Quantum mechanical calculations suggest that lignin polymerization is kinetically controlled

Terry Z. H. Gani¹, Michael J. Orella¹, Eric M. Anderson¹, Michael L. Stone¹, Fikile R. Brushett¹,
Gregg T. Beckham²,* and Yuriy Román-Leshkov¹,*

¹Department of Chemical Engineering, Massachusetts Institute of Technology, 25 Ames Street,
Cambridge, MA 02139, USA

²National Bioenergy Center, National Renewable Energy Laboratory, 15013 Denver West
Parkway, Golden, CO 80401, USA

Corresponding Authors

*email: vroman@mit.edu (Y.R.-L.)
*email: gregg.beckham@nrel.gov (G.T.B.)

Short title: First-principles calculations suggest lignification is kinetically controlled

Categories: Physical Sciences – Chemistry; Biological Sciences – Plant Biology

Keywords: density functional theory, lignin, polymerization, radical coupling

The authors declare no competing financial interest.
ABSTRACT

Lignin is an alkyl-aromatic biopolymer that, despite its abundance, is underutilized as a renewable feedstock because of its highly complex structure. An approach to overcome this challenge that has gained prominence in recent years leverages the plasticity and malleability of lignin biosynthesis to tune lignin structure in planta through genetic approaches. An improved understanding of lignin biosynthesis can thus provide fundamental insights critical for the development of effective tailoring and valorization strategies. Although it is widely accepted that lignin monomers and growing chains are oxidized enzymatically into radicals that then undergo kinetically-controlled coupling in planta, the adequacy of this model for predicting lignin structure remains unclear because of the difficulty of exactly replicating in planta lignification conditions and the widespread notion that radical coupling reactions are typically barrierless. Here, we provide comprehensive computational evidence for the role of kinetically-controlled radical coupling in governing lignin structure, ultimately explaining multiple key structural observations made empirically for native lignin in a variety of biomass types. Specifically, we develop first-principles models for a representative set of monolignol and dilignol radical coupling and transfer reactions that are fully consistent with experimentally observed structural characteristics. We also rationalize the origins of regioselectivity in coupling reactions through structural and activation strain analyses. Overall, our computational findings advance the understanding of lignin biosynthesis while paving the way for predictive first-principles kinetic models of lignification for effective lignin valorization strategies.

SIGNIFICANCE STATEMENT

Lignin is an abundant biopolymer important for plant function while holding promise as a renewable source of valuable chemicals. Although the lignification process in plant cell walls has been long-studied, a comprehensive, mechanistic understanding on the molecular scale remains elusive. A better understanding of lignification will lead to improved atomistic models of the plant cell wall that could, in turn, inform effective strategies for biomass valorization. Here, using first-principles quantum chemical calculations, we show that a simple model of kinetically-controlled radical coupling broadly rationalizes qualitative experimental observations of lignin structure across a wide variety of biomass types, thus paving the way for predictive, first-principles models of lignification while highlighting the ability of computational chemistry to help illuminate complex biological processes.
INTRODUCTION

Lignin is an abundant alkyl-aromatic biopolymer constituting up to 30% by weight of biomass(1) that strengthens cell walls, facilitates water transport, and inhibits microbial attack.(2) It is composed mainly of phenylpropanoid building blocks (i.e., monolignols) that give rise to eight known C-O or C-C linkages of widely varying strengths and characteristics.(2-4) Lignin offers tremendous potential value as a renewable resource(5) for bulk chemical(6-8), specialty chemical(9), and functional polymer(10) production, particularly if it can be efficiently depolymerized into monomeric units rather than lower-value oligomeric fractions or pyrolysis oil.(11) This is a formidable challenge primarily because of the high degree of structural complexity(2-4) that limits achievable monomer yields of selective depolymerization strategies(12, 13), notably those targeting the most abundant and labile β-O-4 (alkyl aryl ether) linkages.(14) Based on statistical arguments alone, monomer yields should increase with the proportion of β-O-4 linkages(15, 16), which should in turn be related to the method of lignin extraction(5) and the relative proportions of monomeric units with different propensities to form β-O-4 bonds.(4) However, recent experiments comparing monomer yields across various native(17) and extracted(18) lignin substrates have revealed surprising trends suggesting that post-depolymerization monomer yields are heavily influenced by additional, still unknown, factors beyond what is expected from a thermodynamically-controlled polymerization given a particular monolignol distribution in native lignin. Hence, a detailed understanding of lignin structure beyond the capabilities of modern analytical methods is critical for rationalizing these discrepancies and informing the development of more efficient depolymerization processes.

While empirical structural models have been developed(19-24) and successfully used to predict pyrolytic product distributions(25), their usefulness is limited by the accuracy of the underlying experimentally measured analytical properties that they are designed to reproduce. Inconsistencies among experimental data sources have necessitated empirical parameter tuning in some of these models, which limits predictive capabilities.(24) The development of predictive, first-principles structural models(26) has so far been hindered by the complexity of lignin biosynthesis, which involves a complex, spatiotemporally controlled interplay of metabolic pathways, monomer transport, and enzymatic and non-enzymatic polymerization reaction steps.(2-4, 27, 28) Briefly, monolignols are biosynthesized in the cytoplasm and transported to the cell wall through mechanisms that are not yet well understood. In the cell wall, laccase and peroxidase enzymes facilitate the oxidation of monolignols and growing chains into phenoxyl radicals. In alignment with the lack of optical activity in lignin(29) and the ease of incorporating nonstandard monolignols into growing chains(30), it is widely accepted that chain growth occurs by kinetically-controlled radical coupling, followed by aromaticity-restoring tautomerization and/or hydration. Spin delocalization throughout the conjugated carbon framework of each radical gives rise to a multitude of reactive sites(31) and possible coupling products. If this model holds, the relative proportions and sequences of observed linkages would then be governed by a combination of the relative propensities of radical coupling at each reactive site and the relative concentrations of reactive intermediates(4), suggesting that first-principles models of lignin structure can be developed by quantifying the rates of these processes.(26)

Due to challenges in studying lignification in planta at the molecular scale, conclusive evidence for kinetically controlled radical coupling has been scarce in the open literature. In vitro polymerization experiments(32-37) have shed light on lignin polymerization mechanisms but are inherently limited by the historical difficulty of quantifying complex product mixtures and the
inability to fully replicate in planta lignification conditions. In this regard, first-principles calculations can provide valuable, complementary insight. For instance, extensive study of the thermochemical properties of lignin linkages(38-46) has helped to identify the labile β-O-4 linkage as a target for depolymerization processes. However, with very few exceptions(47), computational studies of lignin polymerization have so far been limited to the thermodynamics(31, 48) of monolignol coupling reactions or the kinetics(49, 50) of β-O-4 coupling, neither of which provide a complete picture of lignification. In this work, we take a step towards a comprehensive first-principles structural model of lignin by characterizing the thermodynamics and kinetics for a representative set of radical formation and coupling reactions, which are shown to be fully consistent with qualitative experimental observations of lignin structure.

RESULTS AND DISCUSSION

We begin by considering the couplings of sinapyl alcohol (S) and coniferyl alcohol (G) radicals to determine if the product distributions of in vitro dimerization can be correctly predicted by computed relative activation energies, which would strongly suggest kinetic control. The G radical has 3 reactive sites (β, 5, and O) that give rise to 5 possible self-coupling products (β-β, β-O-4, β-5, 5-5, and 5-O-4, excluding O-O that yields an extremely unstable peroxide), whereas the S radical has 2 reactive sites (β and O; the 5 position is occupied by a methoxy group that blocks subsequent rearomatization) that give rise to 2 possible self-coupling products (β-β and β-O-4). These coupling products are known as quinone methides, unstable intermediates that readily undergo highly exothermic(31) hydration and tautomerization reactions in the presence of water to restore aromaticity. As these reactions are likely to be kinetically irrelevant, we instead focus on the initial radical coupling reactions, for which we gratifyingly observe significant differences in activation energies that are in excellent agreement with product distributions from in vitro dimerization experiments (Fig. 1). Notably, G-G couplings involving the β position (β-β, β-5, and β-O-4) have comparable barriers (ca. 2-4 kcal/mol) that are much smaller than the corresponding barriers for 5-5 and 5-O-4 couplings (ca. 10 kcal/mol) (Fig. 1, top). Indeed, horseradish peroxidase (HRP)-catalyzed bulk dehydrogenation polymerization (DHP) of G monomers in aqueous solution, in which dimerization is the dominant process, yields a mixture of β-β, β-5, and β-O-4 dimers in comparable amounts.(35) As the kinetic barriers for reactions at the β position are very low, dynamic effects that are outside the scope of this work may contribute in part to the observed selectivities, but do not affect our key conclusions. Radical coupling thermodynamics alone do not accurately describe selectivity, greatly overpredicting the favorability of β-β linkages, thus necessitating explicit consideration of individual activation energies over widely used Bell-Evans-Polanyi relationships.(51, 52) Interestingly, addition of methoxy groups to the G units disfavors β-O-4 coupling but not β-β coupling, exhibiting a larger difference in barrier heights for the S-S coupling reactions (ca. 1.4 kcal/mol for β-β coupling vs. 7.2 kcal/mol for β-O-4 coupling) (Fig. 1, bottom). This is again consistent with experimental observations of primarily the β-β dimer in bulk DHP of S monomers.(32, 35) Computed activation energies of S and G cross-couplings are also in good agreement with results of bulk polymerization of mixtures of S and G monomers (SI Appendix Fig. S1). DHP experiments in non-aqueous media result in significantly different product distributions(32, 36) that are well-captured computationally by changing the dielectric constant of the implicit solvent (SI Appendix Table S1), lending further support to the selectivity-determining nature of radical coupling.
The notion of kinetically controlled radical coupling appears to contradict the widely held expectation that radical coupling reactions should be intrinsically barrierless(53) because the interaction of two singly occupied molecular orbitals (SOMOs) should produce a doubly occupied MO that is more stable than either SOMO. Although small radical coupling barriers can sometimes result from disruption of intermolecular forces in the reacting complex (RC)(53), this is not the case here as the relative monomer orientations are preserved along the bond-forming reaction coordinates (SI Appendix Fig. S2). We thus applied the activation strain model (ASM)(54) to understand the regioselectivity accompanying these unusual barriers that appear to be intrinsic to bond formation. ASM decomposes the relative electronic energy, $\Delta E$, at any point along a reaction coordinate into a strain component, $\Delta E_{\text{str}}$, corresponding to the energy required to distort the reacting monomeric fragments from their RC geometry to their current geometry, and an interaction component, $\Delta E_{\text{int}}$, corresponding to the stabilization when the fragments interact at this geometry, i.e.,

$$\Delta E = \Delta E_{\text{str}} + \Delta E_{\text{int}}$$

As $\Delta E_{\text{str}}$ is always positive and $\Delta E_{\text{int}}$ is always negative, a transition state (TS) will result only if their rates of change balance each other somewhere along the reaction coordinate(54), and this is indeed the case for all coupling positions (Table 1). Although reactivity differences can sometimes be rationalized solely by differences in the contributions of either component(55), we find that this is not the case here. For instance, the $\beta-\beta$ reaction is most favorable overall despite its relatively large strain energy, and $\beta$-O-4 is much more favorable than 5-O-4 despite both reactions having comparable interaction energies. Rather, the observed regioselectivity of monolignol radical coupling appears to be controlled by a delicate balance between strain and interaction. Reactions at the $\beta$ and 5 positions experience greater strain than reactions at the O position because pyramidalization of the C atoms disrupts $\pi$ orbital conjugation and orbital overlap with adjacent atoms. At the $\beta$ position, this increase in $\Delta E_{\text{str}}$ is compensated by a more favorable interaction energy arising from an increase in C 2$p$ character of the SOMO that enhances spatial orbital overlap during bond formation (SI Appendix Fig. S3). The unusually low strain contribution to the $\beta$-5 reaction is attributable to a second hydrogen bond that enforces geometric similarity between the RC, TS and product (SI Appendix Fig. S2).

Known discrepancies between the structures of synthetic DHP products and native lignin motivate consideration of chain growth reactions (i.e., cross-coupling between monolignol and oligolignol radicals) in addition to monolignol coupling reactions. Notably, there are too few $\beta$-O-4 linkages in most DHPs because monolignol dimerization and its dominant $\beta$-$\beta$/$\beta$-5 coupling modes are overrepresented(4), as evidenced by the strong dependence of cross-coupling selectivity on the rate of monomer addition.(36) Furthermore, the strong preference for addition to the O position of the growing chain over the 5 position(36, 56) suggests an intrinsic difference in regioselectivities of monolignol coupling and chain growth beyond the inability of the oligolignol (growing chain) radical to couple at its $\beta$ position and explains why $\beta$-O-4 is the dominant linkage even in softwood (i.e., high G) lignins.(4) Thus, to evaluate the ability of kinetically controlled radical coupling to correctly predict these reactivity differences, we investigated the coupling of a G radical to a representative G-G $\beta$-$\beta$ resinol dimer (Fig. 2). In accordance with our expectations, we observe a slight preference for $\beta$-O-4 coupling over $\beta$-5 coupling for chain growth (ca. 3.8 kcal/mol vs. 4.4 kcal/mol), with 5-O-4 and 5-5 coupling remaining highly unfavorable. The enhanced $\beta$-O-4 selectivity is a result of slower $\beta$-5 coupling, rather than faster $\beta$-O-4 coupling,
and likely originates from the lack of a hydrogen bond donor to form a second hydrogen bond with the monolignol radical (*vide infra*).

Why, then, is a strong preference for monomer coupling over chain growth in DHP experiments typically observed? The differences between chain growth and monomer coupling barrier heights are too small for this effect to be attributable to intrinsic differences in radical coupling rates. Rather, it has long been hypothesized that monomer coupling dominates when there is a large difference between the concentrations of monolignol and oligolignol radicals.(4, 57) Indeed, chain growth and β-O-4 formation can be favored in DHP experiments by decreasing the rate of monomer addition(33, 36), decreasing the HRP concentration(58), introducing alternative peroxidases with enhanced affinities for growing chains(57, 59), or increasing the solubility of growing chains(37), all of which promote the formation of oligolignol radicals over monolignol radicals. These observations highlight the importance of determining rates of radical formation in first-principles kinetic models of lignification.

To this end, it was recently shown that, notwithstanding enzyme-binding effects, experimentally observed monolignol and dilignol reactivities toward HRP were correlated to the \( p \) orbital density of the phenolic oxygen.(47) Encouraged by this result, we computed activation energies for radical transfer from a \( p \)-coumaric acid radical to representative G and S monolignols and dilignols in order to further shed light on their relative propensities for radical formation (Fig. 3). Besides quantifying intrinsic dehydrogenation reactivities, these reactions also reasonably depict lignification in herbaceous feedstocks in which both growing chains as well as monolignols are oxidized primarily by radical transfer from \( p \)-coumarates.(60) Our calculations confirm that radical transfers to monolignols are facile and slightly faster for S radicals than G radicals. Analysis of TS geometries and electronic structures revealed that noncovalent interactions including \( \pi \) stacking favor asynchronous hydrogen atom transfer (HAT) reactions over stepwise proton-coupled electron transfers that would be more characteristic of phenoxy radicals(61) (SI Appendix Fig. S5). More importantly, we also found that radical transfers to growing chains are much less efficient than to monolignols, thus necessitating a very slow monomer addition rate to achieve reasonable chain lengths in the absence of external factors that specifically promote the oxidation of growing chains. To this effect, oligolignol-specific peroxidases have been identified in hardwoods(59) and softwoods(57) that may help ease the requirement for very slow monomer addition, which would also ostensibly lead to very slow lignification rates. In herbaceous lignins, extensive ferulate cross-linking(62) may also serve to reduce the number of difficult growing chain oxidations required for lignification. We further hypothesize that the recently noted inability of 5-O-4 and 5-5 linkages to undergo chain branching reactions in softwoods(63) can similarly be explained by an intrinsic lack of reactivity towards HAT reactions, and work along these lines is ongoing.

The small but significant fractions of 5-O-4 and 5-5 linkages (ca. 1%(63) and 4%(64), respectively) present in native softwood lignin pose a final test for kinetically controlled radical coupling, as their orders of magnitude are inconsistent with the differences in coupling activation energies calculated so far (e.g., \( \Delta H_a \) of 6 kcal/mol corresponds to a <0.1% ratio of relative rates at 25°C). As DHP of model dilignols in the absence of monolignols does indeed yield such linkages(36), it has been hypothesized that they result from the coupling of two growing chains, which can each only react at the 5 or O positions.(2, 4, 63) These couplings would be vanishingly unlikely in DHP experiments where monolignols can freely diffuse in solution together with growing chains, but may be facilitated by spatial confinement or mass-transfer regulation effects under *in planta* lignification conditions. However, the feasibility of this pathway would require
much lower intrinsic barriers for the coupling of two growing chains that would be predicted from monolignol coupling or chain growth. Indeed, we observe low barrier heights (ca. 2-3 kcal/mol) for the 5-5 and 5-O-4 coupling of two pinoresinol units that are comparable to barrier heights for monolignol coupling at the β position, confirming that growing chains can indeed easily couple if they are both oxidized and suitably oriented.

In conclusion, we have computed activation energies for a representative set of radical coupling and transfer reactions and, contrary to expectations that radical couplings should be barrierless, demonstrated how they can rationalize key structural observations of native lignin and synthetic dehydrogenation polymers. Overall, our findings constitute strong computational evidence that lignin structure is governed by kinetically controlled radical coupling, while confirming that differences between monolignol coupling, chain growth, and oligolignol coupling are important to lignification. Ongoing work in our groups is focused on developing predictive first-principles kinetic models of lignin structure to unravel the effects of monolignol composition and biosynthetic parameters on lignin structure and depolymerization yields.

METHODS

The computational methods are described fully in the SI Appendix. Briefly, all calculations were performed using ORCA 4.0.1. Geometries were optimized at the B3LYP-D3(69)/def2-SV(P)(70) level of theory, and single-point energies were calculated at the M06-2X(71)/def2-TZVP(70) level of theory, which has been shown to perform well for the thermochemistry of lignin models. Solvation contributions were treated implicitly with the CPCM model, and the choice of water (ε = 80.4) as the solvent was motivated by the aqueous environments of enzyme-catalyzed in planta and in vitro polymerizations. Experimentally observed variations in product distributions with solvent polarity are well captured computationally (SI Appendix Table S1). As rigorous conformational analyses of large and flexible oligomers are computationally intractable, we instead identified a single lowest-energy conformer for each reaction by DFT optimizing structures obtained from force field-based searches, as in prior studies. All calculations were spin-unrestricted, and transition states were located by scanning the singlet and broken-symmetry spin surfaces along the radical coupling reaction coordinates, followed by eigenvector following optimizations from the energy maxima.

ACKNOWLEDGMENTS

This work was funded by The Center for Bioenergy Innovation, a U.S. Department of Energy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. This work was carried out using computational resources from the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575, under allocations MCB-090159 and CTS-180027. M.J.O. was supported by the National Science Foundation Graduate Research Fellowship Program (award number 1122374).
REFERENCES


Figure 1. Monolignols and monolignol coupling reactions studied in this work. Each reaction is labeled with its reaction enthalpy ($\Delta H_r$, kcal/mol) and activation enthalpy ($\Delta H_a$, kcal/mol). Reactive positions of monomers and bonds formed in each coupling product are labeled in blue.
Figure 2. Representative chain growth reactions. Each reaction is labeled with its reaction enthalpy ($\Delta H_r$, kcal/mol) and activation enthalpy ($\Delta H_a$, kcal/mol). Reactive positions of coupling units and bonds formed in each coupling product are labeled in blue.
Figure 3. Representative radical transfer reactions. Each substrate is tabulated with its O-H bond dissociation energy (BDE, kcal/mol) and radical transfer activation enthalpy (ΔHₐ, kcal/mol). The O-H BDEs do not correlate well with radical transfer kinetics, suggesting that monolignol reactivity is primarily kinetic in origin.

<table>
<thead>
<tr>
<th>–R</th>
<th>BDE</th>
<th>ΔHₐ</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>83.7</td>
<td>0.9</td>
</tr>
<tr>
<td>S</td>
<td>79.8</td>
<td>0.6</td>
</tr>
<tr>
<td>GG</td>
<td>85.4</td>
<td>6.3</td>
</tr>
<tr>
<td>SS</td>
<td>81.8</td>
<td>2.2</td>
</tr>
</tbody>
</table>
Figure 4. Representative chain coupling reactions. Each reaction is labeled with its reaction enthalpy (ΔH_r, kcal/mol) and activation enthalpy (ΔH_a, kcal/mol). Reactive positions of coupling units and bonds formed in each coupling product are labeled in blue.
Table 1. Strain ($\Delta E_{str}$) and interaction ($\Delta E_{int}$) contributions in kcal/mol to the TS and product energies of radical coupling reactions. Plots of $\Delta E$, $\Delta E_{int}$ and $\Delta E_{str}$ along the bond forming reaction coordinates are provided in SI Appendix Fig. S3.

<table>
<thead>
<tr>
<th>Monomers</th>
<th>Linkage</th>
<th>Transition state</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\Delta E_{int}$</td>
<td>$\Delta E_{str}$</td>
</tr>
<tr>
<td>G-G</td>
<td>$\beta-\beta$</td>
<td>-5.9</td>
<td>8.1</td>
</tr>
<tr>
<td>G-G</td>
<td>$\beta-5$</td>
<td>-3.3</td>
<td>5.9</td>
</tr>
<tr>
<td>G-G</td>
<td>5-5</td>
<td>-0.6</td>
<td>10.5</td>
</tr>
<tr>
<td>G-G</td>
<td>$\beta-O-4$</td>
<td>-0.6</td>
<td>6.2</td>
</tr>
<tr>
<td>G-G</td>
<td>5-O-4</td>
<td>-0.3</td>
<td>10.0</td>
</tr>
<tr>
<td>S-S</td>
<td>$\beta-\beta$</td>
<td>-9.3</td>
<td>12.4</td>
</tr>
<tr>
<td>S-S</td>
<td>$\beta-O-4$</td>
<td>-1.3</td>
<td>9.2</td>
</tr>
</tbody>
</table>