Supporting Information for

Novel Lignin Polymerrization Pathway Leading to Branching in the Structure

Seth Beck and Samir H. Mushrif*

Department of Chemical and Materials Engineering, University of Alberta, 116 St NW, Edmonton, AB, T6G 2E1, Canada

*Corresponding Author, Email: mushrif@ualberta.ca (SHM)

Computational Methodology

A guaiacyl and syringyl monolignol intermediate, containing a QM structure was used as one of the reactants, where the β-O-4 linkage leading to a dimer structure, was terminated with a hydrogen forming a β-OH group. The nucleophilic addition of the alcohol groups in coniferyl and sinapyl alcohols, the γ-OH and the 4-OH, to the monolignol intermediates were investigated. The absence of p-coumaryl alcohol as a monolignol and monolignol intermediate is justified as it is typically present in relatively small fractions in majority of biomass, $< 5\%$ [1].

All-electron DFT calculations were performed using the Gaussian 09 code [2]. For accurate thermochemistry calculations, conformational sampling was performed and details of the procedure is provided in the next paragraph. All data reported is at the M06-2X/6-311++G(d,p) level of theory as it has demonstrated accurate modeling of the carbohydrates present in lignocellulosic biomass [3, 4]. No constraints were implemented on the atoms during the geometry optimizations and were subsequently followed by frequency calculations to compute the thermochemical data and to verify that no spurious frequencies were present in the reactant and product compounds. The reaction free energies were calculated at 298K. The nucleophilic addition to the QM intermediate was previously investigated via step-wise and concerted mechanisms and it was concluded that a step-wise mechanism was kinetically favoured [4]. The step-wise mechanism assumes the phenolic oxygen at the 4 position in the QM intermediate is protonated in a barrier-less step [5] and the α carbon proceeds to undergo nucleophilic addition. The Berny algorithm was used for the transition state (TS) search. TS optimizations were followed by frequency calculations to confirm the presence of a single imaginary frequency, characteristic of a saddle point. A single explicit water molecule was used to help stabilize the proton transfer in the reaction mechanism, as done previously [5]. Intrinsic reaction coordinate (IRC) calculations were performed to ensure the identified TS corresponded to the anticipated reactant and product. The reactant compounds obtained from the IRC calculation underwent full geometry optimization and were used with the TS structures to determine the activation free energy of the reaction. To help discern the local environment of the reaction, the energetics are calculated in the gas phase to model a hydrophobic environment and an implicit solvent to model a solvated environment. The polarizable continuum model (PCM) using the integral equation formalism variant (IEFPCM) was used where solvent considerations are indicated in the reaction energetics. The dielectric constant of water was chosen for the implicit solvent in the continuum model (ε=78.3553) since lignin polymerization is understood to occur in an aqueous medium [6]. All structures in an implicit solvent underwent full geometry optimization and frequency calculations to verify the stationary points.

The conformational space of the product compounds was sampled using a M06-2X/6-31G level of theory and scanning over two torsion angles, depicted in Figure 2. Torsion angle 1 was rotated at 30° intervals, whereas torsion angle 2 was rotated at 90° intervals, to generate all possible structures within the 360° periodicity of each torsion angle. A crude potential energy surface was developed where all observable local minimums were selected for further optimization at the M06-2X/6-311++G(d,p) level of theory where no constraints were imposed. The lowest energy conformer at the M06-2X/6-311++G(d,p) level of theory was re-optimized incorporating an implicit solvent.

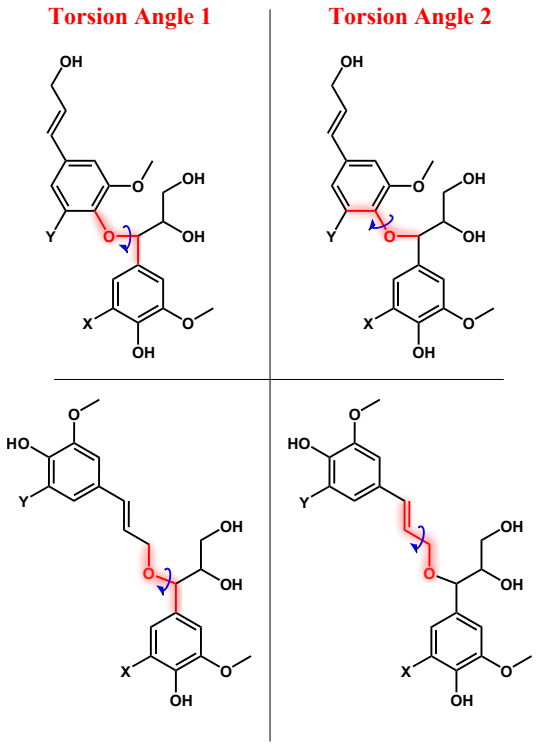

Figure S.1: The schematic of the torsion angles systematically rotated to generate an approximate potential energy landscape of the product compounds. Note, "X" and "Y" either represent a H group or an OCH3 group.

Further Details in the Results and Discussion

The reduced thermodynamic favorability observed in α-O-4 linkage is attributed to the increased rigidity associated with it. The α-O-4 linkage possesses limited flexibility and increases the steric interactions between the reactant moieties comprising the lignin linkage. The limited flexibility decreases the intramolecular hydrogen bonding strength and does not allow any aromatic ring stacking to promote favorable π-π interactions, reducing the thermodynamic stability, compared to what is observed with the α-O-γ lignin linkages (Figures S.2 to S.9).

The presence of solvent decreases the relative thermodynamic feasibility of forming both the α-O-γ and α-O-4 linkages. The reduced reaction free energies in the implicit solvent environment are likely a result of the shielding effect the dielectric medium has on steric interactions, hydrogen bonding and aromatic ring stacking, reducing the stability of the product compounds.

The reaction free energy of forming an α-O-γ lignin linkage possesses a monolignol compositional dependence. This dependence is inversely correlated with the number of methoxy substitutions on the monolignols participating in the reaction. Reducing the number of methoxy substitutions decreases the steric interactions, promoting favorable π-π interactions, providing stability to the product compounds and more favorable reaction free energies.

Visualized Geometries of α-O-4 and α-O-γ Lignin Linkages

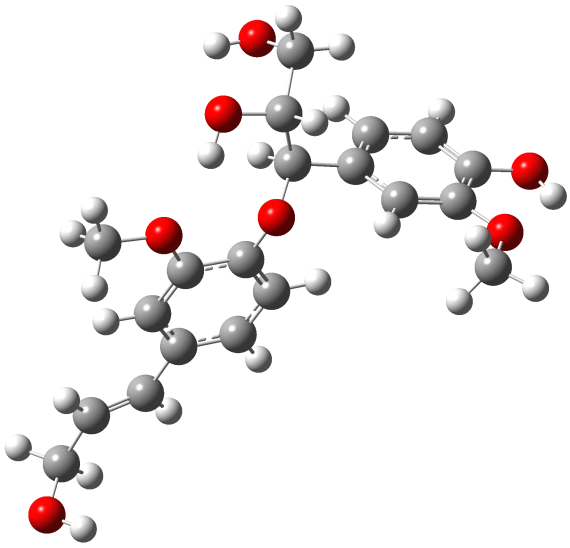

Figure S.2: The visualized product coordinates of the nucleophilic addition of coniferyl alcohol to a guaiacyl quinone methide intermediate forming an α-O-4 lignin linkage

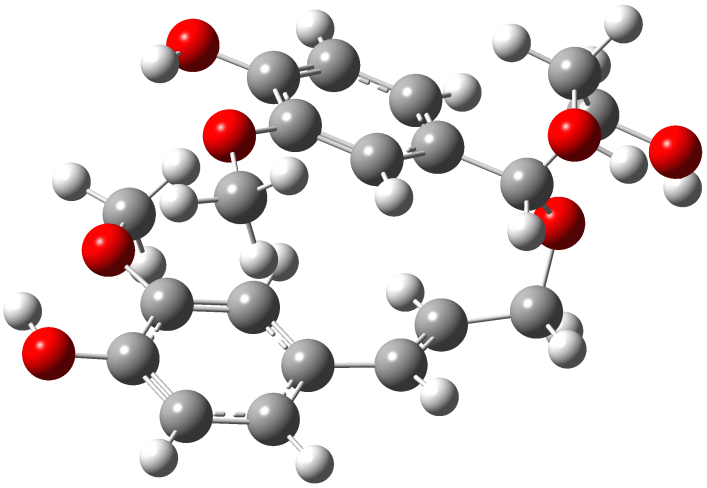

Figure S.3: The visualized product coordinates of the nucleophilic addition of coniferyl alcohol to a guaiacyl quinone methide intermediate forming an α-O-γ lignin linkage

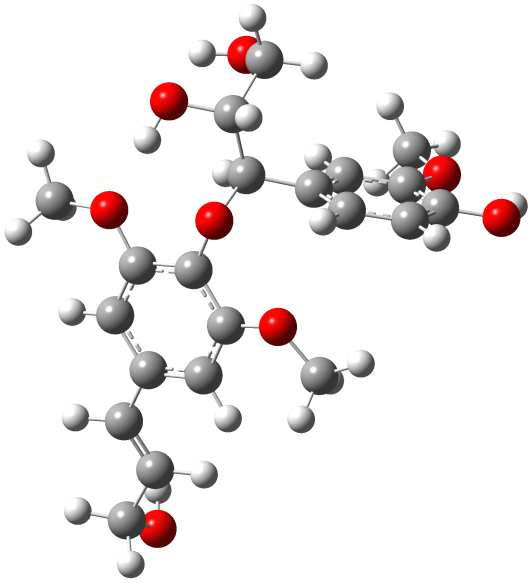

Figure S.4: The visualized product coordinates of the nucleophilic addition of sinapyl alcohol to a guaiacyl quinone methide intermediate forming an α-O-4 lignin linkage

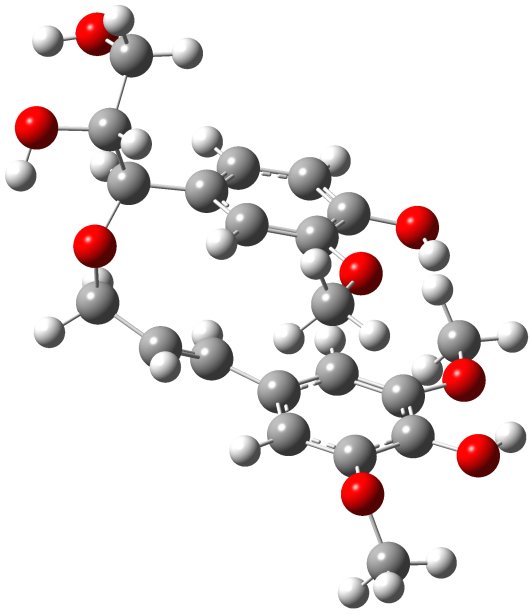

Figure S.5: The visualized product coordinates of the nucleophilic addition of sinapyl alcohol to a guaiacyl quinone methide intermediate forming an α-O-γ lignin linkage

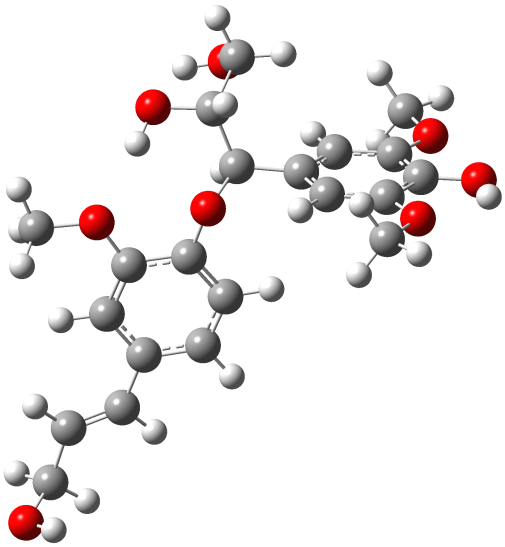

Figure S.6: The visualized product coordinates of the nucleophilic addition of coniferyl alcohol to a syringyl quinone methide intermediate forming an α-O-4 lignin linkage

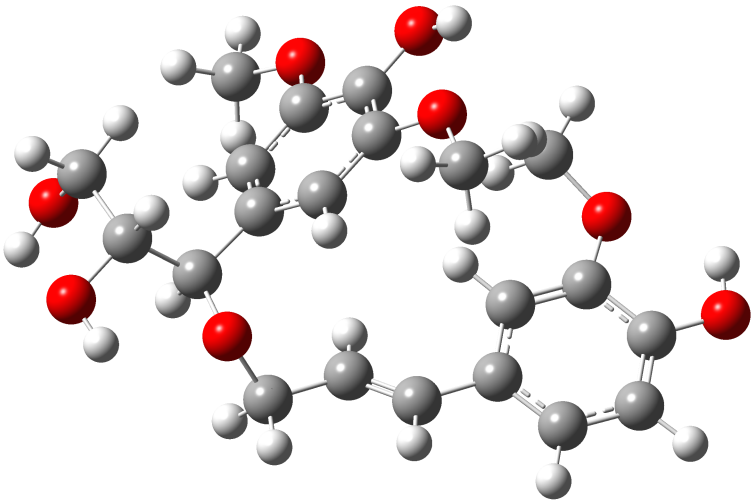

Figure S.7: The visualized product coordinates of the nucleophilic addition of coniferyl alcohol to a syringyl quinone methide intermediate forming an α-O-γ lignin linkage

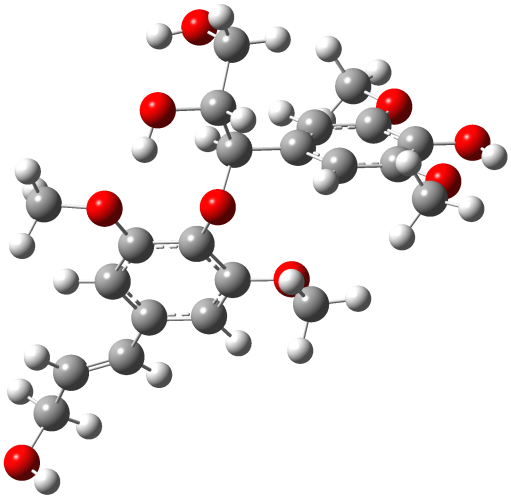

Figure S.8: The visualized product coordinates of the nucleophilic addition of sinapyl alcohol to a syringyl quinone methide intermediate forming an α-O-4 lignin linkage

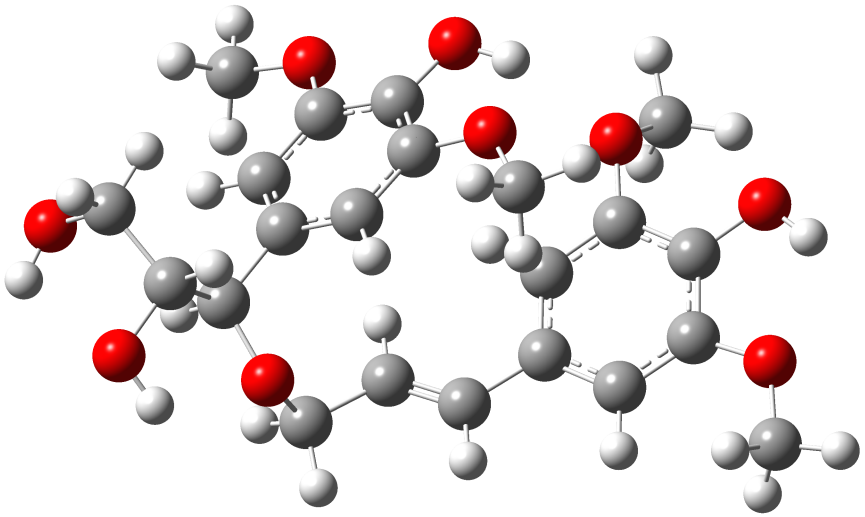

Figure S.9: The visualized product coordinates of the nucleophilic addition of sinapyl alcohol to a syringyl quinone methide intermediate forming an α-O-γ lignin linkage

Reaction Free Energy of Lignin Carbohydrate Complex (LCC) Linkages

Table 1: The reaction free energies (kJ/mol) at a M06-2X level of theory for nucleophilic addition to the QM monolignol intermediate, where the carbon number indicates the location of the nucleophile. Reproduced from [4]. Details of carbon numbering can be found in Figure S.10.

Nucleophile	6-311++G(d,p)	6-311++G(d,p) Solvent
Mannose-C6	-89.0	-58.9
Mannose-C2	-87.1	-66.6
Glucose-C6	-84.0	-54.2
Glucose-C3	-84.0	-59.3
Galactose-C6	-87.9	-60.1
Arabinose-C5	-82.2	-61.7
Xylose-C3	-91.4	-61.4
Glucuronic Acid-C6	-78.6	-50.2

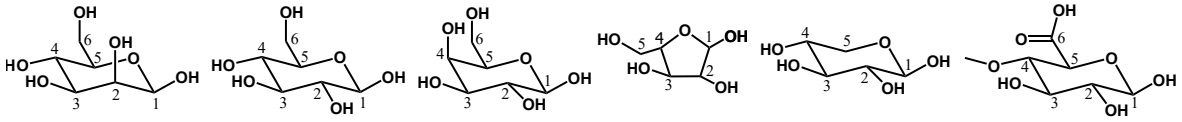

Figure S.10: Structures of the nucleophiles considered to form an LCC linkage with the monolignol QM intermediate and the carbon numbers associated with the compounds from [4].

References

[1] Ralph J, Lapierre C, Boerjan W. Lignin structure and its engineering. Current opinion in biotechnology. 2019;56:240–249.

[2] Frisch M, Trucks G, Schlegel H, Scuseria G, Robb M, Cheeseman J, et al. Gaussian 09, Revision D. 01, 2009, Gaussian. Inc, Wallingford CT. 2009.

[3] Zhao Y, Truhlar DG. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. Theoretical chemistry accounts. 2008;120(1):215–241.

[4] Beck S, Choi P, Mushrif SH. Origins of covalent linkages within the lignin–carbohydrate network of biomass. Physical Chemistry Chemical Physics. 2022;24(34):20480–20490.

[5] Durbeej B, Eriksson LA. Formation of β-O-4 lignin models-a theoretical study. Holzforschung. 2003;57(5):466–478.

[6] Boerjan W, Ralph J, Baucher M. Lignin biosynthesis. Annual review of plant biology. 2003;54(1):519–546.