SUPPORTING INFORMATION

Characterization and Molecular Simulation of Lignin in Cyrene Pretreatment of Switchgrass

Yun-Yan Wang[1], Yunxuan Wang[2], Luna Liang[2], Micholas Dean Smith[3,4], Xianzhi Meng[2], Yunqiao Pu[5,6*], Mitra Mazarei[7], Rupesh Agarwal[3,4], Shalini Jayaraman Rukmani[4], Brian H. Davison[5,6], Arthur J. Ragauskas[1,2,6*]

[1]Department of Forestry, Wildlife, and Fisheries, Center for Renewable Carbon, University of Tennessee Institute of Agriculture, Knoxville, TN 37996 (USA)

[2]Department of Chemical & Biomolecular Engineering, University of Tennessee Knoxville, Knoxville, TN 37996 (USA)

[3]Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, Knoxville, TN 37996 (USA)

[4]Center for Molecular Biophysics, Oak Ridge National Laboratory, Oak Ridge, TN 37831 (USA)

[5]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831 (USA)

[6]Joint Institute for Biological Sciences, Oak Ridge National Laboratory, Oak Ridge, TN, 37831 (USA)

[7]Department of Plant Sciences, University of Tennessee, Knoxville, TN 37996 (USA)

[*] Corresponding author

E-mail address: puy1@ornl.gov; aragausk@utk.edu

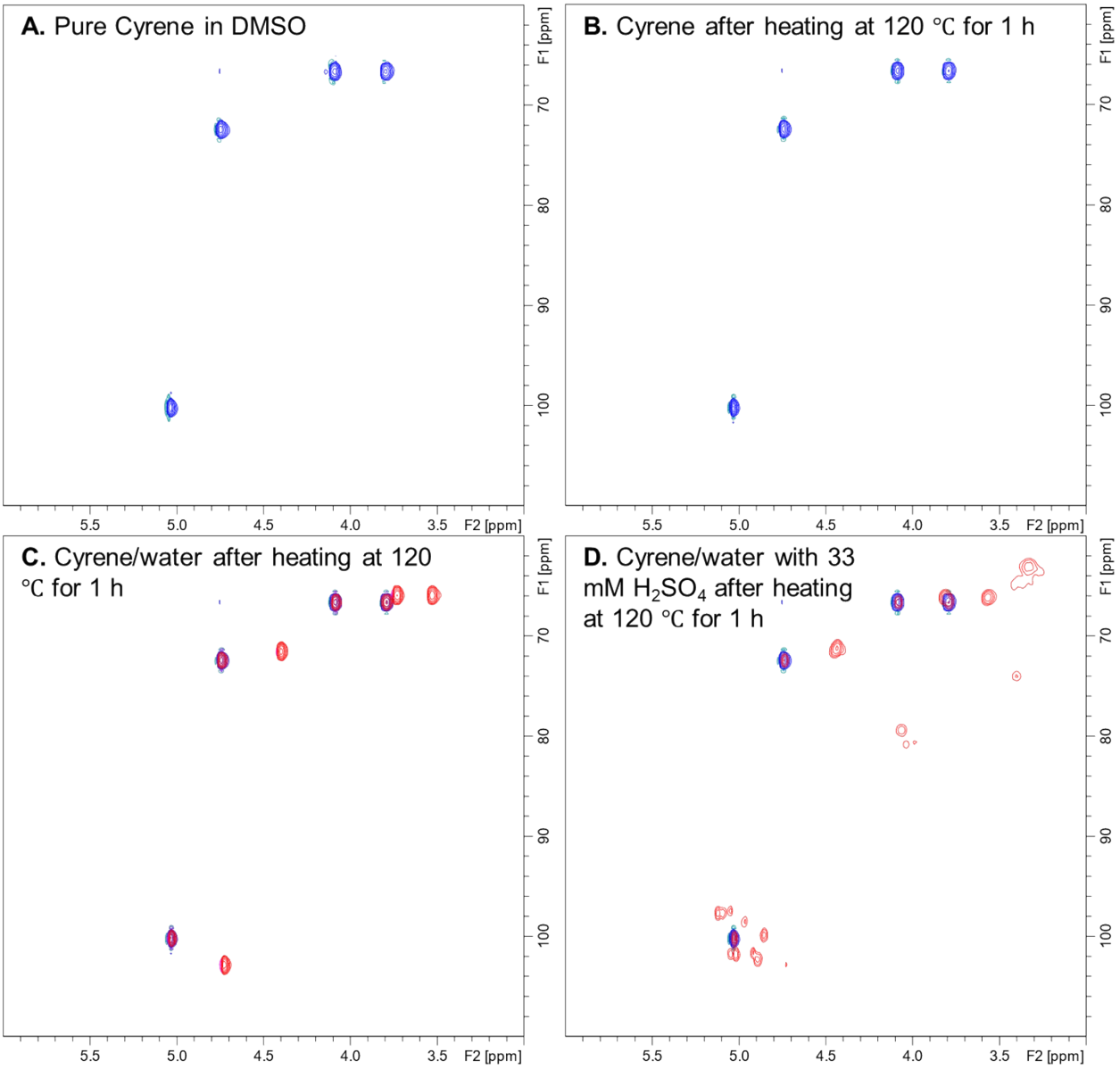

Figure S1. NMR spectra of solvent blank test. A: Pure Cyrene; B: Pure Cyrene after heating at 120 ℃ for 1 h; C: Overlap of the NMR spectra of pure Cyrene and mixture of Cyrene and water at 2:1 volumetric ratio after heating at 120 ℃ for 1 h; D: Overlap of the NMR spectra of pure Cyrene and mixture of Cyrene and water at 2:1 volumetric ratio with 33 mM H_2SO_4 after heating at 120 ℃ for 1 h. All NMR tests were conducted in DMSO.

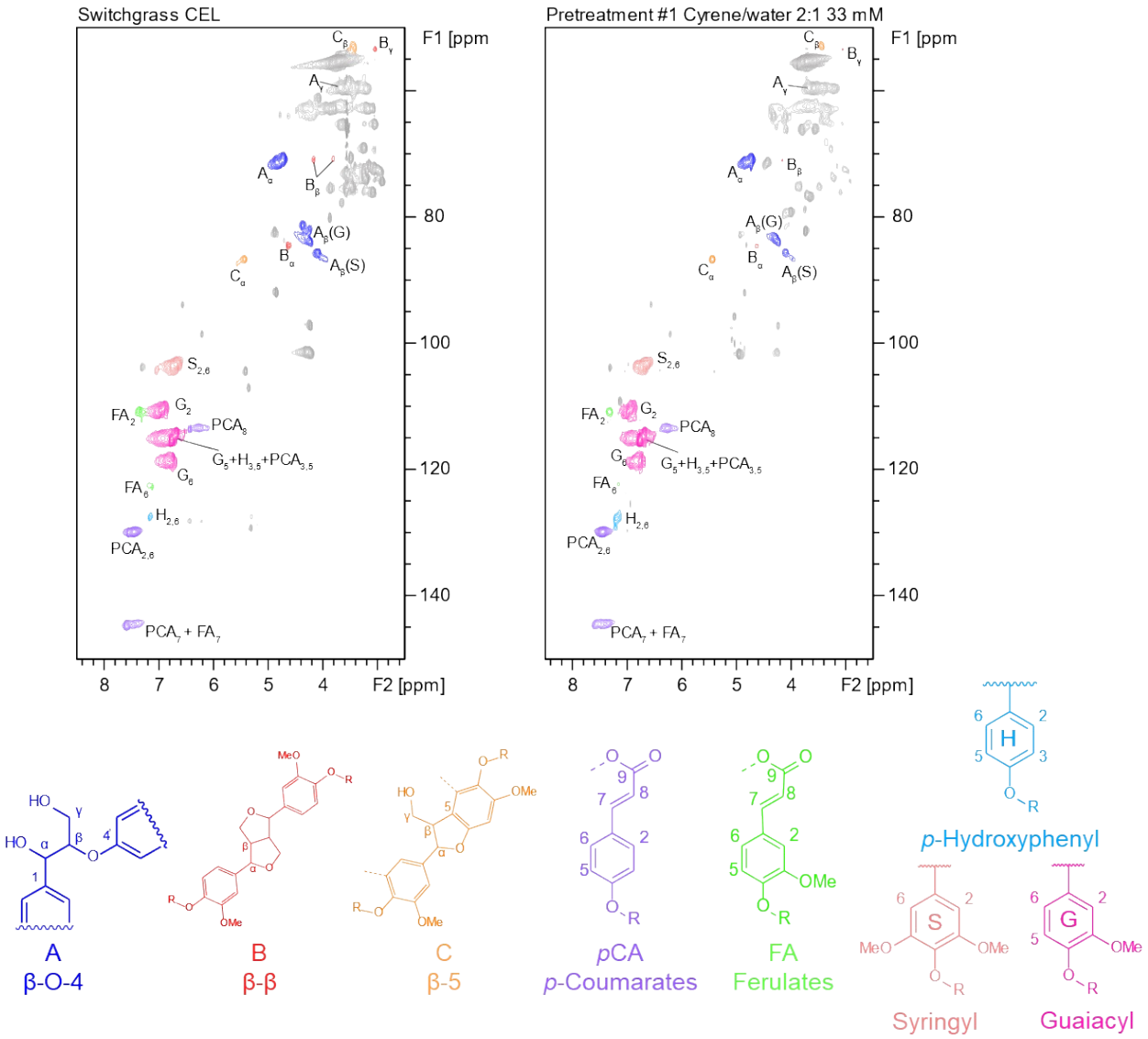

Figure S2. 2D HSQC NMR spectra of switchgrass CEL (left) and the lignin recovered after pretreatment #1, and structures of the corresponding subunits and inter-unit linkages.

Force-field parameter development for Cyrene and testing of said parameters

The following contains a description of force-field parameter development for Cyrene and testing of said parameters for studies with lignin.

In order to perform simulations of lignin in Cyrene, the development and testing of CHARMM-compatible force-field parameters was required. Here we describe the force-field parameter development strategy and two sets of validation simulations: (1) molecular dynamics-based viscosity calculations and (2) simulations of small lignin aggregate disassociation under both neat Cyrene and a previously published 4:1 Cyrene:water (v/v) ratio[1].

Cyrene Parameterization. To obtain a CHARMM-compatible parameterization of Cyrene, the CHARMM-GUI *Ligand Reader*[2] module was initially used to obtain atom-types, charges, bonds, angles, and dihedrals were assignments by analogy with ligands available within CGenFF[3]. The initial assignments showed high penalty scores (greater than 20) for both charges and bonded terms, and as a result further refinement was necessary.

To refine the force-field parameters, the force-field toolkit (FFTK[4]) within VMD was used while taking the parameter estimates provided by the CHARMM-GUI *Ligand Reader* as initial input guesses for Cyrene force-field optimization was undertaken. FFTK optimizes both non-bonded (electrostatic) and bonded (bonds, angles, and dihedrals) in a stepwise fashion, with charges initially fit following the standard CHARMM water interaction protocol, while bonds and angle terms are fit via comparison to quantum mechanical normal mode calculations (performed using Gaussian09 with the MP2/6-31G* basis set). Dihedral terms were obtained by through a direct comparison to quantum-mechanical (density functional theory based) torsional potential energy evaluation. For additional details on the FFTK procedure interested readers are directed to the original FFTK manuscript[4]. One of the most challenging portions of refinement with FFTK can be the assignment of dihedral (torsional) parameters; here we find that, upon parameter optimization, the molecular-mechanics force-field derived for Cyrene has excellent agreement with quantum-mechanical single-point energies across the conformations derived from a dihedral conformational sweep (Figure S3).

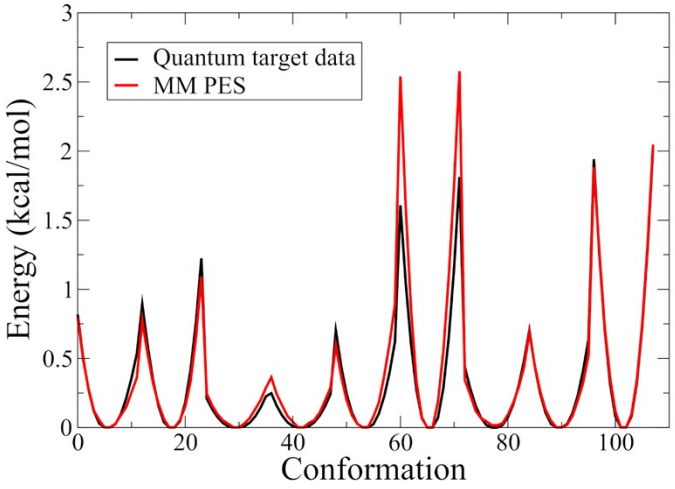

Figure S3. Comparison of Cyrene force-field energies to quantum-mechanical derived energies obtained from several dihedral angle conformations.

Validation Molecular Dynamics Protocol. To further test the validity of the derived Cyrene force-field, as noted above, two series of molecular dynamics (MD) simulations were performed, one focusing on neat Cyrene's viscosity and the other looking at the compatibly of the derived Cyrene force-field with an existing lignin force-field. For both series of simulations, the GROMACS simulation engine (version 2020)

was used[5-7]. All simulations make use of the Particle-Mesh-Ewald (PME) method[8], as implemented within GROMACS, with a short-range cutoff of 1.2nm. Van der waals interactions were treated with a cutoff of 1.2nm and the force-switch modifier with switching at 1.0nm. All simulations make use of the LINCS algorithm[9] to constrain hydrogen bond lengths, except in production simulations containing only Cyrene, where no constraints were used.

MD-based Cyrene Viscosity Estimation. Simulation boxes containing 729 Cyrene molecules with an initial volume of ~123.88nm^3 were generated using the gmx solvate utility within GROMACS. Following box generation, simulations were performed following a four-step procedure: (1) energy minimization, (2) NPT relaxation, (3) long-time NPT production, (4) secondary NVT production. With regards to energy minimization, to remove potential clashes resulting from the initial solvent box generation, energy minimization, using steepest-decent, was performed in single precision with a stop condition of a maximum force of less than 1 kJ mol^{-1} nm^{-1} or until the maximum force reaches (single) machine precision limits. Following energy minimization, a single NPT (isobaric-isothermal) with a target temperature of 293.15K and pressure of 1 bar was performed using the velocity rescaling (v-rescale) thermostat from Bussi[10] and the Berendsen barostat[11], for a total of 20ns with an integration step of 2fs. Post-relaxation, the final configurations of the relaxation simulations were used to seed 10-independent NPT simulations, using the v-rescale thermostat[10] and Parrinello-Rahman barostat[12]. These simulations were performed in double precision for 100ns each with a target temperature of 293.15K and pressure of 1bar and an integration step of 2fs. As an additional test of viscosity estimation, 100 constant volume production simulations, 10 seeded from each of the last 10ns of each NPT production simulation, were also performed using the same settings as the NPT simulations.

Viscosities were estimated directly from the fluctuations within the MD simulations using the Green-Kubo formalism[13, 14] as implemented within the GROMACS gmx energy utility. Figure S4 shows the results of these calculations.

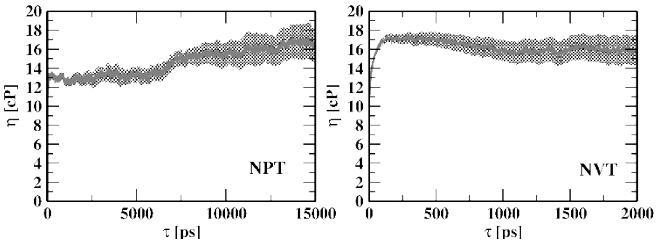

Figure S4. Viscosity estimation for neat Cyrene from NPT and NVT equilibrium simulations. Solid lines correspond to mean estimates over correlation times and shadows are standard errors of the mean estimates derived from the different independent simulation,

Prior experimental characterization of Cyrene at 293.15K suggests a viscosity near 14.5cP[15]. Our simulation results provide estimates between 13 and 16cP, suggesting that the force-field parameters developed are reasonable in representing Cyrene.

Lignin-aggregate dissociation under Cyrene and Cyrene-water conditions force-field test

To confirm the applicability of the developed Cyrene parameters in modeling the solvation of lignin under Cyrene and Cyrene-water conditions were also performed. These campaigns aimed to model the disassociation of low-MW lignin aggregates. For modeling purposes, the aggregates were made up of three, linear, 11-mer lignin chains composed of all G-subunits with a single β- 5, three 5-5, four β-O-4, and two α-O-4 inter-unit linkages and the force-field parameters were taken from Petridis and Smith[16]. Simulations themselves were prepared and performed following a four-stage process: initial solvent placement, energy minimization/clash removal, short NPT relaxation/equilibration, and production. Lignin aggregates are initially centered within boxes of size 729nm^3 and surrounded by (co)solvent using the gmx solvate utility. Post solvent addition, energy minimization is performed to within a 1 kJ mol^{-1} nm^{-1} tolerance or until the maximum force reaches machine precision limits. To further remove clashes and relax the simulation box dimensions, 10 independent NPT simulations, each 20ns in length, were performed using the energy minimized configuration as its initial coordinates and with velocities randomly assigned to correspond to 393.15K (120°C). The relaxation NPT simulations were performed with position restraints placed on each lignin atom and using the v-rescale thermostat and Berendsen barostat. The last frame from each of the 10 independent relaxation simulations were used as initial coordinates and velocities for use in 10 independent 55ns production simulations. Production simulations were performed at 393.15K in the NPT ensemble with the v-rescale thermostat[10] and Parrinello-Rahman barostat[12].

To confirm that lignin dissociation occurs for simulations using the generated Cyrene force-field, the minimal distances between each pair of lignin chain was computed every frame and histogrammed over five sliding 11ns windows. Figure S5 shows the results of this calculation. It can be clearly seen that in both bulk Cyrene and 4:1 v/v Cyrene-water mixtures, lignin readily disassociates, as shown by the peak minimal distance between any two lignin chains is greater than 2nm, within the simulation time Additionally, in both simulation conditions it is clear that the minimal distance distributions come to no longer change, i.e. have converged, after between 20 and 30ns. A similar analysis of the distributions of the radius of gyration (Figure S6) for individual chains shows that chain equilibration occurs within 20ns for mixtures of 4:1 cyrene-water, and near 30ns for under neat Cyrene conditions. Not only does this

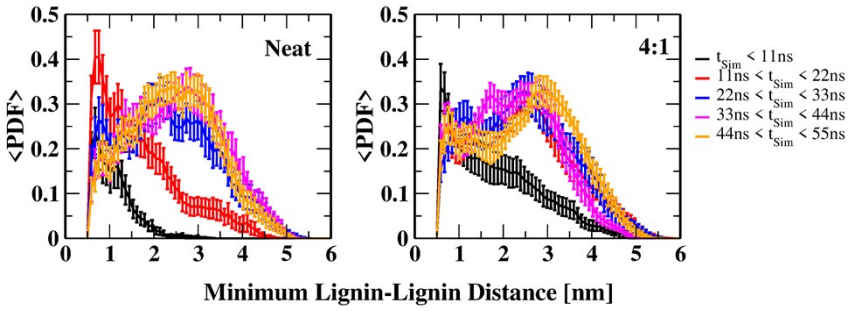

Figure S5. Average inter-lignin-lignin chain minimal distance distributions. Error-bars are standard error of the mean over 10 independent trajectories.

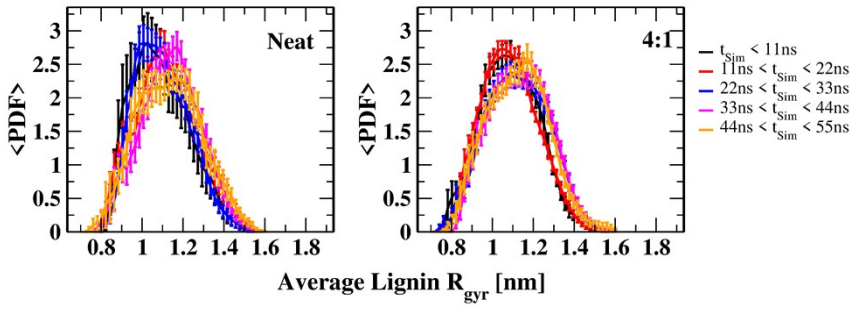

Figure S6. Average individual lignin chain radius of gyration distributions. Error-bars are standard error of the mean over 10 independent trajectories.

demonstrate that the Cyrene simulation parameters are reasonable in reproducing existing interactions with lignin, but further the necessary simulation time to examine a single chain may be as short as 30ns.

Cyrene force-field parameters (in GROMACS format):

```
[ moleculetype ]
; Name      nrexcl
Cyrene   3

[ atoms ]
; nr  type  resnr residue atom cgnr charge  mass
    1      CG321CY     1     CYRE     C1     1    -0.1840    12.0110
    2      CG321CY     1     CYRE     C2     1    -0.1350    12.0110
    3      CG2O5CY     1     CYRE     C3     1     0.4240    12.0110
    4      OG2D3CY     1     CYRE     O1     1    -0.4810    15.9990
    5     CG3RC1CY     1     CYRE     C4     1     0.1720    12.0110
    6     OG3C51CY     1     CYRE     O2     1    -0.3370    15.9990
    7     CG3C52CY     1     CYRE     C5     1     0.0140    12.0110
    8     CG3RC1CY     1     CYRE     C6     1     0.0590    12.0110
    9     OG3C51CY     1     CYRE     O3     1    -0.2520    15.9990
   10       HGA2CY     1     CYRE     H1     1     0.0900     1.0080
   11       HGA2CY     1     CYRE     H2     1     0.0900     1.0080
   12       HGA2CY     1     CYRE     H3     1     0.0900     1.0080
   13       HGA2CY     1     CYRE     H4     1     0.0900     1.0080
   14       HGA1CY     1     CYRE     H5     1     0.0900     1.0080
   15       HGA2CY     1     CYRE     H6     1     0.0900     1.0080
   16       HGA2CY     1     CYRE     H7     1     0.0900     1.0080
   17       HGA1CY     1     CYRE     H8     2     0.0900     1.0080

[ pairs ]
```

; ai aj func

11 12 1

11 13 1

3 11 1

8 12 1

8 13 1

3 8 1

10 12 1

10 13 1

3 10 1

2 17 1

2 7 1

2 9 1

11 17 1

7 11 1

9 11 1

10 17 1

7 10 1

9 10 1

1 4 1

1 5 1

4 12 1

5 12 1

4 13 1

5 13 1

2 14 1

2 6 1

4 14 1

4 6 1

4 9 1

7 14 1

3 7 1

8 14 1

5 15 1

5 16 1

1 15 1

15 17 1

9 15 1

1 16 1

16 17 1

9 16 1

1 6 1

6 17 1

5 17 1

[bonds]

; i j func

1 2 1

1 10 1

1 11 1

1 8 1

2 12 1

2 13 1

2 3 1

3 4 1

3 5 1

5 6 1

5 14 1

5 9 1

6 7 1

7 8 1

7 16 1

7 15 1

8 9 1

8 17 1

[angles]

; i j k func

2 1 10 5

2 1 11 5

2 1 8 5

10 1 11 5

8 1 10 5

8 1 11 5

12 2 13 5

3 2 12 5

1 2 12 5

3 2 13 5

1 2 13 5

1 2 3 5

2 3 4 5

4 3 5 5

2 3 5 5

3 5 6 5

3 5 14 5

3 5 9 5

6 5 14 5

6 5 9 5

9 5 14 5

5 6 7 5

6 7 8 5

6 7 16 5

6 7 15 5

8 7 16 5

8 7 15 5

15 7 16 5

1 8 9 5

1 8 7 5

1 8 17 5

7 8 9 5

9 8 17 5

7 8 17 5

5 9 8 5

[dihedrals]

; i j k l func

10 1 2 12 9

10 1 2 13 9

10 1 2 3 9

11 1 2 12 9

11 1 2 13 9

11 1 2 3 9

8 1 2 12 9

8 1 2 13 9

8 1 2 3 9

2 1 8 9 9

2 1 8 7 9

2 1 8 17 9

10 1 8 9 9

10 1 8 7 9

10 1 8 17 9

11 1 8 9 9

11 1 8 7 9

11 1 8 17 9

12 2 3 4 9

12 2 3 5 9

13 2 3 4 9

13 2 3 5 9

1 2 3 4 9

1 2 3 5 9

4 3 5 6 9

4 3 5 14 9

4 3 5 9 9

2 3 5 6 9

2 3 5 14 9

2 3 5 9 9

3 5 6 7 9

14 5 6 7 9

9 5 6 7 9

3 5 9 8 9

6 5 9 8 9

14 5 9 8 9

5 6 7 8 9

5 6 7 16 9

5 6 7 15 9

6 7 8 1 9

6 7 8 9 9

6 7 8 17 9

16 7 8 1 9

16 7 8 9 9

16 7 8 17 9

15 7 8 1 9

15 7 8 9 9

15 7 8 17 9

1 8 9 5 9

7 8 9 5 9

17 8 9 5 9

[dihedrals]

3 2 1 4 2

3 1 2 4 2

References:

1. Meng, X.; Pu, Y.; Li, M.; Ragauskas, A. J., A biomass pretreatment using cellulose-derived solvent Cyrene. *Green Chemistry* **2020,** *22* (9), 2862-2872.
2. Kim, S.; Lee, J.; Jo, S.; Brooks III, C. L.; Lee, H. S.; Im, W., CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules. *J. Comput. Chem.* **2017,** *38* (21), 1879-1886.
3. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell Jr., A. D., CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010,** *31* (4), 671-690.
4. Mayne, C. G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J. C., Rapid parameterization of small molecules using the force field toolkit. *J. Comput. Chem.* **2013,** *34* (32), 2757-2770.
5. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J., GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005,** *26* (16), 1701-1718.
6. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1*, 19-25.

7. Kutzner, C.; Páll, S.; Fechner, M.; Esztermann, A.; de Groot, B. L.; Grubmüller, H., More bang for your buck: Improved use of GPU nodes for GROMACS 2018. *J. Comput. Chem.* **2019,** *40* (27), 2418-2431.
8. Abraham, M. J.; Gready, J. E., Optimization of parameters for molecular dynamics simulation using smooth particle-mesh Ewald in GROMACS 4.5. *J. Comput. Chem.* **2011,** *32* (9), 2031-2040.
9. Hess, B., P-LINCS: A parallel linear constraint solver for molecular simulation. *Journal of chemical theory and computation* **2008,** *4* (1), 116-122.
10. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *The Journal of chemical physics* **2007,** *126* (1), 014101.
11. Berendsen, H. J.; Postma, J. v.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *The Journal of chemical physics* **1984,** *81* (8), 3684-3690.
12. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981,** *52* (12), 7182-7190.
13. Green, M. S., Markoff Random Processes and the Statistical Mechanics of Time-Dependent Phenomena. II. Irreversible Processes in Fluids. *The Journal of Chemical Physics* **1954,** *22* (3), 398-413.
14. Kubo, R., Statistical-Mechanical Theory of Irreversible Processes. I. General Theory and Simple Applications to Magnetic and Conduction Problems. *J. Phys. Soc. Jpn.* **1957,** *12* (6), 570-586.
15. Salavagione, H. J.; Sherwood, J.; De bruyn, M.; Budarin, V. L.; Ellis, G. J.; Clark, J. H.; Shuttleworth, P. S., Identification of high performance solvents for the sustainable processing of graphene. *Green Chemistry* **2017,** *19* (11), 2550-2560.
16. Petridis, L.; Smith, J. C., A molecular mechanics force field for lignin. *J. Comput. Chem.* **2009,** *30* (3), 457-467.