soxhlet apparatus to remove extractives, followed by washing with water and air drying. The extractives-free poplar was ball-milled using a planetary ball mill (Retsch PM 100) at 580 rpm for 2.5 h. The milled fine biomass powder was then subjected to enzyme treatment using CTec2 and HTec2 enzyme mixture with citrate buffer (1.0 M) solution under continuous agitation (200 rpm) at 50°C for 48 h. The residue was isolated by centrifugation and was hydrolyzed one more time with freshly added enzymes. The residue obtained was recovered by centrifuging and washing with deionized water. The recovered solids were treated with protease (Protease from Streptomyces, Sigma-Aldrich) to remove residual enzymes at 37 °C for 24 h. The protease was deactivated at 100 °C for 10 min prior to freeze-drying. The recovered residue was then extracted with p-dioxane:water (96:4, v/v) under stirring at room

www.pnas.org/cgi/doi/10.1073/pnas.1922883117

temperature for 48 h X 2 in the dark. The extracted mixture was combined, centrifuged and the supernatant was collected, roto-evaporated using a rotary evaporator, and freeze dried for further use. The dried lignin samples were dissolved in a 1:1 acetic anhydride/pyridine mixture and stirred for overnight at room temperature. Ethanol was added to the reaction mixture and the solvents were removed with a rotary evaporator. The addition and removal of ethanol was repeated until acetic acid was completely removed from the sample. The acetylated lignin sample was then dissolved in chloroform and precipitated with diethyl ether. The precipitate was centrifuged, washed with diethyl ether and dried under vacuum. Gel permeation chromatography analysis of the lignin molecular weight is shown in

Table S1, and NMR analysis found a S/G ratio of ~1.6.

Table S1*. Gel permeation chromatography analysis of the lignin molecular weight is shown below and NMR analysis found a S/G ratio of ~1.6.*

	Mn	Mw
Poplar acetylated lignin	1.72×10^3	9.88×10^3

Bacterial cellulose production

Acetobacter xylinus subsp. *sucrofermentans* (ATCC 700178) was purchased from the American Type Culture Collection (Manassas, Virginia, USA). Deuterium oxide (D_2O) was purchased from Cambridge Isotope Laboratories (Andover, Massachusetts, USA). Partially deuterated bacterial cellulose was prepared by growing the bacteria in a D_2O-based growth media with hydrogenated glycerol as carbon source for 5 days at 26˚C, as previously described (56, 57). The pellicle formed at the air-liquid interface was removed and washed rigorously and repeatedly with 1% sodium hydroxide to remove the growth media and bacterial debris. The solvent was then neutralized (pH ~7) by successive washes in H_2O to remove NaOH.

To prepare the cellulose for SANS measurements, the pellicles were frozen at -80˚C for 2 h, followed grinding (6 x 30 sec pulses, 10 secs between pulses) to form a slurry using a blender (Waring Inc, Texas, USA). The mass fraction of the cellulose in the slurry, expressed as weight percent (wt %), was determined by freeze-drying a known volume of the slurry (in triplicate). Finally, the cellulose slurry was solvent-exchanged into a 85/15 D_2O/H_2O mixture (58).

Small-angle neutron scattering

The Bio-SANS instrument was setup to maximize the dynamic Q-range obtained in a single exposure, Q ~ 0.003 – 0.85 $Å^{-1}$. In the case of the in-situ reaction studies, the overall setup that includes the reaction cell limits reliable data to Q_{max} ~ 0.5 $Å^{-1}$. In the in-situ reaction experiments, the sample was heated to 150 °C and the pressure did not exceed 100 bar. All other SANS experiments were conducted in room temperature and pressure. To obtain this large dynamic q-range in a single exposure, the main detector was positioned at 15.5 m from sample; the wing detector at 1° rotation from direct beam; neutron wavelength was set to 6 Å neutrons to obtain maximum flux of neutrons; source and sample apertures were 40 mm and 12 mm, respectively and placed 17.3 m apart (equivalently 'zero' guides in the collimator sections placed in the beam path). The raw SANS data was converted to reduced SANS data by applying instrument specific corrections implemented in the Mantid software(59). The reduced SANS data was obtained following correcting for dark current, detector pixel solid angle, pixel sensitivity, sample thickness, absolute intensity and background or solvent scattering. The 2D scattering patterns were azimuthally averaged to produce 1D SANS scattering profiles presented as intensity versus wave-vector, I(Q) vs Q.

Molecular Dynamics Simulations

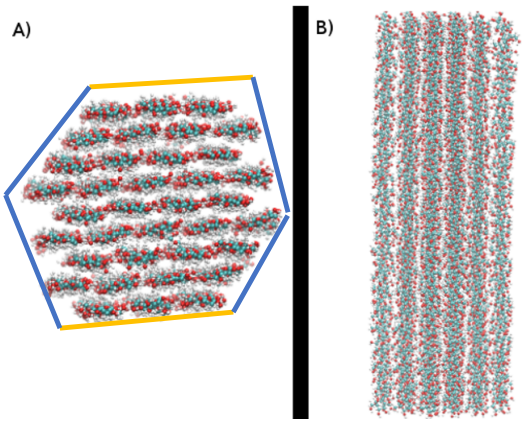

Figure S1*. Simulation model of cellulose: (A) perpendicular to the fibril axis with non-polar surfaces shown in orange and polar surfaces in blue; (B) view parallel to the fibril axis.*

Cellulose MD *Simulations.* To study the solvation of cellulose under THF-water conditions, classical all-atom molecular dynamics simulations of a model cellulose Iβ(60) fibril were performed. The model was constructed to match prior work and consisted of 36 cellulose chains of degree of polymerization 20 (Figure S1)(61). The cellulose model was then solvated using the GROMACS 2016.3 gmx solvate tool in two solvent conditions: 1:3 THF-water volume/volume ratio and pure water(62). The simulations were performed at the same conditions (THF:water ratio and 25 °C) as the SANS experiments in Figures 2 and S2. All simulations were performed using the GROMACS 2016.3 simulation package and force-field parameters were taken from taken from the CHARMM force-field(62-68).

For both pure water and 1:3 THF:water, simulations were performed using a standard three-stage approach: energy minimization, box/solvent relaxation, and production simulation. Energy minimization was performed using steepest-descent with a target of 50 kJ mol^{-1} nm^{-1} or until 17000 steps are reached, whichever is reached first. Following energy minimization, five independent instances were generated using the energy-minimized configuration as initial conditions. These independent instances were then

simulated using an integration timestep of 2fs with the solute (Cellulose) subjected to position restraints while the solvent and box were allowed to equilibrate for 2.5 ns within canonical ensemble using Berendsen thermostat(69). During equilibration half of the instances had temperatures set to 25°C and the other half at 45°C. Following equilibration, five production simulations were conducted at 25°C, with the solute position restraints removed. The temperature and pressure of the production simulations were controlled using the V-rescale thermostat(70) and Parrinello-Rahman barostat(71). All production simulations were carried out for 100 ns, each with an integration time-step of 2 fs and frame saving rate of 20 ps, and the last 50 ns were used for analysis. To allow for the use of the 2 fs integration timestep, in both the equilibration and production simulations, all-bonds lengths were constrained using the LINCS algorithim(72, 73). For all simulations the long-range electrostatics were controlled using PME as implemented within GROMACS(74).

Cellulose Simulation Analysis. Theoretical SANS intensities were calculated employing the SASSENA software package(75), averaging over the last 50 ns from each independent trajectory. To model the deuteration conditions of the experiments, we assigned randomly the scattering length density of deuterium in the following way. (i) For cellulose in the co-solvent we modeled as deuterium two (out of three) exchangeable hydrogen atoms per glucose monomer; four (out of seven) non-exchangeable hydrogen atoms per glucose; 65% of the water hydrogen atoms; all the THF hydrogen atoms. (ii) For cellulose in D_2O we modeled as deuterium two (out of three) exchangeable hydrogen atoms per glucose monomer; four (out of seven) non-exchangeable hydrogen atoms per glucose; 85% of the water hydrogen atoms.

Cellulose solvent spatial density distributions were computed using the VMD Volmap tool(76) for each independent trajectory (with bin sizes of 1.5$Å^3$). The resulting distributions were converted to xyz

grid points with a cutoff of 1/3 the maximum isovalue and subjected to a hierarchical cluster analysis using the WEKA toolkit(77-79) to identify high-density solvent-clusters. Gyration tensors for each cluster were then calculated using VMD.

Table S1*. Sequence of lignin used in the models. L and R refer to the optical chirality of the linkages*

	Unit	Linkages
1	G	β5R
2	G	βO4R
3	S	βO4L
4	S	βO4R
5	S	βO4L
6	G	$\beta\beta$
7	S	O4βR
8	S	O4βL
9	H	O4βR
10	S	O4βL
11	S	O4βR
12	G	

Isolated Lignin Simulations. A model of Poplar lignin was constructed with degree of polymerization of 12, containing seven syringyl (S), four guaiacyl (G) and one *p*-hydroxyphenyl (H) units, consistent with the experimental bulk chemical composition(80). The "sequence" of lignin is shown in Table S2. The lignin was fully acetylated by replacing H in hydroxyl groups with (CH_3)-C=O in a cubic box with a side length of 80 Å, and solvent molecules were added using Packmol(81). Three solvent cases were considered: THF:water 1:1 (v:v), pure THF, and pure water. The number of THF molecules was determined based on its liquid density at 25 ^{o}C (0.8833 g/cm^3). Each system was simulated with three different initial structures (i.e., instances), each instance for 300 ns. The CHARMM force fields were used for lignin(68) and THF(82), the TIP3P water(83) model. The lignin potentials associated with acetylation were parameterized using Ligand Reader & Modeler on CHARMM-GUI(84).

The last 200 ns of each instance were used for analysis, and all the calculated values represent averages over the three instances. The radial distribution functions were analyzed for non-hydrogen atoms between center of mass (COM) of lignin and solvents. Cryson(85) was employed for the SANS calculation using a fraction of 0.95 of D_2O, which yields an equivalent solvent scattering length density to 100% d_8-THF.

Lignin Aggregate Simulations. Systems containing three 12-mer acetylated lignin chains were solvated in each of three solvents: THF, water, and 1:1 v/v THF-water. The simulation settings were similar to those described above. Following energy minimization, five independent solvent relaxation simulations, i.e. simulations with the position restraints set for the lignin molecules, were performed for 2.5 ns within the canonical ensemble at 25 °C (with the temperature controlled using the Bernstein thermostat). Post-relaxation, each of the five relaxation simulations were continued as production simulations with position restraints removed from the lignin molecules within the canonical ensemble were performed using the V-rescale thermostat for 250 ns at 25 °C.

Supplementary Results

SANS of cellulose

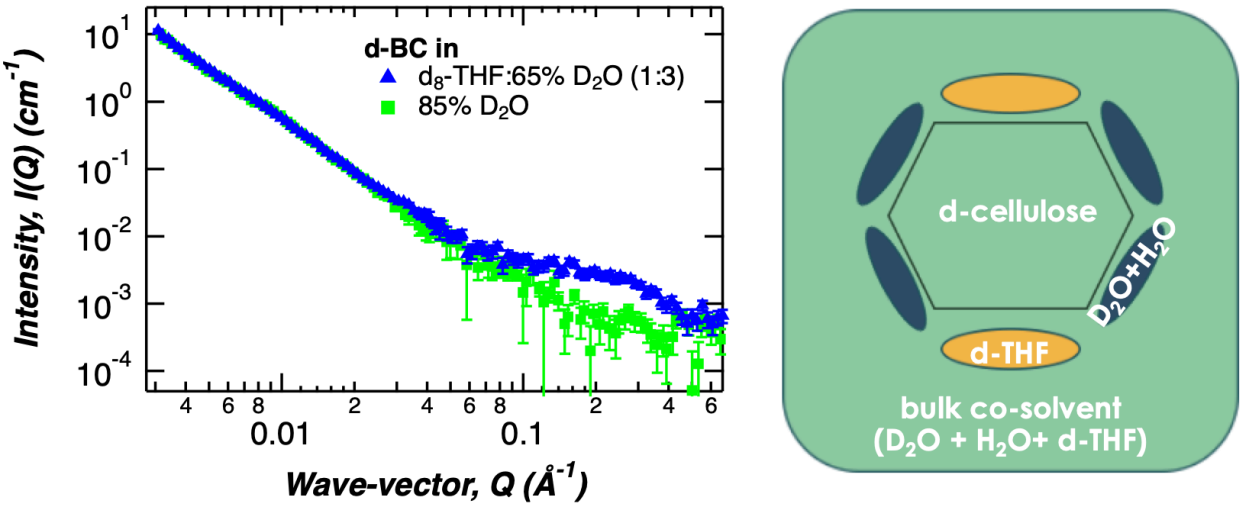

Figure S2*. (Left) Deuterated bacterial cellulose (d-BC) samples measured with SANS at 25 °C at the d-BC contrast match point, but at two different solvent compositions. (I, green) d-BC in 85% D_2O and 15% H_2O. (II, blue) d-BC in a 1:3 v/v mixture of perdeuterated THF (d_8-THF) and 65% D_2O (35% H_2O). The excess scattering plotted in Figure 2A was determined as the scattering intensity of (II)-(I). (Right) Sketch of the contrast conditions corresponding to sample (II). Since the cellulose and bulk solvent have similar scattering length density (green), the excess scattering arises from the THF-rich (orange) and water-rich (blue) nanodomains on the surface of the cellulose. The SANS data of partially deuterated bacterial cellulose in Fig. S2 were obtained after subtracting the scattering profile of bulk solvent mixture (65% D_2O: 100% d_8-THF at 3:1). The solvent scattering profile was scaled by 2-3% to account for the small excess incoherent scattering in the sample. Similarly, 85% D_2O solvent scattering was scaled by 4% to account for excess incoherent scattering in the sample aqueous sample.*

SANS of isolated lignin

SANS data were collected of lignin isolated from Poplar and acetylated in two solvent environments. The data from lignin in d_8-THF:D_2O co-solvent were fit to a Debye model(86) (Fig. S3):

$$I(Q) = A\left[exp\left(-Q^2R_g^2\right) + Q^2R_g^2 - 1\right]/Q^4R_g^4 + Bkg \quad \text{(S1)}$$

where A is a scaling factor, R_g is the radius of gyration and Bkg is the background. The Debye model assumes individually solvated molecules in a "theta" solvent that adopt random (Gaussian) conformations. In the limit of large Q ($Q>>R_g^{-1}$), Equation S1 scales as $I(Q)\sim Q^{-2}$. The fit to the data yields an approximate lignin R_g of 13.6 ±0.3 Å. Similar behavior of lignin, *i.e.* consistent with a Debye model, was found irrespective of whether lignin was stepwise added to the co-solvent, or the co-solvent was added to the lignin, or lignin was diluted from 10 mg/ml to 5 mg/ml. The forward scattering intensity is 0.0118 cm^{-1}, which for an ideal Gaussian coil model of the data, corresponds to a molecular weight of 1.73 kDa and contour length of 103 Å. These values agree with GPC molecular weight of ~ 1.7 kDa and contour length from MD of 120 Å.

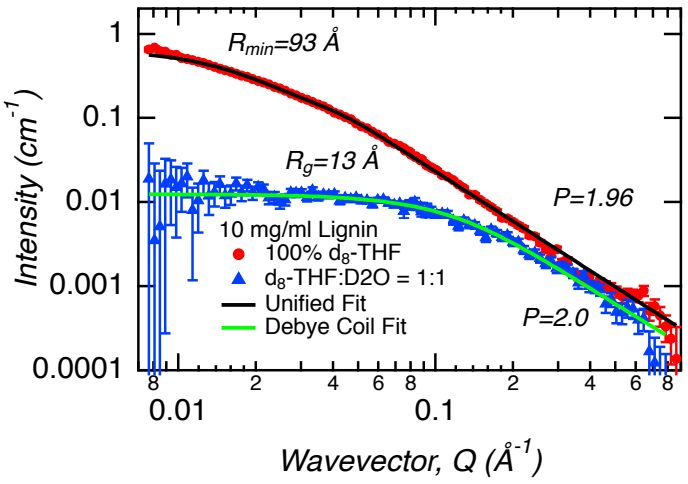

Figure S3. *SANS of acetylated Poplar lignin (10 mg/ml) in 1:1 (v:v) d_8-THF:D_2O co-solvent at 40 °C is shown in blue and a Debye fit (Equation S1) is shown as a green line. SANS of acetylated Poplar lignin (10 mg/ml) in d_8-THF at 40 °C is shown in red and the Unified (Equation S2) fit is shown as a black line.*

A Unified fit (Fig. S3)(87) was fit to the SANS data of lignin in d_8-THF:

$$I(Q) = Gexp\left(-Q^2R_{min}^2/3\right) + B\left\{\frac{[erf(QR_{min}/6^{1/2})]^3}{Q}\right\}^P + I_{Bkg} \quad \text{(S2)}$$

which has five fitting parameters: *G (scalar)* and R_{min} (here, since there is no plateau in the intensity R_{min} is the minimum size of the lignin particle) of the Guinier function; *B* (scalar) and *P* (the power-law exponent) of the power-law function; and I_{bkg} (the background intensity). In general, the exponent *P* indicates the bulk morphology of a polymer in solution: *P*=2 is consistent with a Gaussian chain, *P*=1.62 with a self-avoiding walk and *P*=3 with a collapsed chain.

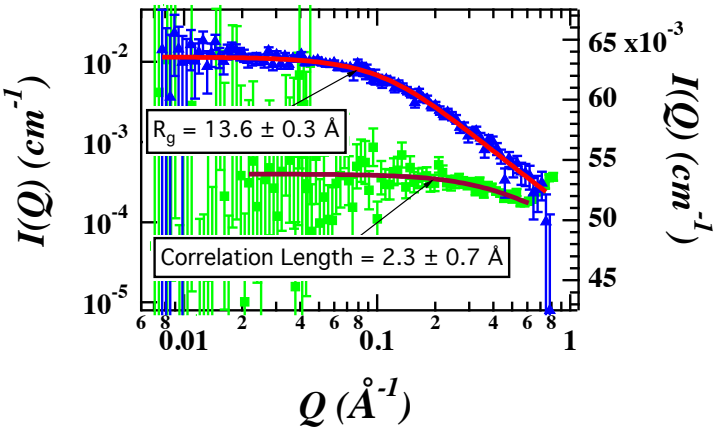

Figure S4*. SANS of acetylated Poplar lignin (10 mg/ml) in 1:1 (v:v) d_8-THF:D_2O co-solvent at 40 °C is shown in blue and a Debye fit (Equation S1) is shown as a light pink line (same as Fig. S3). SANS of 1:1 (v:v) d_8-THF:D_2O co-solvent at 40 °C (without lignin) is shown in green and a fit to the Ornstein-Zernike model is shown as a dark red line. The correlation length for THF:water solvent without lignin at 1:1 ratio (~20 wt%) is 2.3 ± 0.7 Å*

SANS data in 100% d_8-THF suggest of acetylated lignin forms aggregate particles, with a distribution of particle sizes. The R_{min} obtained from the fit is the lower limit of the particle dimension. The Unified fit results indicate an average $R_{min} \geq 93$ Å of the lignin aggregates, but the exact particle size cannot be obtained because there is no I(Q) plateau at small Q. The lignin molecules within these aggregate particles have sufficient solvent penetration and conform to a Gaussian random coil, evidenced by the exponent *P*

= 1.96 in the higher-Q region (similar to the exponent observed in the co-solvent system). The data follows a ~ Q^{-2} dependence at high Q, suggesting a random coil structure at small lengthscales. A Q^{-1} is found fat small Q, suggesting a rod-like aggregates of lignin chains.

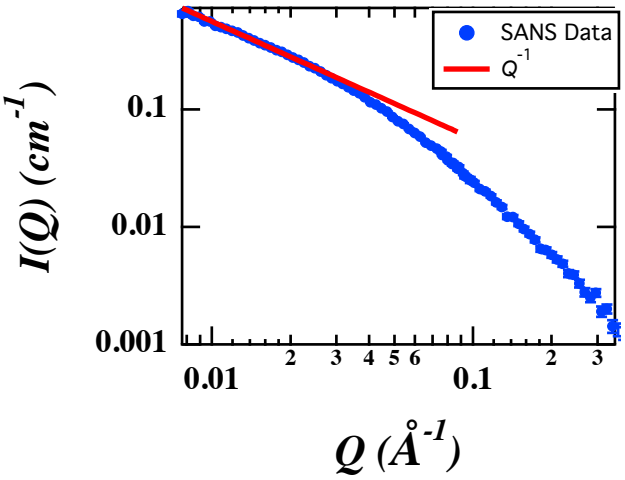

Figure S5*. SANS of acetylated Poplar lignin (10 mg/ml) in in d_8-THF at 40 °C is shown in blue (same as Fig. S3) and a Q^{-1} dependence, indicative of rod-like aggregates, is shown as a red line.*

MD of three lignin molecules

MD simulations were conducted of three acetylated lignin molecules in two solvent conditions: THF and THF:water. The lignins were separated in the beginning of the simulation. The number of inter-molecular contacts, *C* is a measure of their aggregation, defined here as the number of atomic pairs, the atoms belonging to two separate molecules, less than 0.5 nm apart. Averaging over the last 150 ns of the simulations, we obtained C=9.2±4.1 for simulations in THF:water and C=34.3±8.1 in THF.

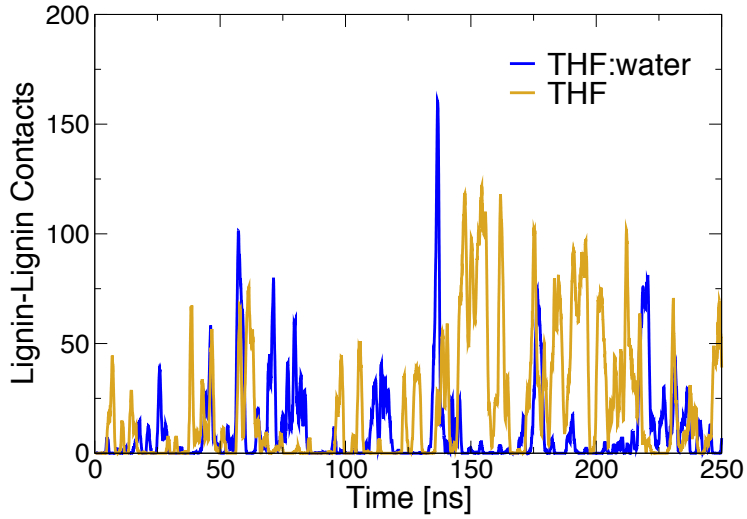

Figure S6*: Lignin-lignin inter-molecular contacts as a function of simulation time. Data represent the average of five simulations per solvent.*

References

53. Cao S, Pu Y, Studer M, Wyman C, & Ragauskas AJ (2012) Chemical transformations of Populus trichocarpa during dilute acid pretreatment. *Rsc Advances* 2(29):10925-10936.
54. Holtman KM, Chang Hm, Jameel H, & Kadla JF (2006) Quantitative 13C NMR characterization of milled wood lignins isolated by different milling techniques. *Journal of Wood Chemistry and Technology* 26(1):21-34.
55. Yoo CG, Li M, Meng X, Pu Y, & Ragauskas AJ (2017) Effects of organosolv and ammonia pretreatments on lignin properties and its inhibition for enzymatic hydrolysis. *Green Chemistry* 19(8):2006-2016.

56. O'Neill H*, et al.* (2015) Production of bacterial cellulose with controlled deuterium-hydrogen substitution for neutron scattering studies. *Methods Enzymol* 565:123-146.
57. He JH*, et al.* (2014) Controlled incorporation of deuterium into bacterial cellulose. *Cellulose* 21(2):927-936.
58. Shah R*, et al.* (2019) Hemicellulose-Cellulose Composites Reveal Differences in Cellulose Organization after Dilute Acid Pretreatment. *Biomacromolecules* 20(2):893-903.
59. Arnold O*, et al.* (2014) Mantid—Data analysis and visualization package for neutron scattering and SR experiments. *Nucl. Instrum. Methods Phys. Res., Sect. A* 764:156-166.
60. Nishiyama Y, Langan P, & Chanzy H (2002) Crystal structure and hydrogen-bonding system in cellulose 1 beta from synchrotron X-ray and neutron fiber diffraction. *J. Am. Chem. Soc.* 124(31):9074-9082.
61. Mostofian B*, et al.* (2016) Local Phase Separation of Co-solvents Enhances Pretreatment of Biomass for Bioenergy Applications. *J Am Chem Soc* 138(34):10869-10878.
62. Abraham MJ*, et al.* (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2:19-25.
63. Guvench O, Hatcher E, Venable R, Pastor R, & MacKerell A (2009) CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *Journal of Chemical Theory and Computation* 5(9):2353-2370.
64. Guvench O*, et al.* (2011) CHARMM Additive All-Atom Force Field for Carbohydrate Derivatives and Its Utility in Polysaccharide and Carbohydrate–Protein Modeling. *Journal of Chemical Theory and Computation* 7(10):3162-3180.
65. Vorobyov I*, et al.* (2007) Additive and Classical Drude Polarizable Force Fields for Linear and Cyclic Ethers. *J. Chem. Theory Comput.* 3(3):1120-1133.
66. Petridis L & Smith JC (2009) A Molecular Mechanics Force Field for Lignin. *J Comput Chem* 30(3):457-467.
67. Jorgensen W, Chandrasekhar J, Madura J, Impey R, & Klein M (1983) Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* 79(2):926-935.
68. Petridis L & Smith JC (2009) A Molecular Mechanics Force Field for Lignin. *J. Comput. Chem.* 30(3):457-467.
69. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, & Haak JR (1984) Molecular-Dynamics with Coupling to an External Bath. *J Chem Phys* 81(8):3684-3690.
70. Bussi G, Donadio D, & Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126(1).
71. Parrinello M & Rahman A (1981) Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. *J Appl Phys* 52(12):7182-7190.
72. Hess B, Bekker H, Berendsen HJC, & Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. *J Comput Chem* 18(12):1463-1472.
73. Hess B (2008) P-LINCS: A parallel linear constraint solver for molecular simulation. *J Chem Theory Comput* 4(1):116-122.
74. Abraham MJ & Gready JE (2011) Optimization of Parameters for Molecular Dynamics Simulation Using Smooth Particle-Mesh Ewald in GROMACS 4.5. *J Comput Chem* 32(9):2031-2040.

75. Lindner B & Smith JC (2012) Sassena - X-ray and neutron scattering calculated from molecular dynamics trajectories using massively parallel computers. *Comput Phys Commun* 183(7):1491-1501.
76. Humphrey W, Dalke A, & Schulten K (1996) VMD: visual molecular dynamics. *Journal of molecular graphics* 14(1):33-38.
77. Witten IH*, et al.* (1999) Weka: Practical machine learning tools and techniques with Java implementations.
78. Holmes G, Donkin A, & Witten IH (1994) Weka: A machine learning workbench.
79. Hall M*, et al.* (2009) The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter* 11(1):10-18.
80. Stewart JJ, Akiyama T, Chapple C, Ralph J, & Mansfield SD (2009) The Effects on Lignin Structure of Overexpression of Ferulate 5-Hydroxylase in Hybrid Poplar. *Plant Physiol* 150(2):621-635.
81. Martinez L, Andrade R, Birgin EG, & Martinez JM (2009) PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J Comput Chem* 30(13):2157-2164.
82. Vanommeslaeghe K*, et al.* (2010) CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* 31(4):671-690.
83. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, & Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79(2):926-935.
84. Kim S*, et al.* (2017) CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules. *J Comput Chem* 38(21):1879-1886.
85. Svergun DI*, et al.* (1998) Protein hydration in solution: Experimental observation by x-ray and neutron scattering. *Proceedings of the National Academy of Sciences of the United States of America* 95(5):2267-2272.
86. Debye P (1947) Molecular-weight determination by light scattering. *J Phys Colloid Chem* 51(1):18-32.
87. Beaucage G (1995) Approximations leading to a unified exponential power-law approach to small-angle scattering. *Journal of Applied Crystallography* 28:717-728.