

The Formation and Deconstruction of Lignin Carbohydrate Complexes (LCCs) in Lignocellulosic Biomass

by

Seth Beck

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Chemical Engineering

Department of Chemical and Materials Engineering

University of Alberta

Abstract

Current biomass fractionation technologies are not atom efficient, which is primarily a result of its recalcitrance. Although covalent linkages between lignin and cellulose/hemicellulose, commonly referred to as lignin carbohydrate complexes (LCCs), have been identified to directly correlate with biomass recalcitrance, their formation mechanism and role in deconstruction remains largely speculative. Hence, in this thesis, the kinetics and thermodynamics of the reaction pathways resulting in the formation of the various types of LCC linkages (benzyl ether, benzyl ester and phenyl glycoside (PG)) are quantified. For the stable LCC linkages that are anticipated to contribute to the biomass' recalcitrance the most, reaction mechanisms and energetics of the various deconstruction pathways are also computed. All-electron density functional theory (DFT) calculations were performed to quantify the reaction energetics as well as elucidate reaction pathways.

Independent of the source of biomass, the predominant lignin linkage is the β -O-4 linkage, which requires the re-aromatization of a quinone methide (QM) intermediate via electrophilic addition followed by nucleophilic addition. The electrophile is commonly assumed to be a proton from the acidic conditions of the plant cell wall and the nucleophile is assumed to be water. In such a case, both reaction sites would be terminated, and lignin would only interact physically with cellulose/hemicellulose. The present work explores the electrophilic and nucleophilic addition of hemicellulose, both of which would lead to LCC linkages as well as the nucleophilic addition of other monoglignols, leading to novel lignin linkages. The electrophilic addition of hemicellulose demonstrates a novel mechanism for PG formation and is kinetically facile and thermodynamically favoured (exergonic) over competing mechanisms at the reducing end of hemicellulose, thereby suggesting it is the likely mechanism resulting in the formation of PG linkages experimentally reported. The nucleophilic addition of hemicellulose is kinetically facile and thermodynamically more favorable (exergonic) than the nucleophilic addition of water, indicating the formation of benzyl ether and ester LCC linkages are a preferred synthesis route during lignin polymerization. However, formation

of the benzyl ester linkage is kinetically limited compared to the formation of benzyl ether linkages, suggesting the benzyl ester linkages are not abundant *in vivo*. Moreover, the nucleophilic addition of other monolignols is kinetically facile and exergonic, suggesting that the predominant β -O-4 linkage comprising lignin can act as a branching point via an α -O- γ lignin linkage.

Given the thermodynamic preference for benzyl ether linkages to occur along the hemicellulose chain and PG linkages to occur at the reducing end of hemicellulose, these linkages represent the most probable LCC linkages contributing to biomass recalcitrance. Therefore, the reaction mechanisms, kinetics and thermodynamics associated with the deconstruction of the benzyl ether and PG LCC linkages in biomass under acidic conditions are elucidated. Competing reactions are also essential to identify to ensure the desired deconstruction products are obtained during biomass deconstruction. As such, the possible reaction pathways identified include degrading the lignin structure, degrading the hemicellulose structure, or cleaving the LCC linkage. The deconstruction energetics demonstrate that breaking PG linkages is kinetically and thermodynamically favored in acid catalyzed deconstruction, indicating that PG linkages are unlikely to be the LCC linkage significantly contributing to biomass recalcitrance. Cleaving the benzyl ether linkages demonstrated the lowest activation barriers, however, possessed positive reaction energies (endergonic). Comparatively, the degradation of hemicellulose possessed higher activation barriers and the greatest thermodynamic feasibility, with negative reaction energies (exergonic). Computing the reaction energetics as a function of temperature suggested that increasing the temperature during the deconstruction of lignocellulosic biomass would result in the activation barriers associated with hemicellulose degradation being surmounted before favorable reaction free energies would be established for cleaving benzyl ether linkages. As a result, benzyl ether linkages are likely the LCC linkage significantly contributing to biomass recalcitrance and cleaving these linkages to obtain lignin and carbohydrates in their chemically intact form is a thermodynamically controlled process. The present work utilizes computational tools to pinpoint

the origin of biomass recalcitrance, and compute the preferred deconstruction pathways, providing insight into the convoluted structure of cellulose, hemicellulose and lignin within lignocellulosic biomass. Ultimately, this work will lay the foundation for designing novel deconstruction technologies and play a key role in making bio-derived specialty chemicals and materials a sustainable reality.

Preface

This thesis is the original work by the author, Seth Beck, under the supervision of Professor Samir H. Mushrif.

A version of Chapter 1 has been published as S. Beck, P. Choi, and S. H. Mushrif, “Physico-chemical interactions within lignocellulosic biomass and their importance in developing solvent based deconstruction methods,” *Reaction Chemistry & Engineering*, vol. 7, no. 12, pp. 2471–2487, 2022. Seth Beck conducted the literature review and wrote the manuscript. Phillip Choi and Samir Mushrif participated in writing of the manuscript.

A version of Chapter 2 has been published as S. Beck, P. Choi, and S. H. Mushrif, “Origins of covalent linkages within the lignin–carbohydrate network of biomass,” *Physical Chemistry Chemical Physics*, vol. 24, no. 34, pp. 20480–20490, 2022. Seth Beck conceptualized the project, performed the simulations, analyzed the data and wrote the manuscript. Phillip Choi and Samir Mushrif participated in conceptualizing the project, analyzing the data and writing of the manuscript.

A version of Chapter 3 has been published as S. Beck and S. H. Mushrif, “Reaction pathways and energetics of the deconstruction of lignin carbohydrate complexes (LCCs) in lignocellulosic biomass,” *Sustainable Energy & Fuels*, vol. 8, pp. 3113–3123, 2024. Seth Beck conceptualized the project, performed the simulations, analyzed the data and wrote the manuscript. Samir Mushrif participated in conceptualizing the project, analyzing the data and writing of the manuscript.

A version of Chapter 4 has been submitted as S. Beck and S. H. Mushrif, “Mechanistic insights into the formation and deconstruction of phenyl glycoside linkages in lignocellulosic biomass,” to *Reaction Chemistry & Engineering*. Seth Beck conceptualized the project, performed the simulations, analyzed the data and wrote the manuscript. Samir Mushrif participated in conceptualizing the project, analyzing the data and writing of the manuscript.

A version of Chapter 5 has been published as S. Beck and S. H. Mushrif, “Novel lignin polymerization pathway leading to branching in the structure,” *Organic Biomolecular Chemistry*, vol. 21, no. 22, pp. 4596–4600, 2023. Seth Beck conceptualized the project, performed the simulations, analyzed the data and wrote the manuscript. Samir Mushrif participated in conceptualizing the project, analyzing the data and writing of the manuscript.

A PhD is just an excuse to drink alcohol

-Logan Woolf

Acknowledgments

I extend my gratitude to supervisor Professor Samir Mushrif for his mentorship throughout my PhD. I am grateful for the high degree of academic freedom he provided, that allowed me to explore various research areas and topics at my own convenience. His willingness to discuss ideas and provide constructive feedback to me throughout the program created an invaluable opportunity for professional growth. Through his mentorship I have substantially elevated my knowledge base, critical thinking and research abilities as a whole.

I would like to thank Professor Phillip Choi for the many thought provoking discussions around research and philosophy that has largely shaped my perspective within academia. Special thanks to colleague Shang Jiang, for performing numerous molecular dynamic simulations whenever requested to assist in my exploration of new topics. Thanks goes out to Daniel Barker-Rothschild for our many discussions of lignin and sharing his joy of the field. I would like to thank Filipe Dos Santos De Araujo for being my best friend, in Brazil. I would also like to acknowledge the many individuals at the University of Alberta that sat with me during coffee breaks and lunches that listened to my jokes; however, only a special thanks goes out to the individuals who would laugh. Finally, I would like to express appreciation for my wife, Angelyn Beck, as well as my parents, Corey and Jan Beck, who have provided immense amounts of support throughout my PhD.

The work conducted in this thesis was funded by the Natural Sciences and Engineering Research Council (NSERC) of Canada, Canada First Research Excellence Fund as part of the University of Alberta's Future Energy Systems research initiative and the University of Alberta. I would also like to acknowledge financial support by Flora Greenhalgh, Gerald J Maier and NOVA Chemicals Corporation, Heng-Joo and Ku-Leong Ng as well as Dr William Youdelis. Computational resources were provided by Digital Research Alliance of Canada (the Alliance).

Table of Contents

1	Introduction	1
1.1	Motivation	1
1.2	Lignocellulosic Biomass	4
1.2.1	Composition	4
1.2.2	Lignin Synthesis	6
1.2.3	Microstructure: Cellulose, Hemicellulose & Lignin Interactions	8
1.2.3.1	Physical Interactions	8
1.2.3.2	Cellulose & Hemicellulose	8
1.2.3.3	Cellulose & Lignin	10
1.2.3.4	Lignin & Hemicellulose	11
1.2.3.5	Chemical Interactions (Lignin Carbohydrate Complexes (LCCs))	12
1.3	Deconstruction of Lignocellulosic Biomass	17
1.4	Summary of Chapter 1	20
1.5	Objectives & Scope of the Thesis	21
1.6	Organization of the Thesis	23
1.6.1	Chapter 2: Formation of Benzyl Ether LCC Linkages	24
1.6.2	Chapter 3: Deconstruction of Benzyl Ether LCC Linkages	25
1.6.3	Chapter 4: Formation & Deconstruction of Phenyl Glycoside LCC Linkages	25
1.6.4	Chapter 5: Branching in the Lignin Structure	26
2	Formation of Benzyl Ether LCC Linkages	27
2.1	Introduction	27
2.2	Computational Methodology	28
2.2.1	Model Compounds	28
2.2.2	Computational Procedure	28
2.3	Results & Discussion	31
2.3.1	Model Compounds	31
2.3.2	Benchmarking DFT Level of Theory	32
2.3.3	Thermochemistry of LCC Linkage Formation	34
2.3.4	Structural Isomerism	37
2.3.5	Kinetics & Mechanism of LCC Linkage Formation	38
2.3.6	Discussion	40
2.4	Conclusions	42

3	Deconstruction of Benzyl Ether LCC Linkages	43
3.1	Introduction	43
3.2	Computational Methodology	44
3.2.1	Model Compounds	44
3.2.2	Computational Procedure	45
3.3	Results & Discussion	46
3.3.1	Kinetics & Thermochemistry of Benzyl Ether LCC Deconstruction . .	46
3.3.2	Composition Dependence of Benzyl Ether LCC Deconstruction	49
3.3.3	Reaction Energetics Temperature Dependence	51
3.3.4	Discussion	54
3.4	Conclusions	57
4	Formation & Deconstruction of Phenyl Glycoside LCC Linkages	59
4.1	Introduction	59
4.2	Computational Methodology	61
4.2.1	Model Compounds	61
4.2.2	Computational Procedure	62
4.3	Results & Discussion	63
4.3.1	Hemi-Acetal & Transglycosylation Formation	63
4.3.2	Electrophilic Addition of Hemicellulose	65
4.3.3	Competing LCC Linkages	68
4.3.4	Deconstruction of PG LCC Linkages	71
4.3.5	Discussion	75
4.4	Conclusions	77
5	Branching in the Lignin Structure	79
5.1	Introduction	79
5.2	Computational Methodology	80
5.2.1	Model Compounds	80
5.2.2	Computational Procedure	81
5.3	Results & Discussion	82
5.3.1	Thermochemistry of α -O-4 & α -O- γ Lignin Linkages	82
5.3.2	Kinetics of α -O-4 & α -O- γ Lignin Linkages	83
5.3.3	Discussion	85
5.4	Conclusions	86
6	Conclusions	87
A	Molecular Modeling Techniques	111
A.1	Molecular Mechanics	112
A.2	Electronic Structure Methods	116
A.2.1	Hartree-Fock	116
A.2.2	Density Functional Theory (DFT)	119
A.2.3	Carr-Parrinello Molecular Dynamics & Metadynamics	122

B	Supporting Information for Chapter 2	124
B.1	Computational Details for Conformational Screening	124
B.2	Kinetic Data	127
C	Supporting Information for Chapter 3	129
D	Supporting Information for Chapter 4	131
D.1	Identifying Barrier via Scanning	131
D.2	Reaction Energetics Temperature Dependence	132

List of Figures

1.1	Visual depiction summarizing the different strategies to deconstruct lignocellulosic biomass.	4
1.2	a) Cellulose repeat unit. b) An example of a hemicellulose segment from softwood. The structure is not meant to represent a repeating unit, as the polymer shows significant chemical diversity. The orange wavy lines represents the remaining units in the hemicellulose structure. c) The three primary monolignols that polymerize to form lignin in secondary plant cell walls. . .	6
1.3	The five types of LCC linkages proposed. The LCC linkage is colored red. . .	14
1.4	The radical coupling reaction between a monolignol and ferulate moiety to form an LCC linkage.	15
1.5	The radical coupling reaction between two monolignols forms a lignin dimer, where the lignin intermediate forms a quinone methide (QM) structure that reacts with cellulose or hemicellulose to form a LCC linkage. A benzyl ether linkage is depicted.	16
1.6	An overview of the thesis organization.	24
2.1	General workflow to identify the lowest energy conformers for the model LCC moieties, <i>i.e.</i> , the simplified monolignol bonded to the monomer forms of cellulose and hemicellulose.	29
2.2	Benzyl ether formation with a mannose moiety. (a) The step-wise reaction mechanism, where the first step is protonation followed by the nucleophilic addition to the QM intermediate. (b) The concerted reaction mechanism, where protonation and nucleophilic addition to the QM intermediate occurs simultaneously. Note, dashed lines between molecules indicate hydrogen bonding.	30
2.3	Structures of the nucleophiles considered to form an LCC linkage with the monolignol QM intermediate and the carbon numbers associated with the compounds. Red groups indicate the reaction sites considered for the nucleophilic addition to the QM intermediate.	32
2.4	The relative reaction free energies for the nucleophilic addition of hemicellulose components to the QM intermediate, with the nucleophilic addition of water taken as the reference, as a function of the basis set. Data corresponds to gas phase calculations. The color coding of molecules is the same as in Figure 2.3.	33

2.5	The nucleophilic addition to the QM intermediate of all the compounds considered in the present chapter. (a) The thermochemical data corresponding to the most favorable reaction site for each compound in the gas phase and implicit solvent and the associated activation free energy. Each reaction is labeled with its reaction free energy (ΔG_r , kJ/mol) and the activation energy corresponding to the step-wise mechanism (ΔG_a^{SW} , kJ/mol). Where un-bracketed numbers indicate gas phase calculations and bracketed numbers indicate implicit solvent calculations. Reaction sites and LCC linkages are highlighted in red. (b) Reaction free energies of the formation of LCC linkages, relative to the nucleophilic addition of water. The color coding of molecules is the same as in Fig. 2.3.	36
2.6	Activation free energies (ΔG_a) for the step-wise (SW) and concerted (C) mechanisms for the nucleophilic addition of mannose to the QM intermediate. Un-bracketed numbers indicate gas phase calculations and bracketed numbers indicate implicit solvent calculations.	40
3.1	Structure of the (a) primary benzyl ether LCC linkage with mannose, (b) secondary benzyl ether LCC linkage with mannose, (c) secondary benzyl ether LCC linkage with xylose and (d) primary benzyl ether LCC linkage with arabinose as well as atomic numbering used to distinguish the reaction pathways, <i>e.g.</i> , the LCC linkage would be the $C_\alpha - O_1$ bond in every structure.	45
3.2	(a) The reaction pathways considered in the acid catalyzed deconstruction of the primary benzyl ether LCC linkage with mannose. (b) The activation free energy (ΔG_a , kJ/mol) and (c) reaction free energy (ΔG_r , kJ/mol), in gas phase, water, tetrahydrofuran (THF) and anisole, of the reaction pathways illustrated in (a).	48
3.3	(a) The two deconstruction pathways accessible, across varying hemicellulose compositions, when the oxygen participating in LCC linkages, <i>i.e.</i> , the O_1 reaction site, is protonated. (b) The activation free energy (ΔG_a , kJ/mol) and (c) reaction free energy (ΔG_r , kJ/mol), in gas phase, water, tetrahydrofuran (THF) and anisole, of the reaction pathways illustrated in (a).	50
3.4	The free energy barriers for cleavage of the various LCC linkages as well as the competing hemicellulose degradation pathways as a function of temperature, in the gas phase. The solid lines are calculated assuming the enthalpic (ΔH_a) and entropic (ΔS_a) contributions are constant with the values computed at 298K. Note, since no significant differences were observed in the reaction energetics when an implicit solvent was incorporated, only gas phase data is reported.	53
3.5	The reaction free energies for cleavage of the different LCC linkages as well as the competing hemicellulose degradation pathways as a function of temperature, in the gas phase. The solid lines are calculated assuming the enthalpic (ΔH_r) and entropic (ΔS_r) contributions are constant with the values computed at 298K. Note, since no significant differences were observed in the reaction energetics when an implicit solvent was incorporated, only gas phase data is reported.	54

4.1	Mechanisms proposed in the literature for the formation of phenyl glycoside (PG) linkages in biomass (a) hemi-acetal formation (b) transglycosylation (c) monolignol glucosides, and (d) a newly proposed mechanism of electrophilic addition of hemicellulose to the lignin quinone methide (QM) intermediate. The carbon numbering of the hemicellulose moiety (mannose) is shown in the hemi-acetal mechanism (a).	61
4.2	The free energy pathway (in kJ/mol) of phenyl glycoside (PG) linkage formation with mannose and xylose via the hemi-acetal mechanism. Note (a): Activation free energy increases to greater than 150 kJ/mol without an acid catalyst present.	64
4.3	The free energy pathway (in kJ/mol) of phenyl glycoside (PG) linkage formation with mannose and xylose via the transglycosylation mechanism.	65
4.4	Re-aromatization of the quinone methide (QM) lignin intermediate via electrophilic addition at the carbonyl group (C ₄ position) followed by nucleophilic addition at the α carbon.	66
4.5	Free energy pathway (in kJ/mol) of the re-aromatization of the quinone methide (QM) intermediate via electrophilic addition of mannose at the different carbon locations, forming a phenyl glycoside (PG) linkage, followed by nucleophilic addition of water to the α carbon on the lignin intermediate.	67
4.6	Free energy pathway (in kJ/mol) of the re-aromatization of the quinone methide (QM) intermediate via electrophilic addition of xylose at the different carbon locations, forming a phenyl glycoside (PG) linkage, followed by nucleophilic addition of water to the α carbon on the lignin intermediate. Note (a): The barrier was determined via scanning the distance between the C ₁ location on xylose and the oxygen at the C ₄ location on lignin. For further computational details on the calculation of this barrier please see Appendix D.	68
4.7	Free energy pathway (in kJ/mol) for the formation of a PG LCC linkage via electrophilic addition of mannose at the C ₁ location followed by nucleophilic addition of water to the lignin QM intermediate, compared to the reaction energetics for the formation of a benzyl ether LCC linkage via electrophilic addition of a proton followed by nucleophilic addition of mannose at the C ₁ location to the lignin QM intermediate. Note (a): The electrophilic addition of a free proton is a barrierless step and occurs spontaneously, therefore, the structures denoted by BE0 and BE1 are equivalent and only used for illustration purposes.	70

4.8	The free energy pathway (in kJ/mol) for the formation of a PG LCC linkage via electrophilic addition of xylose at the C ₁ location followed by nucleophilic addition of water to the lignin QM intermediate, compared to the reaction energetics for the formation of a benzyl ether LCC linkage via electrophilic addition of a proton followed by nucleophilic addition of xylose at the C ₁ location to the lignin QM intermediate. Note (a): The barrier was determined via scanning the distance between the C ₁ location on xylose and the oxygen at the C ₄ location on lignin. For further computational details on the calculation of this barrier please see the SI. Note (b): The electrophilic addition of a proton is a barrierless step and occurs spontaneously, therefore, the structures denoted by BE0 and BE1 are equivalent and only used for illustration purposes.	71
4.9	(a) The reaction pathways considered in the acid catalyzed deconstruction of the PG LCC linkage with mannose and xylose. (b) Activation free energies (in kJ/mol) and (c) reaction free energies (in kJ/mol) of the reaction pathways illustrated in (a).	74
5.1	The reaction mechanism to synthesize the β-O-4 linkage as well as the re-aromatization of the QM intermediate that leads to various α-O-H/α-O-4/α-O-γ based structures. Note, “X” and “Y” either represent a H group or an O – CH ₃ group.	80
5.2	The schematic of the torsion angles systematically rotated to generate an approximate potential energy landscape of the product compounds.	82
5.3	The nucleophilic addition of coniferyl alcohol (CA), sinapyl alcohol (SA) and water to a guaiacyl (G) and syringyl (S) QM intermediate. The nucleophilic addition of the γ-OH as well as the 4-OH of coniferyl and sinapyl alcohol are considered. (a) Each reaction is labeled with its reaction free energy (ΔG _r , kJ/mol) and activation free energy (ΔG _a , kJ/mol) with the labels corresponding to the legend in (b). Un-bracketed numbers indicate gas phase calculations and bracketed numbers indicate implicit solvent calculations. (b) Reaction free energies leading to α-O-4 and α-O-γ lignin linkages, relative to the nucleophilic addition of water to each respective monolignol intermediate. The reaction free energies reported for the formation of benzyl ether (BE) LCC linkages as well as the nucleophilic addition of water to the guaiacyl QM intermediate, is taken from Chapter 2. Although the reaction free energy of BE LCC linkage formation depends on the composition of nucleophile, the most favored BE linkage site is selected.	84
A.1	The multiscale nature of the research problem pertaining to understanding the biomass microstructure and to develop solvent based deconstruction methods. Modeling tools that can be used for different time and length scales are also listed.	112
A.2	A visual depiction of the bonded and non-bonded energy terms accounted for in a force field.	113

B.1	A schematic of the collective variables (CVs) used in the CPMD-metadynamics calculations. Depending on the connection of the sugar molecule, torsion angle 2 was varied.	125
B.2	The schematic of the torsion angles systematically rotated to generate an approximate free energy landscape of glucuronic acid.	127
C.1	The product structure resulting from the $C_6 - O_1$ bond breaking pathway. Carbon, oxygen and hydrogen are colored in gray, red and white, respectively. The approximate location of the positive charge is highlighted in a blue circle and the hydrogen bonds are depicted with black dashed lines.	129
C.2	The product structure resulting from the $C_2 - OH$ bond breaking pathway. Carbon, oxygen and hydrogen are colored in gray, red and white, respectively. The approximate location of the positive charge is highlighted in a blue circle and the hydrogen bonds are depicted with black dashed lines.	130
C.3	The product structure resulting from the $C_3 - OH$ bond breaking pathway. Carbon, oxygen and hydrogen are colored in gray, red and white, respectively. The approximate location of the positive charge is highlighted in a blue circle and the hydrogen bonds are depicted with black dashed lines.	130
D.1	The relative electronic energy as a function of distance between the carbon 1 location on xylose and the oxygen at the carbon 4 location on lignin (reaction coordinate) with the energy of the complex separated by 2.97 Angstroms taken as reference. Energies reported are at a M06-2x/6-311++G(d,p) level of theory	132
D.2	The free energy barriers for the deconstruction pathways of a phenyl glycoside (PG) LCC linkage as a function of temperature. The circles denote the computationally computed activation barriers at 298K, while the lines are calculated assuming the enthalpic and entropic contributions are constant with the values computed at 298K.	133
D.3	The reaction free energies for the deconstruction pathways of a PG LCC linkage as a function of temperature. The circles denote the computationally computed reaction energies at 298K, while the lines are calculated assuming the enthalpic and entropic contributions are constant with the values computed at 298K.	133

List of Tables

1.1	Selected major government funded programs on biomass to biofuel research in Canada. Some of the recent funding (for example, low carbon economy fund – 2 billion, strategic innovation fund) are for clean energy and the specific share for biofuels is not clearly stated.	2
1.2	The relative weight fractions of the three main biopolymers in various biomass material. Adapted from McKendry [1].	4
2.1	Reaction free energies (kJ/mol), in the gas phase and the implicit solvent, of the nucleophiles that are capable of forming a benzyl ether linkage at the primary and secondary alcohol locations.	38
B.1	The calculated activation energy, rate constant as well as reaction half-life for the nucleophilic addition to the QM intermediate of the most thermodynamically favorable reaction sites on the model compounds considered. Unbracketed numbers indicate gas phase calculations and bracketed numbers indicate implicit solvent calculations.	128

Abbreviations

BE	Benzyl Ether
CA	Coniferyl Alcohol
DHP	Dehydrogenation Polymer
DFT	Density Functional Theory
DMSO	Dimethyl Sulfoxide
FT-IR	Fourier-Transform Infrared Spectroscopy
FES	Free Energy Surface
GVL	γ -Valerolactone
GC	Gas Chromatography
G	Guaiacyl
HMQC	Heteronuclear Multiple Quantum Coherence
HSQC	Heteronuclear Single Quantum Coherence
HMF	Hydroxymethylfurfural
H	Hydroxyphenyl
IRC	Intrinsic Reaction Coordinate
LCC	Lignin-Carbohydrate Complex
LCB	Lignocellulosic Biomass
LUMO	Lowest Unoccupied Molecular Orbital
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
OP	Organosolv Pretreatment
PG	Phenyl Glycoside
QM	Quinone Methide
SA	Sinapyl Alcohol
S	Syringyl
TS	Transition State

Chapter 1

Introduction

1.1 Motivation

The growing concern for CO₂ emissions compounded with the decreasing fossil reserves has been a driving force towards establishing a clean and renewable carbon source that is capable of alleviating the reliance on petroleum feedstock [1]. On the forefront of meeting these requirements is second generation lignocellulosic biomass (LCB), which is abundantly produced in nature and has a carbon neutral (potentially negative) effect on global CO₂ concentrations [1, 2, 3, 4]. Compared to the many other renewable energy sources such as wind, solar, geothermal, *etc.* only biomass has the advantage of offering an alternative synthesis route for commodity and specialty chemicals and materials. As such, biomass is the only renewable energy source that is also an alternative source of carbon for valorization [5]. From the viewpoint of availability of LCB within the Canadian context, the country produces ~80 million dry-tonnes of agriculture residues per year [6], and presently the majority is used as an inefficient source of energy, either by direct burning or by converting it into fuel, and not as a source of renewable carbon for the chemical industry, completely discounting its economic value. Despite more than \$2 billion spent over the last 10 years on biomass-to-biofuels research in Canada (*cf.* Table 1.1), very little progress has been made beyond bioethanol and bio-diesel. Converting biomass to fuels and fuel additives using current technology is not market competitive due to the low and highly fluctuating costs of their fossil derived counterparts. Though biological transformations are selective towards potential high-value chemicals, the cost of the biological agents and slow turnover affect the economics adversely. Technology with proven commercial success have struggled to remain cost competitive. For example, the cost of succinic acid is over \$2500 per ton when derived from biomass sources, whereas it falls to less than \$2000 per ton from petroleum sources. This resulted in the 10 000 metric ton capacity bio-based succinic acid plant by BASF-carbon-Puran inoperative since 2019 [7]. Similarly, bio-Amber, a bio-based succinic acid plant, declared bankruptcy in

Table 1.1: Selected major government funded programs on biomass to biofuel research in Canada. Some of the recent funding (for example, low carbon economy fund – 2 billion, strategic innovation fund) are for clean energy and the specific share for biofuels is not clearly stated.

Project/Program Title	Year	Funding (\$)
ecoENERGY for Biofuels Program [9]	2008-2017	1.5 B
Next-generation Biofuels Fund [10]	2012-2022	500 M
Alberta Bioenergy Producer Credit Program [11]	2007-2014	150 M
Bioenergy Producer Program [12]	2017-2020	63 M
Biomass Cluster led by the BioFuelNet Canada Network [13]	2019	10.1 M
Woodland Biofuels [14]	2019	4.7 M
Production and Conversion of Biorefinery Cellulose to Advanced Fuels, Biochemicals and Biomaterials [15]	2012-2017	1.1 M
Lignin-to-Drop-In BioJetfuels and Chemicals [16]	2012-2018	1 M
Biomass-rich Waste Conversion into Drop-in Fuels [17]	2012-2017	0.73 M

2018 [8]. Hence, developing and diversifying biomass deconstruction and conversion efforts, via inorganic pathways, to target specialty chemicals as well as materials that either cannot be obtained or are relatively difficult/expensive to synthesize from petroleum sources is essential to improve the economic viability of bio-based products and sway the paradigm away from petroculture.

Presently, a wide variety of deconstruction techniques are implemented, such as steam explosion, liquid hot water, ammonia fibre/freeze explosion, alkali or acidic based processes, *etc.*; [18, 19, 20] however, these methodologies employ severe operating conditions with elevated temperatures/pressures and/or concentrated chemicals, that partially destroy lignin and hemicellulose, leaving cellulose as the target component for valorization. Despite cellulose being the primary component in biomass, it still comprises less than half of the viable carbon ¹ that can be obtained in the majority of feedstocks [1]. Thus, in many cases neglecting the economic value of hemicellulose and lignin, results in over 50% of the carbon

¹Taking standard composition ranges of feedstocks [1] and assuming the carbon present in cellulose, hemicellulose and lignin is a result of being comprised entirely of glucose, xylose and coniferyl alcohol, respectively

being inefficiently utilized, reducing revenue by over five times [21]. An additional implication of the current conversion techniques is that the severe operating conditions means the technology cannot be decentralized safely. Since biomass energy content per hectare is low (the lower heating value is ~ 17 MJ/kg) [1], agriculture residues need transporting to a centralized refinery for treatment, costing \$60–90 per ton for a distance of 100 km [22], with absolutely no benefit for the local agricultural and rural communities [23].

Without understanding the biomass microstructure and the physico-chemical interactions that govern the biomass recalcitrance, resorting to severe conditions in a trial-and-error fashion to combat such recalcitrance is inevitable [23, 24, 25]. However, if the biomass microstructure is understood and the phenomena dictating the recalcitrance adequately quantified, this would allow a methodical approach in rationally designing less severe deconstruction methods. Such developments would circumvent the poor utilization of the feedstock, preserving the viable carbon that can be obtained and chemically transformed to desired products as well as allow a level of decentralization to be implemented. “At-source” biomass fractionation is crucial and likely the only way to offer social benefits to the rural communities that produce it, while simultaneously, making transport economically feasible. Additionally, processing under milder conditions would also allow control over the deconstruction, thereby providing a basis to tailor the deconstruction methodology depending on the desired products. This in turn ensures an efficient synthesis route is incorporated and unnecessary processing expenses are avoided. Therefore, understanding the biomass microstructure, thus the origin of biomass recalcitrance, is the core impediment in systematically designing deconstruction technology that would revolutionize the synthesis of specialty chemicals and materials derived from biomass, making it a niche, commercially feasible industry (*cf.* Figure 1.1).²

Chapter 1 will discuss the existing knowledge and knowledge-gaps about the microstructure and linkages among constituents in biomass, and will establish the often overlooked link between biomass deconstruction and the biomass structure. With the relevant background and motivation presented, the objectives, scope and organization of the thesis will then be discussed.

²The color scheme of cellulose units as blue, hemicellulose units as orange/red/yellow and lignin units as green will be maintained for Chapters 1-4 of the thesis.

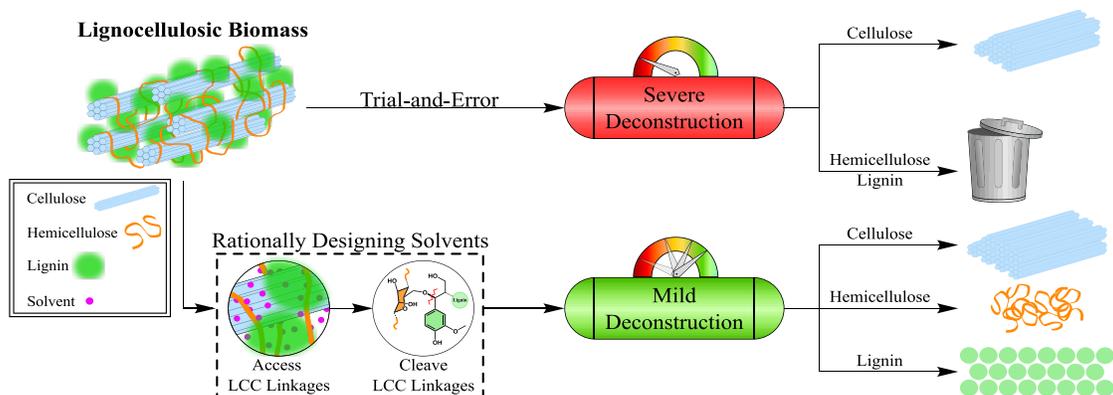


Figure 1.1: Visual depiction summarizing the different strategies to deconstruct lignocellulosic biomass.

1.2 Lignocellulosic Biomass

1.2.1 Composition

LCB is primarily composed of three biopolymers: cellulose, hemicellulose and lignin, with trace amounts of alkali and alkaline earth metals, as well as other components such as proteins [26, 27]. The relative composition of biomass components and the exact polymeric structure (repeating monomer/oligomer units) of each biopolymer largely depends on the source of biomass, *i.e.*, hardwoods (oak, ash, beech) vs. softwoods (pine, fir, spruce) vs. herbaceous plants. The relative composition of the biopolymers is provided in Table 1.2.

Table 1.2: The relative weight fractions of the three main biopolymers in various biomass material. Adapted from McKendry [1].

Biomass Source	Cellulose	Hemicellulose	Lignin
Hardwood	0.45 - 0.50	0.20 - 0.25	0.20 - 0.25
Softwood	0.35 - 0.40	0.25 - 0.30	0.27 - 0.30
Wheat Straw	0.33 - 0.40	0.20 - 0.25	0.15 - 0.20
Switchgrass	0.30 - 0.50	0.10 - 0.40	0.05 - 0.20

Cellulose is the main constituent of the plant wall and is a linear polymer composed of β -D-glucopyranose linked via β -(1,4) glycosidic bonds (Figure 1.2a). Within LCB, cellulose chains pack together to form crystalline polysaccharides that come together through a multitude of favorable interactions that include hydrogen bonding as well as hydrophobic stacking to create rigid microfibrils within the cell wall [28, 29]. Hemicellulose is a branched polymer

within plant cell walls that has a β -(1,4)- linked backbone with an equatorial configuration [30] and is composed of pentoses (xylose/arabinose), hexoses (mannose/ glucose/galactose) and sugar acids [31] (Figure 1.2b). The composition and frequency of linkages as well as the branching characteristics of hemicellulose are all dependent on the biomass source [30, 31], with the vast majority of biomass being sufficiently approximated by being composed of either xylans (*e.g.*, hardwoods) or galactoglucomannan (*e.g.*, softwoods) or a combination of the two [31, 32]. Traditionally, lignin was assumed to resemble a highly branched/network polymer; however, the extent of crosslinking (if any) within the lignin structure has been subject to recent debate [33, 34, 35]. The extent of branching in the lignin structure will be explored further in Chapter 5. Regardless, lignin is synthesized by the polymerization of monolignols. The three dominant monolignols, p-coumaryl, coniferyl (CA) and sinapyl (SA) alcohol, simply differ by the degree of methoxy substitution (Figure 1.2c). The polymerization of the monolignols yield the various lignin composition, hydroxyphenyl (H), guaiacyl (G) and syringyl (S) phenylpropanoid units, that define the macromolecule. Depending on the source of biomass, the lignin composition can vary. For example, hardwoods mostly consist of G and S units with traces of H units, whereas softwoods are almost entirely G units with traces of H units [36]. Although evidence has demonstrated the incorporation of other ‘monomeric’ units into the lignin structure, the current literature is still developing in this regard [36, 37, 38].

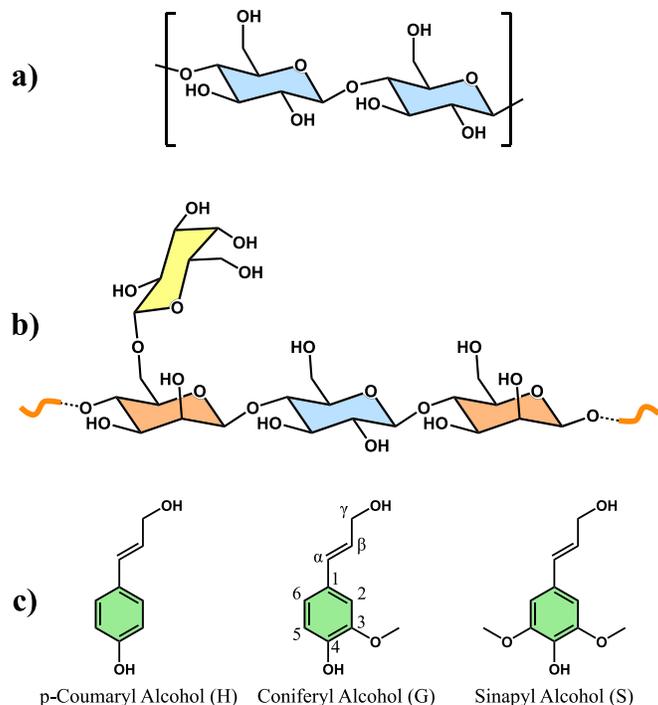


Figure 1.2: a) Cellulose repeat unit. b) An example of a hemicellulose segment from softwood. The structure is not meant to represent a repeating unit, as the polymer shows significant chemical diversity. The orange wavy lines represents the remaining units in the hemicellulose structure. c) The three primary monolignols that polymerize to form lignin in secondary plant cell walls.

1.2.2 Lignin Synthesis

Our understanding of the biomass microstructure is largely qualitative and far from complete. A layer like superposition of cellulose, hemicellulose and lignin [36, 39, 40, 41] has been commonly assumed, where cellulose fibers are covered with adsorbed hemicellulose and then lignin solidifies the carbohydrate matrix. Although this simple model offers a starting point for analysis, the physical and chemical interactions occurring at the interface between the three biopolymers are neglected [42, 43, 44]. Despite cellulose and hemicellulose having fairly well-defined chemical structures that are known with confidence, the primary challenge in elucidating the details of the microstructure stems from lignin, due to its variable structure, organization and convoluted synthesis that leads to covalent linkages between biopolymers, which plays a key role in the overall recalcitrant microstructure.

Lignification of the cell wall can be broken down into three stages: monolignol biosynthesis, transport of monolignols and polymerization of monolignols. The first two stages occur outside the secondary cell wall, therefore, do not directly pertain to the final microstructure that is the topic of the present discussion. Lignin polymerization begins by the enzymatic

oxidation of the monolignols to generate radicals [45], where radical coupling between the oxidized monolignols builds the lignin structure. As a result of the resonance stabilization of the radicals, a distribution of linkages between the lignin units can form [46], β -O-4, β -5, β - β , 5-O-4, 5-5, β -1 and α -O-4 [35] with the β -O-4 linkage comprising the dominant linkage in all of lignin, independent of the biomass source [36, 37]. As the polymerization of monolignols is central to understanding the structure of lignin as well as the organization of the biomass microstructure, extensive work has been done to understand the polymerization process. Previously, it has been suggested that the lignin polymerization process was strictly controlled through proteins [47] and although this proposed model has been readily refuted [48, 49], it remains that the polymerization process is not truly statistically random [37, 50]. In fact, synthetic dehydrogenation polymers (DHPs), *i.e.*, synthetic lignin, when synthesized with no external controls, shows a reduced quantity of β -O-4 linkages when compared to ‘native’ lignin, suggesting additional factors need to be considered. Experimental *in vitro* investigations with DHPs have demonstrated monomer supply composition [51], pH [52, 53, 54], solvent [55], polysaccharide matrix [56, 57], as well as the nature (location, activity, *etc.*) of the oxidation enzyme [58, 59] all play a role in the polymerization. Furthermore, the implication of the confinement effects in monolignol polymerization has also been investigated, *e.g.*, a radical transfer/shuttling mechanism has been proposed as a possible explanation for how large enzymes responsible for the oxidation of monolignols generate the compact structure of the lignified plant cell wall [60, 61].

The polymerization of monolignols has recently begun to be explored through computational investigations as well. *Ab initio* calculations have been used to assess the energetics of radical coupling of the various available pathways. Sangha *et al.* [62] used density functional theory (DFT) to report the reaction enthalpies for the self and cross dimerization of monolignols and found the formation of β -O-4, β -5, β - β linkages were between 5–20 kcal/mol more favorable than formation of 5-O-4, 5-5 and β -1 linkages in gas phase, and up to \sim 15 kcal/mol more favored in an implicit solvent. Additionally, the reaction enthalpies also showed a dependence on the composition of monolignols participating in the reaction. Gani *et al.* [63] used DFT to report the reaction enthalpy and activation enthalpy of radical reactions representative of lignin polymerization and determined radical coupling is indeed kinetically controlled. Dimerization forming the β - β linkage was determined to be the most thermodynamically and kinetically favorable. However, this linkage is sterically inaccessible for chain growth, and the formation of the β -O-4 linkage showed the greatest favorability in growing chains, agreeing with experimental observations reporting the dominant fraction of the β -O-4 linkage. Furthermore, algorithms developed to model monolignol polymerization have shown the monolignol supply rate [64, 65] influences the distribution of linkages formed, in

addition to the composition of the monolignols. Overall, experimental and computational investigations have adequately established that different physiological factors play a role in the polymerization of monolignols. However, a detailed understanding of how these factors influence the formation of the lignin structure in the context of the cellulose/hemicellulose matrix, thus governing the physico-chemical interactions between biopolymers responsible for the refractory nature of biomass, is lacking.

1.2.3 Microstructure: Cellulose, Hemicellulose & Lignin Interactions

1.2.3.1 Physical Interactions

Due to the complexity of the biomass structure, experiments are inherently limited in their capabilities in elucidating molecular level insights. Commonly, changes in macroscopic observables, *e.g.*, product yields from hydrolysis or pyrolysis, are used as evidence to infer about the underlying molecular phenomena. As such, computational tools play a key role in contextualizing experimental results. Despite a hierarchical structure being assumed, where cellulose and lignin do not interact directly, rather both biopolymers interact with hemicellulose, this would require cellulose microfibrils to be completely covered by hemicellulose. However, this is unlikely to be the case, since it would require an organized deposition of hemicellulose, contrary to what is known of the cell wall synthesis as well as what would be anticipated from its branched structure. Therefore, evaluating each combination of interactions, cellulose-hemicellulose, cellulose-lignin and hemicellulose-lignin is imperative.

1.2.3.2 Cellulose & Hemicellulose

The interactions between cellulose and hemicellulose are believed to be purely physical and to the best of my knowledge, no investigation has been reported indicating chemical interactions between these biopolymers. Many computational studies have focused on the physical interactions between cellulose and hemicellulose to quantify the atomistic level interaction details occurring in the cell wall. Using molecular mechanics, early studies [66, 67, 68], provided a starting point for the physical interactions between xyloglucan and cellulose. Hanus and Mazeau [69] investigated the effect the side chain length of xyloglucan had on cellulose adsorption. They found xyloglucan adsorbed non-specifically onto the cellulose surface, with the side chains of xyloglucan influencing the organization, with longer side chains making the adsorption less efficient. Zhang *et al.* [70] investigated the influence of explicit water

and side chain variation. The presence of explicit water decreased the overall interaction energy between the biopolymers and van der Waals interactions dominated over electrostatic interactions in the adsorption. Side chain variation demonstrated negligible alteration to the adsorption energies, however, it influenced the adsorption structure of xyloglucan. Mazeau and Charlier [71] modeled a monolayer of xylan oligomers on a cellulose surface and showed extended conformations were preferred. Additionally, the interaction energy per monomer between xylan and cellulose decreased with the monolayer coverage as a result of the xylan chains interacting with cellulose as well as with each other. Silveira *et al.* [72] investigated the effect hemicellulose composition has on its interaction with cellulose in a solvated environment. Branches of arabinose, glucuronic acid as well as glucuronate strengthen the hemicellulose-cellulose interactions as a result of strongly coordinating to hydrogen bond donor sites on the cellulose surface. The xylan-cellulose organization has been further investigated by incorporating multiple xylan oligomers (DP = 14) on a cellulose face, as well as the role calcium ions play in the biopolymers interactions [73]. It was shown for single chain adsorption that the positioning of the side chain substitution was more important than the chemical nature of the substitution (glucuronic acid and arabinose) for the stability of adsorption, as a result of influencing the flexibility of the xylan chain. Additionally, calcium ions mediated cross-links between xylan chains stabilizing the system. Zhang *et al.* [74] investigated the binding pattern of xylan between two cellulose microfibrils as well as the strength of interactions at the interfaces under shear loadings. The contact area and corresponding hydrogen bonds between cellulose and hemicellulose gave rise to the shear strength of the composite. Furthermore, the presence of explicit water lubricated the cellulose and hemicellulose interface, softening the entire matrix, agreeing with previous reports [70]. Similarly, Khodayari *et al.* [75] modeled five different hemicellulose models on cellulose to characterize the binding strength. Although composition did influence the binding energy, all free energies were within 28 kJ/mol of one another, suggesting hemicellulose composition only plays a moderate role in the interaction energy. Moreover, investigating the influence explicit water has on the biopolymer interactions during shearing modes, again, agreed with previous reports [70, 74]. Kumar *et al.* [76] investigated the interactions of a linear xylan chain (DP = 30) on a cellulose fiber at 303 K as well as 445 K and found that xylan binds stronger and irreversibly at elevated temperatures because of increased hydrogen bonding. The work was completed in conjunction with experimental work to help contextualize the molecular mechanism of cellulose-hemicellulose interactions that was inhibiting enzymatic conversion of cellulose. However, the work directly pertained to re-localized hemicellulose, *i.e.*, hemicellulose re-depositing on cellulose after the biomass structure was destroyed, and not to the recalcitrant biomass microstructure. From a deconstruction perspective, our

atomistic understanding of the cellulose and hemicellulose interactions are unlikely to be substantially improved via further computations. Since such two biopolymer systems have been extensively explored and computational limitations such as cell size and selection of hemicellulose model compounds are inherent in the methodology, new insight becomes increasingly challenging to obtain from these tools alone. Overall, the aforementioned studies show that cellulose and hemicellulose possess favorable interactions because of hydrogen bonding at the interface. However, no strong evidence has been presented that would indicate the physical interactions between cellulose and hemicellulose would give rise to biomass recalcitrance.

1.2.3.3 Cellulose & Lignin

Moreover, computational investigations have modeled lignin adsorption on cellulose using molecular dynamics at various length scales. Houtman and Atalla [77] modeled the adsorption of a lignin monomer (monolignol) and trimer on a cellulose surface in the presence of water and found both molecules adsorbed to the cellulose surface with favorable interactions, regardless of the starting configuration. Additionally, the phenyl rings of lignin molecules showed a preference to orient themselves parallel to the cellulose surface. Shevchenko and Bailey [78] investigated the interaction between individual linear chains of lignin and cellulose. The minimum energy conformers suggested that lignin closely associates with cellulose and could potentially coil around carbohydrates in the cell wall (helical arrangement). Beombes and Mazeau [79] examined a lignin dimer adsorbing on the different surfaces of the cellulose crystal. Favorable adsorption occurred independent of the surface orientation, and the phenyl rings tended to orient themselves parallel to the surface, agreeing with previous investigations [77]. This work was extended [80] to explore the effects of cellulose surface coverage, structural parameters of lignin, as well as explicit water has on the adsorption characteristics. Lignin dimers were used to create a monolayer on the cellulose surface and no lignin aggregation was observed; additionally, the phenyl rings of lignin oriented preferentially parallel to the surface. However, when a bilayer was created, the phenyl ring orientation was lost in the second layer. Furthermore, the presence of explicit water reduced the adsorption interaction energy, but no major changes in the organization or structure of the biopolymers' physical interaction was observed. Finally, when a branched lignin oligomer was created, the branching and types of linkages determined the flexibility of the chain, which influenced whether lignin aggregated on the surface or could form an extended structure to maximize surface contacts. Lindner *et al.* [81] performed large scale molecular dynamics examining the adsorption of 52 lignin polymers, each with a degree of polymerization of 61,

on a cellulose fiber in aqueous solution. It was determined lignin strongly associates with itself as well as cellulose, however, crystalline sections of cellulose demonstrated a greater tendency to associate with lignin than non-crystalline regions of cellulose. Similarly, Vermaas *et al.* [82] used large scale molecular dynamics to create an atlas of lignin-cellulose interactions across different crystalline faces of cellulose and lignin chemistries. The hydrophobic face showed the most effective binding to lignin and increasing the proportion of sinapyl alcohol relative to coniferyl alcohol (referred to as the S/G ratio) decreased the binding affinity between lignin and cellulose. Furthermore, the contact surface area between lignin and cellulose was the primary factor in determining the overall interaction energy. Similar to cellulose interacting with hemicellulose, physical interactions between cellulose and lignin have been quite extensively assessed, with new insights becoming increasingly challenging to obtain. Although, evidence has indicated that physical interactions between cellulose and lignin can be problematic, that is typically in the context of enzymatic access to cellulose [83] and not directly pertaining to the recalcitrant biomass microstructure.

1.2.3.4 Lignin & Hemicellulose

Computational investigations characterizing the physical interaction between lignin and hemicellulose have been relatively scarce, however, the work of Silveira *et al.* [84] have shown that the interactions between the biopolymers are hydrophobic and driven entropically by expelling water from the interaction surfaces.

The extensive computational investigations have demonstrated that cellulose, hemicellulose and lignin have favorable physical interactions. Additionally, cell wall environmental considerations such as solvent effects, confinement effects, multiple polysaccharides, *etc.*, have recently begun to be incorporated, bringing cell wall models closer to reality. However, none of the physical interactions between cellulose, hemicellulose and lignin have demonstrated such favorability that would indicate physical interactions between biopolymers are responsible for the refractory nature of biomass. As such, the origin of biomass recalcitrance is unlikely to arise from physical interactions alone, rather, emerging evidence [85] has indicated that chemical interaction between biopolymers is responsible for such recalcitrance. Such considerations have not been entertained in the mentioned studies and without them, investigations are forced to continue to speculate into biomass recalcitrance and are incapable of providing the framework to guide the design of environment-friendly, mild and potentially decentralized deconstruction technology [42].

1.2.3.5 Chemical Interactions (Lignin Carbohydrate Complexes (LCCs))

Although a preliminary understanding of lignin synthesis exists to date, it has been identified that the lignin synthesis pathway not only creates a lignin polymer, it is also responsible for creating covalent linkages with other biopolymers in the cell wall. The term lignin carbohydrate complexes (LCCs) has been coined to describe the covalent or chemical interactions between lignin and other biopolymers. The difficulty in unequivocally identifying LCC linkages is a result of the severe fractionation techniques employed to isolate them, leaving the possibility that they are extraction artifacts [86]. The first step in isolating ‘native’ lignin and LCCs is a mechanical treatment, typically some form of ball milling, which consequently generates mechano-radicals that increase the carbonyl and phenolic OH groups in lignin [87]; as well as results in depolymerization via cleavage of β -aryl ether bonds [88]. Furthermore, the overall sample yield, structure and quantity of LCC fragments in the sample have been shown to be influenced by milling conditions, *i.e.*, time and supplied energy [89]. As such, it is difficult to ascertain whether the identified LCC fragments were released from the broken carbohydrate matrix or were formed through radical quenching reactions from the ball milling [87, 88]. The subsequent step(s) for extracting and isolating LCCs commonly utilize a series of separations and various solvent treatments with the solvent choice depending on the target component for analysis. Complications also arise in the wet chemistry methods often implemented to selectively degrade LCC fragments since the reaction environment can dramatically change depending on the chemical nature of the degraded components dictating if repolymerization or depolymerization occurs and similarly if the formation or degradation of LCC linkages occurs [87, 90]. Wet chemistry methods can be coupled with non-destructive techniques such as nuclear magnetic resonance (NMR), improving the insight that can be obtained, however, these non-destructive methods are constrained by experimental resolution, signal overlap and require a standard for calibration [91]. This further reinforces that no one LCC isolation techniques allows access to all the details of the LCC structure, composition, quantity, *etc.*, further complicating comparative studies in literature. Although experimental work has laid the foundation for the existence of LCC linkages, developing an unquestionable technique for LCC isolation and analysis is laborious task, therefore, LCCs existence and frequency in native biomass is largely inferred and remains a controversial topic.

Despite LCC linkages being hypothesized in 1866 by Erdmann [92], there are very few investigations that have provided concrete evidence of LCCs. Ralph *et al.* [93] used long-range heteronuclear multiple bond correlation (HMBC) nuclear magnetic resonance (NMR) analysis to provide evidence that ferulates and lignin undergo radical coupling *in vivo* within the cell-wall of ryegrass. Furthermore, the experimental results indicated that the linkages are nearly exclusively occurring between ferulates and lignin monomers, not with lignin

oligomers. This result is used as evidence to indicate that ferulates act as initiation sites for lignin polymerization, which has gathered more evidence in support [94, 95]. Nishimura *et al.* [96] conducted a similar analysis on the wood of Japanese red pine, to provide evidence that lignin was covalently bonded to a mannose residue in glucomannan. This result supports LCC formation through the re-aromatization via nucleophilic addition of hemicellulose to the lignin quinone methide (QM) intermediate. Although these studies shed light on the existence of LCCs linkages, information such as the frequency of these linkages remains debated [44, 96]. However, emerging evidence has indicated that these LCC linkages play a prominent role in biomass recalcitrance [57, 85], making them an imperative consideration from a deconstruction perspective. Therefore, elucidating the mechanistic details of LCC formation could be a key step in pinpointing the origin of biomass recalcitrance and laying the foundation for the design of mild deconstruction technology.

There are primarily five different types of LCC linkages proposed, as depicted in Figure 1.3: benzyl ethers [97, 98], benzyl esters [99], phenyl glycoside (PG) [87, 100], acetal linkages [101] and ferulate linkages [93]. However, not all LCC linkages are equally likely to contribute to biomass recalcitrance. The evidence for the existence of acetal LCC linkages remains extremely scarce [42, 102], indicating such linkages are unlikely to be strong contributors to biomass recalcitrance and are not explored further. Although substantial evidence exists for the existence of PG linkages, their formation mechanism is rarely discussed and remains largely unknown [86, 87, 103]. As a result, PG linkages likely occur in a large enough frequency to contribute to biomass recalcitrance, however, their chemistry requires special attention and further discussion of PG linkages will be expounded upon in Chapter 4. The remaining three LCC linkages, ferulate, benzyl ether and benzyl ester, have enough evidence to suggest they could contribute to biomass recalcitrance as well as enough chemistry identified to be discussed further.

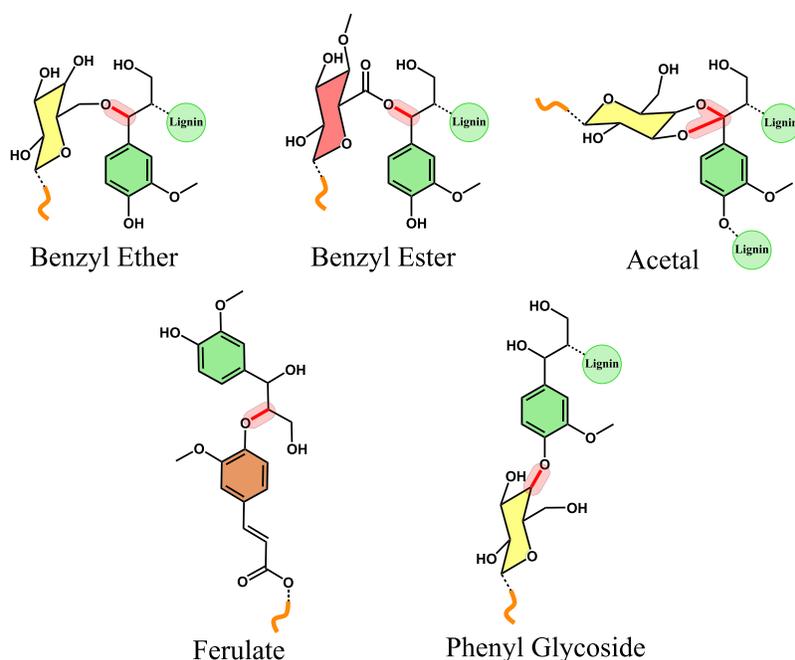


Figure 1.3: The five types of LCC linkages proposed. The LCC linkage is colored red.

The current theories for the formation of the ferulate, benzyl ether and benzyl ester LCC linkages are a result of side reactions occurring during monolignol polymerization in the plant cell wall. Lignin forms by radical coupling polymerization, and radicals are generated through enzymatic oxidation. However, the enzymes that oxidize monolignols possess weak substrate specificity for monolignols and can interact with monolignol-related compounds (ferulates) [95, 104, 105]. Alternatively, radicals could be generated nonenzymatically through a radical transfer/shuttling mechanism on the ferulate compounds. In either case, ferulate compounds are oxidized to generate a radical. The first LCC linkage mechanism suggest that the oxidized ferulate substitutions on hemicellulose participate in lignin polymerization, *i.e.*, LCC linkages form via radical coupling, refer to Figure 1.4 [93]. Although the existence of ferulate linkages are well established, the frequency of such linkages are completely dependent on the ferulate concentration, which is primarily predominant in grasses [36, 32]. Therefore, it is unlikely that ferulate LCC linkages are solely responsible for biomass recalcitrance across the variety of feedstocks. As such, ferulate linkages are not explored further given their strong compositional dependence.

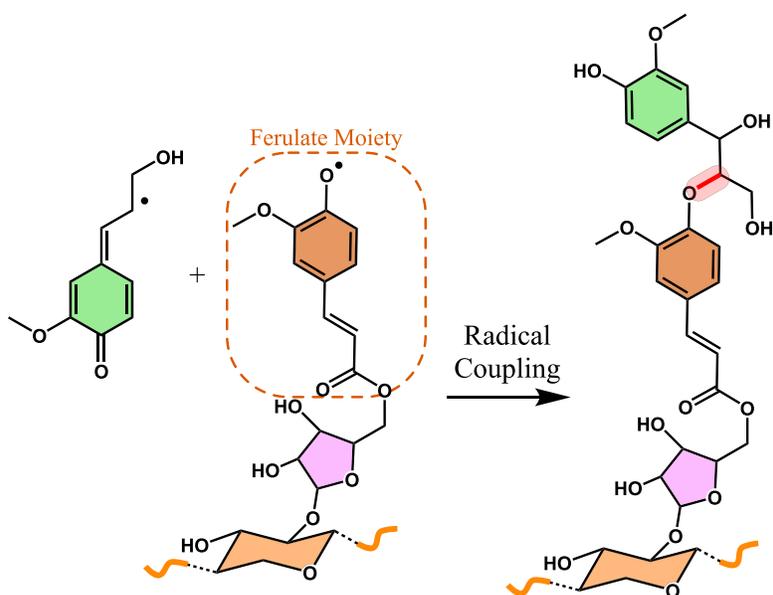


Figure 1.4: The radical coupling reaction between a monolignol and ferulate moiety to form an LCC linkage.

The second mechanism leading to a LCC linkage is a result of the radical coupling to form the dominant β -O-4 linkage in lignin. When two oxidized monolignols couple, one at the β position and one at the 4 position, a QM intermediate is formed. For the β -O-4 linkage formation to complete, the nucleophilic addition at the α position of the monolignol is required for the rearomatization of the QM intermediate. Traditionally the nucleophile was assumed to be water, thereby adding a hydroxyl group at the α position [32]. However, theoretically any compound containing a nucleophile such as a hydroxyl or carboxylic acid group could participate in the reaction. In other words, every hemicellulose component, independent of specific composition, possesses functional groups that could act as a nucleophile. Therefore, this mechanism represents the likely pathway for the formation of benzyl ether as well as benzyl ester LCC linkages, refer to Figure 1.5 for mechanistic details.

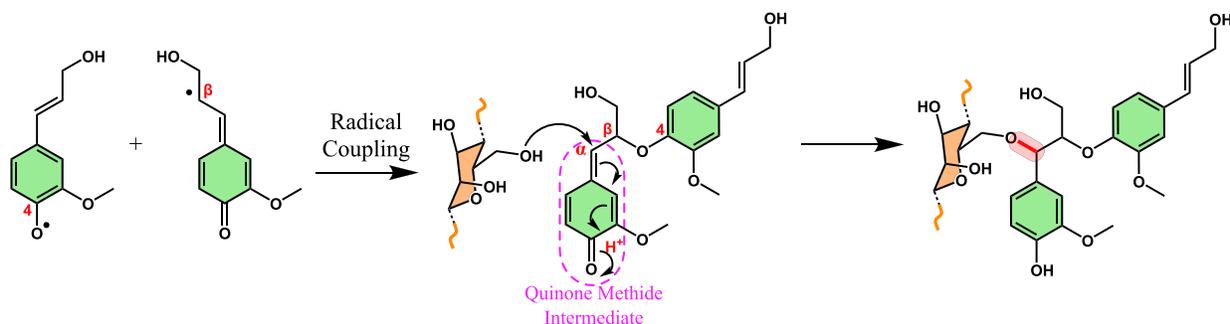


Figure 1.5: The radical coupling reaction between two monolignols forms a lignin dimer, where the lignin intermediate forms a quinone methide (QM) structure that reacts with cellulose or hemicellulose to form a LCC linkage. A benzyl ether linkage is depicted.

Computational methods offer an efficient tool to complement and accelerate our current understanding of the role LCC linkages play in the cell wall. As mentioned, many computational investigations have analyzed the physical interactions between biopolymers; however, limited studies have considered the LCC linkages (chemical interactions) between biopolymers.

One of the first computational investigations for LCC linkages was performed by Remko [106], where a semi-empirical quantum chemical method [107] was used to study model compounds for the conformations of benzyl ether and ester LCC linkages. The work of Durbeej and Eriksson [108] used DFT to determine the thermodynamics and kinetics to form a β -O-4 linkage. The initial radical coupling of monolignols was determined to be thermodynamically as well as kinetically feasible and showed great agreement with more recent investigations that rigorously demonstrated the favorable thermochemistry and facile kinetics of radical coupling of monolignols [63]. Although Durbeej and Eriksson [108] did not directly investigate LCC linkages they modeled the reaction mechanism for the nucleophilic addition of water to the QM intermediate with and without an acid catalyst. It was determined that the energy barrier for the dissociation of water, without an acid catalyst, was insurmountable under physiological conditions. This suggested an acid catalyst is necessary and present when monolignol polymerization occurs. When the mechanism was investigated using three water molecules and an excess proton, the energy barrier for the proton addition was found to be insignificant and the subsequent dissociation of water became kinetically facile. As a result, the study provided a great starting point for investigating the nucleophilic additions to the QM intermediate in lignification. Zhang *et al.* [109] used electronic structure calculations, semiempirical, Hartree-Fock and DFT, to benchmark appropriate levels of theory with geometry characteristics and investigate the hydrogen bonding between cellulose and hemicellulose as well as the LCC linkages between hemicellulose and lignin. However,

the computational methodology limits the utility of the results outside the studies objectives since conformer effects were not considered in the molecules and the B3LYP functional [110, 111, 112] was the most sophisticated level of theory used. Charlier and Mazeau [113] modeled each biopolymer, *i.e.*, cellulose, hemicellulose and lignin, with water in one system. The system composed consisted of two crystalline cellulose surfaces, followed by two layers of xylan (hemicellulose), a middle section of lignin, that varied in composition (G, S, H and coumaryl units) as well as structure (degree of polymerization, linkage types and branching), and lastly by the insertion of water molecules in the voids. A ferulate moiety was arbitrarily added to the xylan, and the various lignin structure generated were covalently linked to xylan at this point to model a LCC linkage. A well-defined interface occurred between cellulose and xylan, whereas xylan interpenetrated lignin, demonstrating favorable physical interactions. Each lignin structured generated showed amorphous characteristics with variable density profiles dependent on the model. The majority of the studies that have incorporated LCC linkages have simply included them as given by experimental work, with limited insight into the linkages themselves.

1.3 Deconstruction of Lignocellulosic Biomass

Biomass deconstruction or fractionation refers to the separation of cellulose, hemicellulose and lignin. Whereas the phrase “pretreatment” can be described as the first step in biomass processing to alter the chemical and/or physical structure of biomass to reduce the refractory nature and improve the effectiveness of further processing [18, 19]. Therefore, pretreatment does not necessarily mean the biopolymers are separated/isolated. However, for the present discussion the intent is to discuss pretreatment as a deconstruction technology implemented prior to any biomass transformation intended for valorization of the components, as is commonly implied throughout literature. Note, typically there is no biomass pretreatment for thermochemical conversion technologies such as torrefaction, pyrolysis, gasification, *etc.*, as such, these technologies are excluded from the current discussion. Therefore, as pretreatment is the first step or series of steps in biomass processing it is the necessary starting point to develop a mild, decentralized deconstruction technology with carbon efficient separation, to yield the constituents available for valorization. An ideal pretreatment step should meet the following technical criteria.

1. The pretreatment should separate and isolate individual biomass components (cellulose, hemicellulose and lignin) while preserving their chemical/structural architecture

so as to maximize the viable carbon that can be utilized.

2. The pretreatment step should be able to be decentralized and it should possess a minimal severity factor *i.e.*, avoid using high temperature or pressure and highly concentrated, difficult to separate chemicals.
3. The pretreatment should be effective for a range of chemical architectures present in the various biomass feedstocks.

As previously mentioned, a variety of pretreatment methodologies exist and are briefly outlined. Steam pretreatment (SP) or also known as steam explosion takes physically degraded biomass (chipped, ground, etc.) and treats it with saturated steam at temperatures ranging from 433–523 K and pressures ranging from 0.7–4.8 MPa [18, 19, 114]. During this treatment, cellulose fibrils coalesce, hemicellulose undergoes hydrolysis and lignin partially phase separates from hemicellulose [115]. Acids released from biomass via SP mediate the hydrolysis of hemicellulose [18, 114]. Liquid hot water pretreatment (LHW) is similar to SP, however, it uses pressurized liquid water at elevated temperatures ranging between 413–533 K and uses a variety of reactor configurations, batch, semi-batch, continuous and semi-continuous [18, 19, 116]. Ammonia fibre/ freeze explosion (AFEX) uses liquid ammonia at moderate temperatures of 333–393 K, and pressures above 2–3 MPa [18, 117, 118]. The biomass and ammonia are brought into contact in a pressurized vessel and heated to the desired temperature. After several minutes, the pressure is released rapidly, resulting in flash vaporization of ammonia and a drop in the system temperature. Treatment results in a solid state phase transition and coalescence of cellulose fibrils, as well as changes in the lignin structure that results in its aggregation; thus improving the accessibility to cellulose for further treatment [18, 19, 115]. Dilute acid pretreatment (DA) brings the feedstock in contact with a dilute acid (concentration less than ~ 4 wt%) in a reactor vessel and heats the system to temperatures ranging between 413–493 K [18, 19, 119]. The system can be heated directly using steam as in SP or indirectly through the reactor walls. Hydrochloric acid, nitric acid, phosphoric acid as well as sulfuric acid have all been used; however, sulfuric acid is commonly adopted [18]. The effect that DA has on the native biomass structure is similar to that of SP; however, more hemicellulose is lost as a result of acid catalyzed hydrolysis [115]. These aforementioned methodologies employ severe operating conditions and are not versatile enough to deal with the various chemical architectures present in different types of biomass feedstocks. As such, these methods result in significant carbon loss from the original feedstock, largely hampering the atom economy.

Organic solvents or organosolv pretreatment (OP) demonstrates the greatest potential for the efficient deconstruction of the biomass components for a range of compositions at modest operating conditions [120, 121]. OP utilizes a variety of organic solvents (methanol, ethanol, acetone, ethylene glycol, triethylene glycol, or tetrahydrofurfuryl alcohol) or aqueous-organic solvent mixtures (tetrahydrofuran or γ -valerolactone) [122, 123] with or without a catalyst such as inorganic acids (hydrochloric or sulfuric acid) or organic acids (oxalic, acetylsalicylic, or salicylic acid) [124]. Typically, the solvent is selected to maximize the solubility of a biopolymer, *i.e.*, operate as a theta or good solvent. Treatment is performed at temperatures between 373–523 K, with a catalyst unnecessary at temperatures above 458 K, since released organic acids from biomass will self-catalyze the deconstruction [18, 120]. Although solvents are drained, recovered and recycled to reduce the operating costs of the process, the cost of chemicals/catalysts plus the additional safety features that must be considered due to the inherent fire and explosion hazard, can still make it more expensive than other techniques. Despite the inherent advantages of OP, it is still in its infancy of technological development [121, 125]. To help push OP closer to commercialization, the optimization of deconstruction conditions, such as temperature, concentrations, and catalysts, that can be used for varying biomass feedstocks and loadings is crucial [123, 125, 126, 127, 128]. Central to achieving this level optimization is understanding the interplay between physical and chemical interactions occurring between the solvent and the biomass microstructure during the pretreatment process [24].

Currently, experimental investigations are forced to proceed without this knowledge which requires relying on a trial-and-error approach of using different operating conditions, feedstocks, catalysts, reactor designs and solvents while monitoring product yields to evaluate the efficacy of the processing [18, 19, 121]. Furthermore, with almost every possible pretreatment variable altered between investigations including the reported results, *e.g.*, whether cellulose, hemicellulose or lignin yield/degradation is reported, or alternatively if the results are reported in terms of enzymatic hydrolysis efficiency [121] (and references therein), direct comparison of studies is near impossible. At best, the same feedstock is selected with only a few changes in operating conditions, *i.e.*, concentration of solvent, choice of catalyst, temperature and treatment time. However, even in such a case, crude extrapolation for fundamental insights is inevitable [121]. As a result, experimental outcomes are limited to identifying candidate solvents and catalysts with limited efficacy, while speculating on the underlying molecular phenomena responsible for the observed results [120, 121]. In some circumstances, pretreatment conditions can be optimized for a particular solvent and feedstock combination, however, such strategies are resource intensive and fail to establish any foundational knowledge that would allow a systematic approach in improving OP technologies.

Rationally designing solvent technologies to overcome biomass recalcitrance and deconstruct biomass into its desired constituents requires a molecular understanding of the interactions between cellulose, hemicellulose and lignin that give rise to the refractory nature. Such crucial molecular insights have remained elusive from the experimental approach. Therefore, rigorous computational investigations are required to compliment the current experimental literature.

The majority of computational studies characterizing organic solvent interactions have focused primarily on the interactions occurring within an individual biopolymer, *i.e.*, just cellulose [129, 130], hemicellulose [131], or lignin [132]. However, such a simplification entirely neglects the interaction details between the biopolymers that give rise to biomass recalcitrance in the first place. Investigations between multiple biopolymers interacting in an organic solvent are relatively scarce. Patri *et al.* [83] used molecular dynamics to investigate the influence tetrahydrofuran (THF)–water mixture had on lignin–cellulose interactions. The results demonstrated the THF–water mixture was an excellent theta solvent for lignin, such that, lignin separates from the cellulose surface and formed a random coil conformation. However, all the aforementioned computational studies [83, 129, 130, 131, 132] have focused on the physical interactions between solvent and biopolymer(s), following the mainstream approach of characterizing the relative solubility of the biopolymers in different environments; with chemical interactions between the solvent and biomass microstructure not being considered [24]. It remains unlikely that if the central issue was simply characterizing physical interactions and identifying solvents that maximize the solubility of the biopolymers, the refractory nature of biomass would have withstood such intense global scrutiny. Therefore, quantifying the physical and chemical interaction between solvent and biomass is a crucial step in developing an efficient deconstruction technology; however, this necessitates the physico-chemical interaction details among cellulose, hemicellulose and lignin, giving rise to the refractory nature of biomass to be understood first.

1.4 Summary of Chapter 1

As a result of growing environmental concerns, utilization of second-generation lignocellulosic biomass as a sustainable alternative to the petrochemical industry has gained tremendous momentum. However, a key impediment in realizing the full economic potential is developing deconstruction technology to recover all the carbon in biomass. Since the difficulty in the efficient recovery of carbon from biomass lies in its recalcitrant nature, identifying the origin of biomass recalcitrance is a crucial first step and the primary objective of the

thesis.

Existing literature has provided a foundation for describing physical interactions among biopolymers; however, chemical interactions (LCC linkages) between lignin and cellulose / hemicellulose is an active topic that requires further research. Lignin polymerization remains central in quantifying LCC formation. However, emphasis on understanding lignin polymerization needs to be done with cell wall conditions implemented, specifically lignin polymerization at the interface of cellulose/hemicellulose. Such models are necessary to characterize the various mechanisms as well as the associated energetics that lead to different types of LCC linkages. Such outcomes will help identify which linkages are primarily contributing to biomass recalcitrance, due to their abundance and/or their chemical strength.

With the recalcitrant LCC linkages identified, designing and optimizing solvent/catalyst systems to deconstruct the linkages becomes a systematic process. Existing literature characterizing the relative solubility of the biopolymers provides a starting point for optimizing solvent conditions that result in the physical dissociation of the biopolymers. Therefore, solvent technology can be further developed to deconstruct LCC linkages (chemical interactions) and dissociate the biopolymers (physical interactions). Using such knowledge provides a foundation for identifying the solvent/catalyst characteristics or descriptors; whether from the macroscopic properties, *e.g.*, density, viscosity, shape, polarity, dielectric constant, or from molecular level interactions. Molecular level interactions include influencing reactant interaction/adsorption with a catalyst [133], microscopic heterogeneities in solvent (changes in local reaction environment) [134], relative stabilization of the reactant [135] and transition state [136], solvent dynamics and non-equilibrium solvation (reorientation speed of solvent as reaction proceeds) [137, 138], entropic effects due to confinement (promotes association of compounds) [139], direct participation in reaction mechanism [140], generation of new/more active catalyst species [141], and preferential physical solvation (of specific functional groups) [142].

1.5 Objectives & Scope of the Thesis

Majority of the previous investigations pertaining to biomass recalcitrance have attempted to take macroscopic observations and extrapolate for fundamental insight. However, such an approach is futile when attempting to characterize the molecular level biopolymer interactions responsible for biomass recalcitrance. Therefore, the entire thesis employs electron structure calculations, *i.e.*, DFT, to elucidate the molecular level insight necessary to quantify the specific interactions responsible for biomass recalcitrance.

A significant hurdle in effectively separating the biopolymers within biomass is understanding the recalcitrant interactions that need to be targeted during deconstruction. Therefore, the primary objective of the thesis is to identify the origin of biomass recalcitrance. In other words, answer the following question:

What role do LCC linkages play in biomass recalcitrance?

This is achieved by characterizing the various chemical interactions, *i.e.*, LCCs, between lignin and cellulose/hemicellulose. Considering previous experimental evidence to assess which LCC linkage could occur in a large enough frequency across the diverse variety of biomass feedstocks, there are three types of LCC linkages that show potential to strongly contribute to biomass recalcitrance: benzyl ether, benzyl ester and PG. Therefore, the thermodynamics and kinetics of forming each of these LCC linkages is quantified, and compared to competing reactions, to ascertain which linkages are energetically favorable to form in the native biomass structure. After the energetically favored LCC linkages are identified, the reaction mechanisms, kinetics and thermodynamics associated with the deconstruction pathways of these linkages is quantified to identify the deconstruction chemistry that is predominately contributing to biomass recalcitrance.

The solvation of lignin, including the accessibility of reaction sites during biomass deconstruction is largely determined by the lignin structure. Therefore, the secondary objective of the thesis is to determine whether branching occurs in the lignin structure. In other words, answer the following question:

Does branching occur in the lignin structure?

Although previous literature provides strong evidence to suggest that branching does occur in the lignin structure, the main proponents of a linear lignin structure are because many of the previously assumed branching points in lignin, are in fact, terminal units. As such, if the lignin structure is indeed branched, an unrecognized branching point must exist in lignin. A novel reaction pathway that could lead to branching in the lignin structure is identified and quantified. As mentioned, DFT calculations are used to compute the reaction energetics and deconvolute the mechanistic details associated with the formation and deconstruction of LCC linkages as well as the novel reaction pathway that could lead to branching in the lignin structure.

1.6 Organization of the Thesis

This thesis is comprised of six chapters. Chapter 1 serves to provide a thorough introduction to the thesis, *i.e.*, provide a literature review of the work completed in the relevant fields, establish the motivation of the research projects completed as well as outline the content and organization of the remaining chapters. Chapters 2-5 are composed of the original works completed and are organized as follows. Each Chapter begins with an introduction to the relevant literature and is followed by a computational methodology section to describe the model systems as well as the computational procedure implemented to obtain the results of the respective Chapter. After the computational methodology the results and discussion are presented to provide an analysis of the data, highlight the key insights and discuss the implications of the work within the context of the current literature. Finally, each chapter ends with a conclusion to summarize the work completed and key insights of the respective chapter. Chapter 6 provides a conclusion to the thesis.

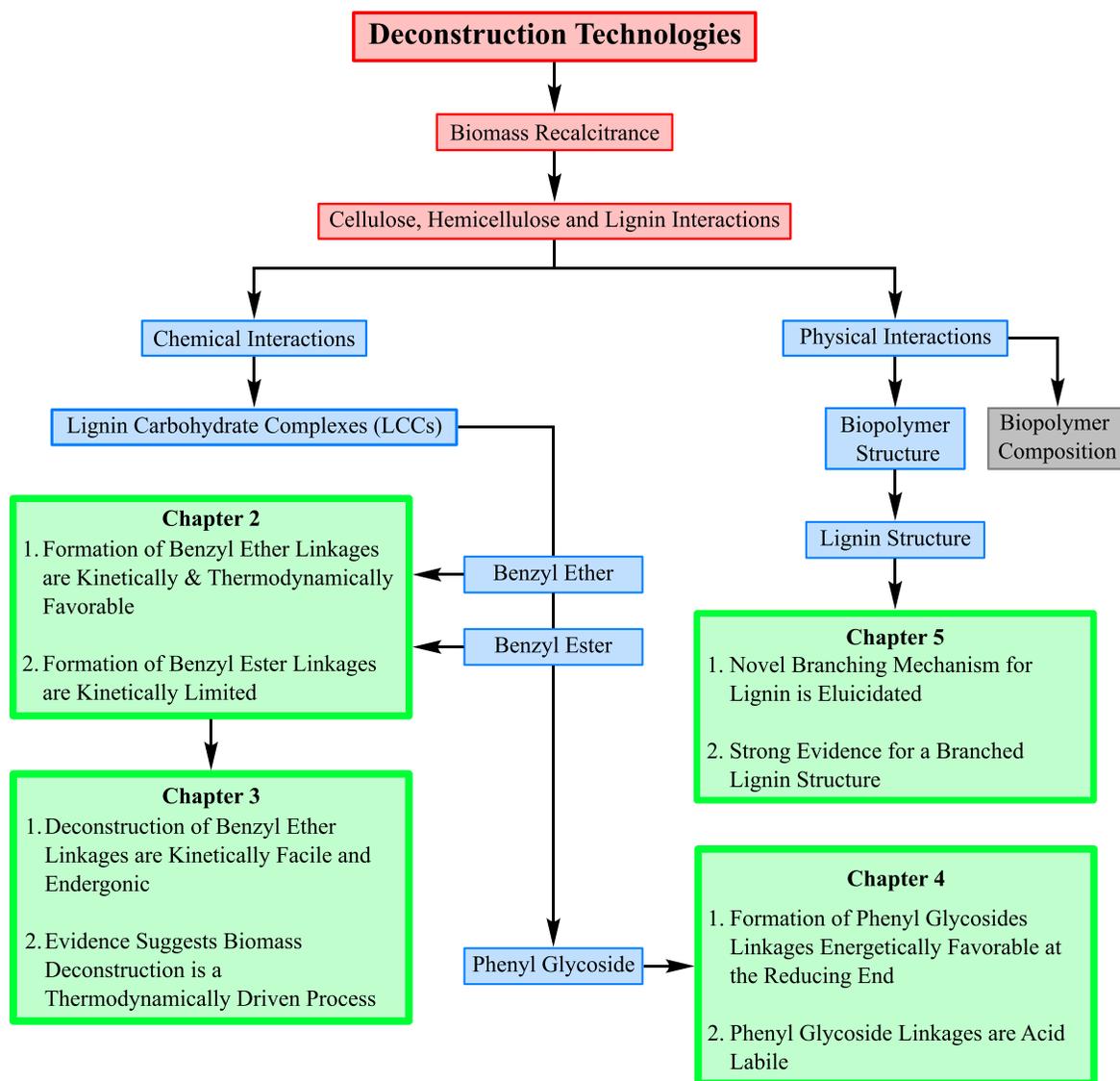


Figure 1.6: An overview of the thesis organization.

1.6.1 Chapter 2: Formation of Benzyl Ether LCC Linkages

The thermodynamics and kinetics of forming benzyl ether and benzyl ester linkages via nucleophilic addition to the QM intermediate was quantified. The formation of these LCC linkages showed thermodynamic favorability, while remaining kinetically facile, compared to the traditionally assumed mechanism of the QM intermediate reacting with water. However, formation of the benzyl ester linkage is kinetically less competitive compared to the formation of benzyl ether linkages, suggesting the benzyl ester linkages are not abundant *in vivo*. The favorable reaction energetics of benzyl ether linkages across the variety of hemicellulose compositions suggest that these LCC linkages are likely abundant. Additionally, from a

deconstruction perspective, the reverse reactions of the linkages require overcoming a positive reaction free energy and higher activation free energy barrier, providing evidence that benzyl ether LCC linkages could contribute strongly to biomass recalcitrance.

1.6.2 Chapter 3: Deconstruction of Benzyl Ether LCC Linkages

The reaction mechanisms, kinetics and thermodynamics associated with the deconstruction of the benzyl ether LCC linkages in biomass under acidic conditions is quantified. The possible reaction pathways identified include degrading the lignin structure, degrading the hemicellulose structure, or cleaving the LCC linkage. Cleaving the LCC linkages demonstrated the lowest activation barriers, however, possessed positive reaction energies. Comparatively, the degradation of hemicellulose possessed higher activation barriers and the greatest thermodynamic feasibility, with negative reaction energies. Computing the reaction energetics as a function of temperature suggested that increasing the temperature during the deconstruction of lignocellulosic biomass would result in hemicellulose degradation before favorable reaction free energies would be established for cleaving LCC linkages. As a result, cleaving benzyl ether LCC linkages to obtain lignin and carbohydrates in their chemically intact form is a thermodynamically controlled process.

1.6.3 Chapter 4: Formation & Deconstruction of Phenyl Glycoside LCC Linkages

The reaction mechanisms, kinetics and thermodynamics associated with the formation and deconstruction of the PG LCC linkage is quantified. The two previously proposed mechanisms, hemiacetal formation and transglycosylation, were associated with a significant activation barrier, suggesting these pathways are kinetically limited. A new mechanism is proposed, the electrophilic addition of hemicellulose to a lignin QM intermediate, that possesses facile kinetics and is exergonic, suggesting it could be the pathway responsible for the significant fraction of PG linkages observed. Moreover, PG formation showed a composition dependence suggesting that xylans will have higher fractions of PG linkages compared to mannans, explaining why softwoods and hardwoods have different reported LCCs. Additionally, the deconstruction energetics demonstrate that breaking PG linkages is kinetically and thermodynamically favored in acid catalyzed deconstruction, indicating that PG linkages are unlikely to be a LCC linkage significantly contributing to biomass recalcitrance

1.6.4 Chapter 5: Branching in the Lignin Structure

The re-aromatization of the QM intermediate via nucleophilic addition of monolignols is explored. The nucleophilic addition of the γ hydroxyl group, forming a α -O- γ lignin linkage, as well as the 4 position hydroxyl group, forming a α -O-4 lignin linkage is quantified. Of the available nucleophiles to re-aromatize the QM intermediate (water, cellulose/hemicellulose, monolignols), the thermodynamically and kinetically favored pathway is the nucleophilic addition of monolignols at the γ hydroxyl group, forming a α -O- γ lignin linkage. The α -O- γ linkage has the added implication that it acts as a branching point in the lignin structure, providing supporting evidence for the branched nature of lignin.

Chapter 2

Formation of Benzyl Ether LCC Linkages

2.1 Introduction

Lignin polymerization begins by the enzymatic oxidation of monolignols to generate radicals [45], where radical coupling between the oxidized monolignols builds the lignin structure. Although a distribution of lignin linkages are formed, the β -O-4 linkage is the dominant linkage in all of lignin, regardless of the biomass source [36, 37]. When two oxidized monolignols combine, one at the β position and one at the 4 position, a quinone methide (QM) intermediate is formed. To complete the β -O-4 linkage formation, the re-aromatization of the QM intermediate at the α position of the monolignol via nucleophilic addition is required. Typically, the nucleophile is assumed to be water, which adds a hydroxyl group at the α position and leads to physical interactions between biopolymers [32]. However, theoretically, any compound containing a hydroxyl or carboxylic acid group could act as a nucleophile; therefore, this pathway represents the accepted mechanism for the formation of benzyl ether and benzyl ester LCC linkages.

Since experimental techniques can be cumbersome in uncovering the molecular level details of the interactions between biopolymers, computational methods can be used to complement and accelerate our current understanding. Previous computational studies have arbitrarily incorporated LCC linkages, while investigating the biomass microstructure [113, 106, 109], without any investigation into the chemistry of their formation. Understanding the pathway and energetics associated with LCC linkage formation helps quantify the interface between biopolymers and provides the foundation to better understand how the convoluted lignin macrostructure is synthesized in the context of the secondary cell walls. The present Chapter evaluates the thermodynamic and kinetic feasibility of the formation of benzyl ether and

benzyl ester LCC linkages. The relative reaction energetics and kinetics of benzyl ether and benzyl ester formation allows their relevance and frequency to be assessed, providing a starting point for pinpointing biomass recalcitrance.

2.2 Computational Methodology

2.2.1 Model Compounds

The monomer forms of cellulose, hemicellulose and lignin are used as model compounds for LCC linkage formation. Specifically, a guaiacyl monolignol reacts with mannose, glucose, galactose, arabinose, and xylose to model benzyl ether linkages and a guaiacyl monolignol reacts with glucuronic acid to model benzyl ester linkages.

2.2.2 Computational Procedure

To ensure that the reaction free energies of LCC linkage formation are not skewed by conformer effects, an efficient technique must be employed to sample the conformational space of the molecules considered. As such, Car-Parinello molecular dynamics [143] and metadynamics (CPMD-metadynamics) were used to calculate the free energy surface (FES) as a function of torsion angles to generate and identify sample conformers. Torsion angles were selected to promote the widest sampling of the system that would not typically occur at reasonable computational times with thermal energy alone. All conformers corresponding to observed minima on the FES were selected for subsequent all-electron density functional theory (DFT) optimization. The general workflow of generating and screening conformers is shown in Figure 2.1. For rigorously correct reaction energetics, a Boltzmann population analysis is required to consider the distribution of conformers. However, this is unlikely to change the results reported, as concluded previously with a carbohydrate thermochemical investigation [144]; therefore, it is not considered presently. For the details of the conformational procedure employed for the reactant and product compounds, please refer to Appendix B.

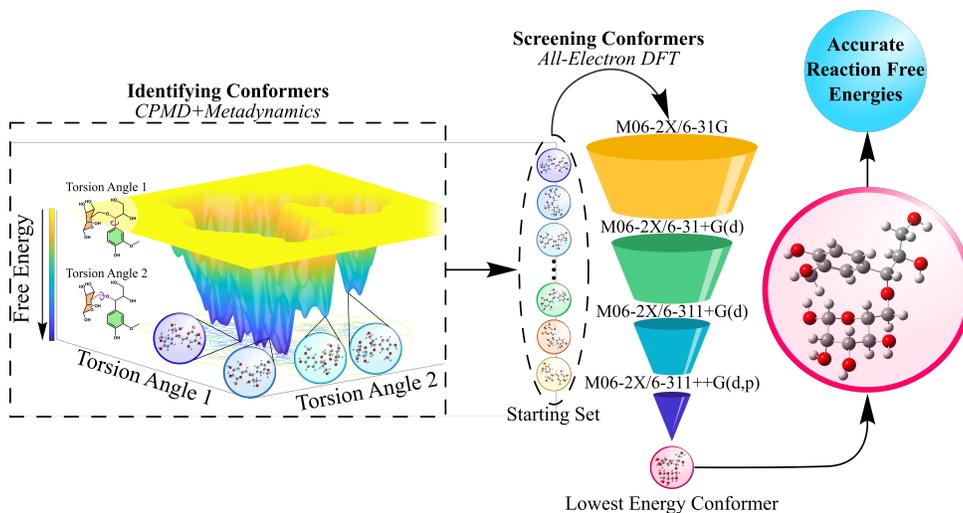


Figure 2.1: General workflow to identify the lowest energy conformers for the model LCC moieties, *i.e.*, the simplified monolignol bonded to the monomer forms of cellulose and hemicellulose.

All-electron DFT calculations were performed using the Gaussian 09 code [145] and were used to compare the relative stabilities and further screen the CPMD-metadynamics identified sample conformers. Due to the large number of starting conformers generated, *e.g.*, many of the model LCC moieties possessed more than 50 conformers, the basis set was chosen to follow a step-wise improvement where the output of a less sophisticated basis set was used as the input for a more sophisticated basis set. This provided a systematic approach for conformational screening, with the lowest energy conformers screened at each level of theory. Simultaneously, comparing how the relative stability of the conformers changed, *i.e.*, the ordering of the lowest energy conformers, and comparing the relative reaction free energies as the level of theory increased, allowed benchmarking at an appropriate level of theory. Every conformer underwent full geometry optimization at each level of theory discussed. No constraints were implemented on the atoms during geometry optimizations. The optimizations were followed by a frequency calculation to provide the desired thermochemical data *i.e.*, free energy, and ensure no spurious frequencies were present in the reactant and product compounds. The hybrid meta M06-2X functional is used at every stage of optimization since it has demonstrated to be sufficiently accurate for the modeling of organic compounds [146, 147]. All data reported is at the M06-2X/6-311++G(d,p) level of theory, unless stated otherwise. The reaction free energies were calculated at 298 K (eqn (2.1)).

$$\Delta G_r^\circ(298K) = (\Delta G)_{Products} - (\Delta G)_{Reactants} \quad (2.1)$$

The formation of LCC linkages was considered via two reaction mechanisms to assess the favored mechanistic pathway, step-wise and concerted, depicted in Figure 2.2. The step-wise

mechanism assumes the phenolic oxygen at the 4 position in the monolignol is protonated in a barrier-less step [108] and the α carbon subsequently undergoes nucleophilic addition (Figure 2.2(a)). Comparatively, the concerted reaction mechanism assumes the nucleophilic addition to the α carbon is occurring simultaneously with the protonation of the phenolic oxygen at the 4 position of the monolignol (Figure 2.2(b)). To compute the activation barrier associated with LCC linkage formation, a similar computational strategy outlined by Durbeej and Eriksson [108] was followed.

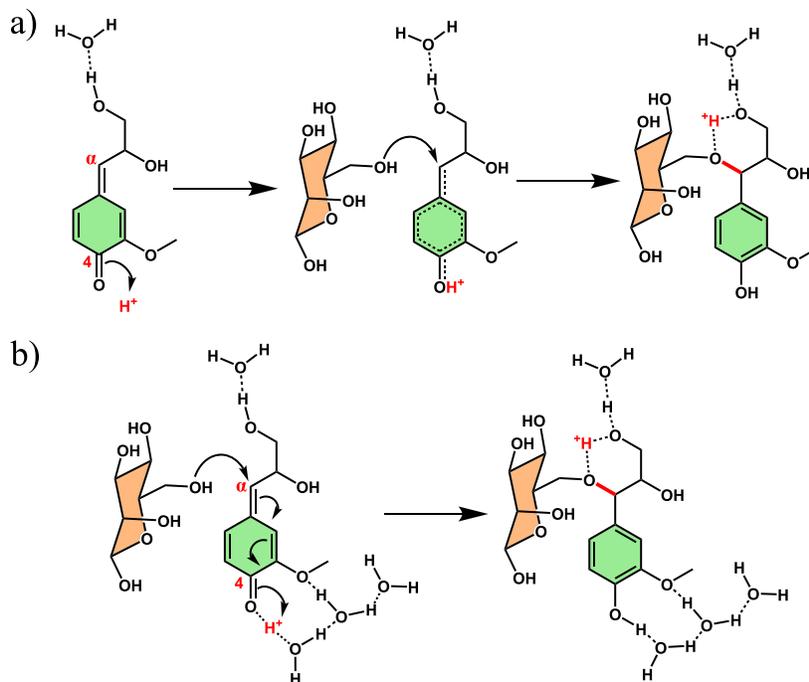


Figure 2.2: Benzyl ether formation with a mannose moiety. (a) The step-wise reaction mechanism, where the first step is protonation followed by the nucleophilic addition to the QM intermediate. (b) The concerted reaction mechanism, where protonation and nucleophilic addition to the QM intermediate occurs simultaneously. Note, dashed lines between molecules indicate hydrogen bonding.

Constrained simulations were performed where the reaction coordinate, in this case, the distance between the oxygen acting as a nucleophile and the α carbon, is varied in small increments and the highest energy structure is used as an input for the transition state (TS) search with no constraints. The Berny algorithm was used for the TS search [148, 149]. Geometry optimizations were followed by frequency calculations to obtain the structural free energy and ascertain if the structure indeed corresponded to a saddle point, *i.e.*, by the presence or absence of a single imaginary frequency. Explicit water molecules were used to help stabilize the proton in the reaction mechanisms, as performed previously [108]. Due to the “free” proton in the concerted mechanism, three water molecules were required

to provide sufficient stabilization for the appropriate identification of a TS, where a single water molecule provided the necessary stabilization of the TS structure corresponding to the step-wise mechanism. Intrinsic reaction coordinate (IRC) calculations were also performed to verify that the computed TS corresponds to the desired reactant and product. The reactant compounds obtained from the IRC were subsequently geometry optimized and used with the TS structures to determine the free energy barrier of the reaction.

To better understand the local environment of LCC formation, the reaction energetics are calculated in both the gas phase, to model a hydrophobic environment, and an implicit solvent, to model a solvated environment. The polarizable continuum model using the integral equation formalism variant was used where solvent considerations are indicated in the thermodynamic and kinetic data. Water was chosen as the implicit solvent in the continuum model ($\epsilon = 78.3553$) since the polymerization of monolignols occurs in an aqueous medium [36].

2.3 Results & Discussion

2.3.1 Model Compounds

An individual guaiacyl monolignol containing a QM structure (monolignol intermediate) is used, where the ether branching point for a dimer structure is terminated with hydrogen, forming an alcohol group. The choice of guaiacyl as the monolignol is justified since it forms nearly the entirety of softwood and the primary fraction of the most lignocellulosic biomass [36].

A total of 7 nucleophiles were considered to cause the re-aromatization of the QM intermediate at the α position of the simplified monolignol: water and the monomer form of the hemicellulose and cellulose compounds, *i.e.*, D-mannose, D-glucose, D-galactose, L-arabinose, D-xylose and 4-O-methyl glucuronic acid [32, 30]. As each hemicellulose sugar possesses multiple bonding sites that could potentially act as the nucleophile, multiple sites per sugar molecule were investigated. To discern between the various linkage sites on each sugar, the carbon number that each alcohol group is attached to will be used to distinguish between the LCC bonds. Carbon 1 for each sugar moiety begins next to the oxygen in the ring and continues clockwise around the compound. The alcohol groups at the 1 and 4 positions are excluded from considerations since they would comprise the backbone of the polymer chain. The QM intermediate and hemicellulose components as well as the carbon numbering used for each are depicted in Figure 2.3.

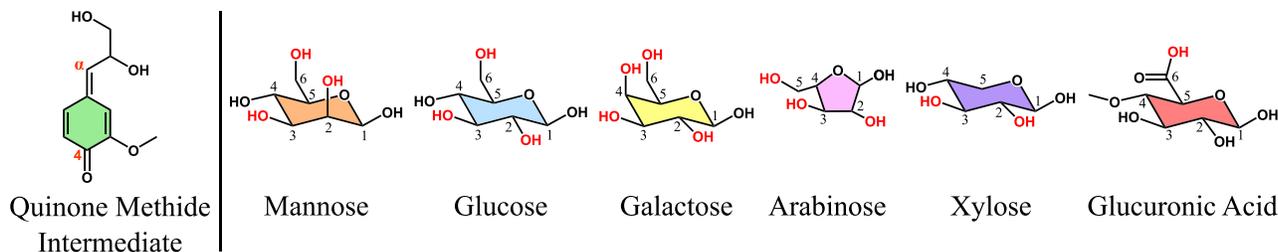


Figure 2.3: Structures of the nucleophiles considered to form an LCC linkage with the monolignol QM intermediate and the carbon numbers associated with the compounds. Red groups indicate the reaction sites considered for the nucleophilic addition to the QM intermediate.

2.3.2 Benchmarking DFT Level of Theory

In order for DFT calculations to properly capture the chemistry, it is essential that the level of theory employed provides a suitable description of the system of interest. As such, benchmarking an appropriate level of theory is necessary to ensure that the calculated results are physically relevant [150]. Figure 2.4 shows the relative reaction free energies for the nucleophilic addition of the 6 hemicellulose compounds to the QM intermediate (refer to Figure 2.3), with the reaction free energy for the nucleophilic addition of water to the QM intermediate taken as the reference. As observed in Figure 2.4, the nucleophilic addition of all sugar compounds is more favorable than that with water. Moreover, the relative trend of reaction free energies changes from a 6-31G basis set to a 6-31+G(d) basis set. Due to the relatively large structure of the LCC moiety, *i.e.*, dimensions span over 10 Å, and the presence of many oxygen atoms, the compound would contain non-covalent interactions and dipole moments. As such, it would be anticipated that polarization and diffuse functions on heavy atoms would be essential to begin capturing such discerning interactions, which is reinforced by the dramatic change in relative energetics between the 6-31G and 6-31+G(d) basis sets. Furthermore, the relative reaction free energies remain qualitatively consistent from a 6-31+G(d) to a 6-311+G(d) basis set, *i.e.*, the most thermodynamically favorable components to react with the QM intermediate, in order, are xylose, galactose, mannose, glucose, arabinose, water and then glucuronic acid. Changing a splitvalence double-zeta basis set to a triple-zeta would provide a systematic improvement in the computed energetics for all compounds. However, it would not be expected to improve the accuracy of the discerning interactions, such as the positioning and stabilization of dipole moments that would be unique to each compound and quantified more aptly by polarization and diffuse functions, which is demonstrated in the present data. Moving from a 6-311+G(d) to a 6-311++G(d,p)

basis set reverses the trend for mannose and galactose as well as for glucuronic acid and water. Overall, good qualitative agreement is maintained between the 6-311++G(d,p) basis set and the 6-31+G(d) basis set. In other words, once polarization and diffuse functions are included on heavy atoms, the majority of discerning interactions for the system are captured. Including polarization and diffuse functions on light atoms further refines the reported quantitative data; however, qualitative trends remain relatively consistent. As such, including polarization and diffuse functions are essential for the present system, as concluded previously with carbohydrate chemistry [151, 152].

Moreover, the present data suggests that the thermochemistry of LCC linkages can be captured qualitatively correctly while only including polarization and diffuse functions on heavy atoms, when comparing chemically similar systems, *e.g.*, relative energetics of benzyl ether formation. However, when comparing chemically different systems, *i.e.*, benzyl ether vs. benzyl ester formation, this trend is not captured adequately until polarization and diffuse functions are included on light atoms (hydrogen) as well. This logically follows since any ‘inaccuracy’ when describing hydrogen orbitals for chemically similar systems would be consistent, whereas the ‘inaccuracies’ for an inadequate description of hydrogen orbitals in the calculations for chemically different systems would not be systematic. Therefore, this establishes the suitable level of theories for all-electron DFT calculations for various objectives that future investigations looking into the reaction energetics of LCC linkages can confidently utilize.

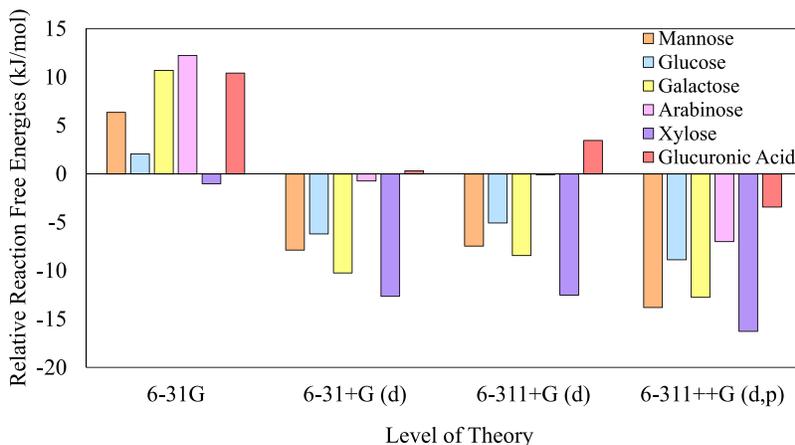


Figure 2.4: The relative reaction free energies for the nucleophilic addition of hemicellulose components to the QM intermediate, with the nucleophilic addition of water taken as the reference, as a function of the basis set. Data corresponds to gas phase calculations. The color coding of molecules is the same as in Figure 2.3.

2.3.3 Thermochemistry of LCC Linkage Formation

Figure 2.5(a) shows the nucleophilic addition to the QM intermediate of all the compounds considered in the present investigation and the reaction energetics associated with the most thermodynamically favorable reaction site of each component. Figure 2.5(b) shows the reaction free energies for LCC formation relative to the nucleophilic addition of water to the QM intermediate. Benzyl ether formation occurs when the nucleophilic addition of mannose, glucose, galactose, arabinose, or xylose reacts with the QM intermediate, whereas benzyl ester formation occurs when the nucleophile is glucuronic acid, both of which constitute an LCC linkage. Alternatively, if water acts as the nucleophile, no LCC linkage is formed, rather water chemically reacts with the QM intermediate adding a hydroxyl group to the growing lignin. As the nucleophilic addition of water has been implicitly assumed to predominately occur during monolignol polymerization [32, 36], the reaction energetics of the formation of an LCC linkage must be compared to the energetics of the nucleophilic addition of water to establish the relative thermodynamic feasibility. Since the QM intermediate possesses a single reaction site, if water reacts with it chemically, carbohydrates are left to interact physically (van der Waals, hydrogen bonding, *etc.*) and *vice versa*.

To gain a preliminary understanding of solvation effects, an implicit solvent is also incorporated. Although lignin polymerization, thus LCC formation, occurs in an aqueous medium, the exposure of the reaction site to the solvent is currently speculative. It would follow intuitively the incorporation of an implicit solvent would represent reality for a solvated system better; however, previous computational results utilizing the statistical-mechanical, 3D reference interaction site model with the Kovalenko-Hirata closure approximation [153, 154], have indicated that lignin and hemicellulose interactions are hydrophobic and entropy driven as a result of expelling water from the interaction surface [84]. Based on this study, the probability of water molecules being present within the hydrophobic interaction interface is very low, thus suggesting that the LCC linkage formation may occur in the hydrophobic environment and gas phase calculations would capture the reality better. Hence, all reaction energetics are calculated in both the gas phase and implicit solvent.

As observed in Figure 2.5, benzyl ether formation is 7.0-16.3 kJ/mol and 13.5-20.8 kJ/mol more exergonic than that of water in the gas phase and implicit solvent, respectively. Whereas benzyl ester formation is 3.4 kJ/mol and 4.4 kJ/mol more exergonic than that of water in the gas phase and implicit solvent, respectively. Therefore, the formation of LCC linkages is thermodynamically favored compared to the traditionally assumed nucleophilic addition of water to the QM intermediate that would result in physical interactions between biopolymers. However, the formation of benzyl ester linkages is noticeably less favored compared to the formation of benzyl ether linkages. This is a result of the benzyl ester LCC

moiety having to accommodate the increased sterics due to the carbonyl (C=O) group, reducing the flexibility of the ester linkage. The decreased flexibility of the ester linkage reduces the intramolecular hydrogen bonding strength, thereby lowering the relative thermodynamic stability, compared to that of benzyl ether LCC moieties. Furthermore, the incorporation of an implicit solvent increases the relative stability of LCC linkages. This difference is likely due to the shielding effect the dielectric medium has on the steric interactions between hemicellulose and lignin structures. Since the LCC linkage formation is thermodynamically favored in both the gas phase and implicit solvent, the reaction free energies do not provide evidence to discern the local reaction environment for the LCC linkage formation. However, the present results confirm that, regardless of the local environment, the formation of LCC linkages is thermodynamically favorable.

Moreover, the present computational data also fits within the context of the current literature. It has been suggested that monolignol polymerization begins by so-called initiation sites that are located in specific regions of the cell wall [155]. Although the exact nature of these sites has not been fully understood [93, 56, 156], the concept of initiation sites provides strong evidence that monolignols are interacting with other polysaccharides prior to polymerization. As such, when monolignols begin to polymerize, the QM intermediate formed would be within proximity to the cellulose/hemicellulose components for a reaction. Therefore, conclusions drawn from the present thermochemical data fit within the framework of our current knowledge of lignin polymerization and supports previous claims that LCC linkages are abundant [44].

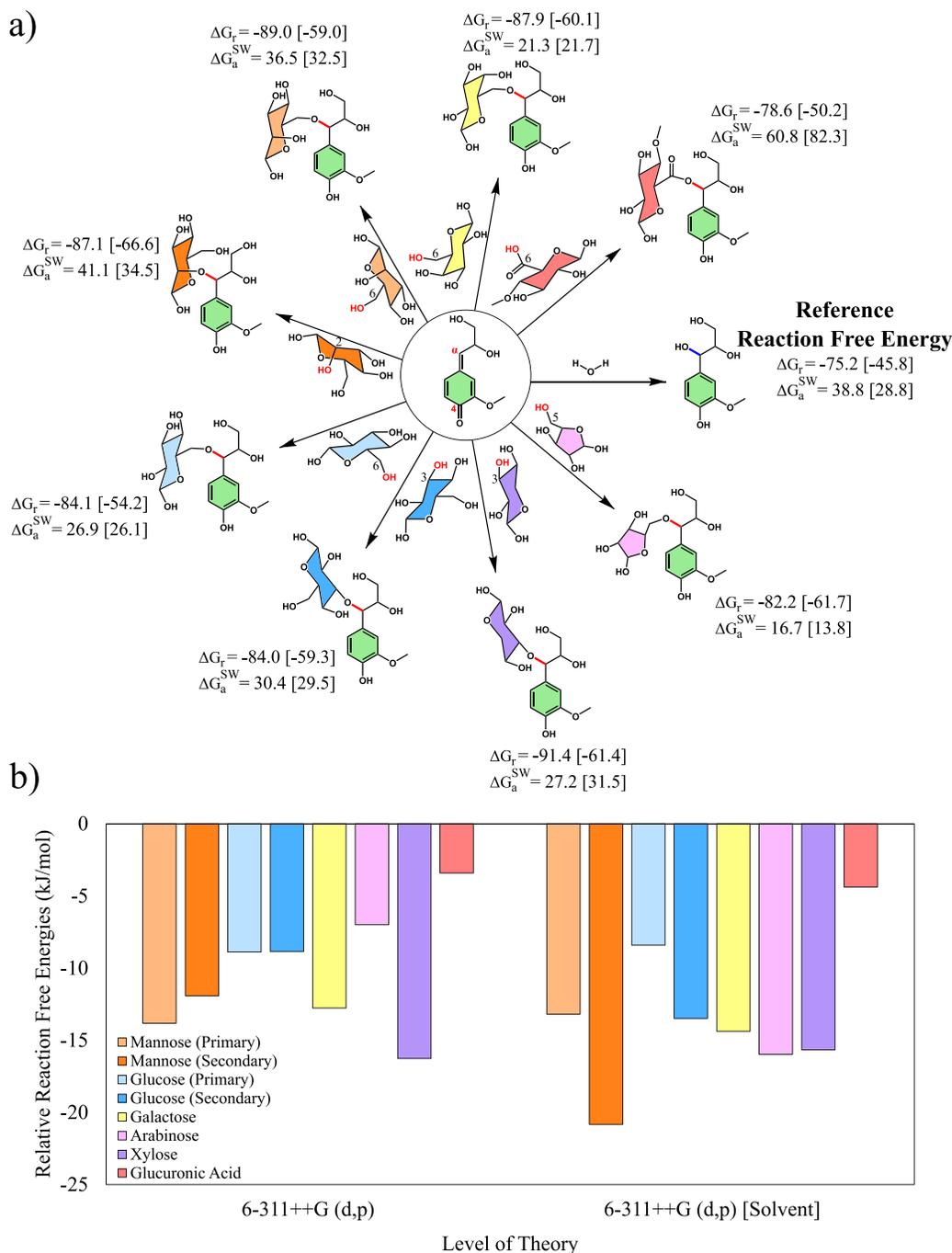


Figure 2.5: The nucleophilic addition to the QM intermediate of all the compounds considered in the present chapter. (a) The thermochemical data corresponding to the most favorable reaction site for each compound in the gas phase and implicit solvent and the associated activation free energy. Each reaction is labeled with its reaction free energy (ΔG_r , kJ/mol) and the activation energy corresponding to the step-wise mechanism (ΔG_a^{SW} , kJ/mol). Where un-bracketed numbers indicate gas phase calculations and bracketed numbers indicate implicit solvent calculations. Reaction sites and LCC linkages are highlighted in red. (b) Reaction free energies of the formation of LCC linkages, relative to the nucleophilic addition of water. The color coding of molecules is the same as in Fig. 2.3.

2.3.4 Structural Isomerism

Table 2.1 shows the calculated reaction free energies for the nucleophilic addition of mannose, glucose, galactose and arabinose, all of which possess a primary and a secondary alcohol reaction site. Specifically, mannose, glucose and galactose contain a primary alcohol at carbon 6 whereas arabinose contains a primary alcohol at carbon 5. Additionally, mannose, glucose and arabinose contain a secondary alcohol at carbons 2 and 3 whereas galactose contains a secondary alcohol at carbons 2, 3 and 4. For carbon numbering, please refer to Figure 2.3. For the hemicellulose components presented in Table 2.1, more than one secondary alcohol exists on each molecule that is capable of acting as a nucleophile to react with the QM intermediate. However, the secondary alcohol that demonstrated the most thermodynamically favorable reaction free energy for the nucleophilic addition to the QM intermediate for each molecule is reported.

As shown in Table 2.1, the reaction free energies at the primary and secondary alcohol locations on mannose, glucose, galactose and arabinose possess similar thermodynamic favorability in the gas phase. The reaction free energy at the primary alcohol location is marginally, within 2 kJ/mol, more favorable for mannose, glucose and galactose and moderately, within 6 kJ/mol, more favorable for arabinose. Hydrogen bonding largely contributes to the stabilization of the LCC moieties, with the effectiveness determined by the number and strength of hydrogen bonds. The LCC moieties that form via the nucleophilic addition of a primary alcohol, compared to the moieties formed with a secondary alcohol, have higher flexibility in the structure to form strong hydrogen bonds, *i.e.*, optimize the distance and angle of hydrogen bonding. As such, it follows that the primary alcohol locations would possess favorable reaction free energies, compared to the secondary alcohol locations, since the product compounds, the LCC moieties, are better stabilized. However, when the implicit solvent is incorporated, the reaction free energy at the primary vs. secondary alcohol shows an increased margin, over 5 kJ/mol; with mannose and glucose demonstrating a more favorable reaction site at the secondary alcohol position. The slightly unexpected switch in thermodynamic favorability between the primary and secondary alcohols for mannose and glucose is a consequence of the hydrogen bonding strength being reduced in an implicit solvent environment. The LCC moieties with mannose and glucose show fewer but stronger hydrogen bonds when bonded at the sugar’s primary alcohol compared to the secondary alcohol. Hence, the reduced hydrogen bonding strength in the implicit solvent results in the secondary alcohol being a preferred reaction site for mannose and glucose.

The present computational data agrees with the experimental work performed to date on benzyl ether LCC linkages. Nishimura *et al.* [96] provided direct evidence that lignin was covalently bonded to a mannose residue; specifically, the oxygen forming the benzyl ether LCC

linkage was connected to carbon 6 (primary alcohol) of the mannose residue. Additionally, the comparable reaction energetics for the nucleophilic addition of primary and secondary alcohol groups to the QM intermediate suggest that both sites have a similar propensity to form LCC linkages. Although the literature commonly assumes that benzyl ether linkages occur at the primary alcohol, the potential to form the LCC linkage with the secondary alcohol cannot be ruled out.

Table 2.1: Reaction free energies (kJ/mol), in the gas phase and the implicit solvent, of the nucleophiles that are capable of forming a benzyl ether linkage at the primary and secondary alcohol locations.

	Mannose	Glucose	Galactose	Arabinose
Primary Alcohol (Gas Phase)	-89.0	-84.1	-87.9	-82.2
Secondary Alcohol (Gas Phase)	-87.1	-84.0	-86.4	-76.4
Primary Alcohol (Solvent)	-59.0	-54.2	-60.1	-61.7
Secondary Alcohol (Solvent)	-66.6	-59.3	-54.8	-55.3

2.3.5 Kinetics & Mechanism of LCC Linkage Formation

To assess the relative feasibility of the formation of LCC linkages, the reaction free energies need to be evaluated within the context of the kinetic data as well. Thus, the activation barriers need to be computed to estimate if the reaction is kinetically feasible under physiological conditions. However, the exact nature of the mechanism for the nucleophilic addition to the QM intermediate has not yet been determined. Previous computational results demonstrated that the nucleophilic addition of water to the QM intermediate without an acid catalyst required an activation barrier of over 129 kJ/mol, whereas an acid catalyst reduced this barrier below 48 kJ/mol [108]. The authors assumed a step-wise mechanism and showed that the protonation of the QM structure occurred in a barrier-less step [108]. This was used as evidence to indicate the presence of an acid catalyst in lignin polymerization, which is biologically consistent with the acidic environment of the plant cell wall [157]. Therefore, based on the work of Durbeej and Eriksson [108] and the previous reports of plant cell walls, it is likely that the LCC linkage formation is an acid catalyzed reaction. However, whether the protonation of the QM structure proceeds before the nucleophilic addition to the α carbon (step-wise) or simultaneously with the nucleophilic addition (concerted), requires investigation. As such, both are evaluated in the present study.

Figure 2.6 shows the activation free energies for the nucleophilic addition of mannose at different carbon locations and that of water via step-wise and concerted mechanisms. For the mechanistic details of the reactions, please refer to Figure 2.2. The activation free

energies via the concerted mechanism are 12.8 [3.0], 3.0 [10.2] and 16.3 [10.4] kJ/mol greater for mannose at the primary alcohol (carbon 6), mannose at the secondary alcohol (carbon 2) and water than the step-wise mechanism, respectively. As the nucleophilic addition to the α carbon proceeded with a lower barrier via a step-wise mechanism in each system, it is anticipated that the step-wise mechanism would dominate *in vivo*; therefore, the concerted mechanism was not explored further.

Figure 2.5(a) shows the activation energies for the second step (nucleophilic addition to the QM intermediate) in the step-wise mechanism. Although many of the nucleophiles possessed multiple reaction sites, only the reaction sites that possessed the most thermodynamically favorable reaction free energy in the gas phase and implicit solvent were selected to computationally identify the corresponding activation free energy. Please see Figure 2.3 for the carbon numbers of the reaction site(s) considered for each hemicellulose component. As can be observed with the activation energy corresponding to the different reaction sites on mannose (carbons 2 and 6) and glucose (carbons 3 and 6), there is less than a 5 kJ/mol change in the barrier between the two sites on each respective hemicellulose compound. Therefore, it is anticipated that the change in kinetics with the various reaction sites on the hemicellulose compounds is negligible.

The activation energy values for the nucleophilic addition of mannose, glucose, galactose, arabinose, xylose and water via the step-wise mechanism are less than 42 kJ/mol; barriers for each molecule are provided in Table B.1 in Appendix B. Using Eyring’s equation and assuming that the reaction is first order, the half-life of the reaction can be determined. In the gas phase, the half-life of the reactants getting consumed, *i.e.*, nucleophilic addition to the QM intermediate, which leads to benzyl ether formation is between 9.6×10^{-11} and 1.8×10^{-6} seconds, whereas the half-life for the consumption of water via nucleophilic addition to the QM intermediate is 6.9×10^{-7} seconds. Therefore, under the physiological conditions of plant growth, *i.e.*, the temperatures at which plants grow [158], the activation free energy of <42 kJ/mol is minimal and easily surmountable (note, this is based on the intrinsic reaction kinetics, and the actual kinetics could be governed by several other factors, *e.g.*, transport, not considered in the present investigation). As such, the nucleophilic addition of water and the cellulose/hemicellulose components leading to a benzyl ether linkage show the kinetically facile reaction energetics under physiological conditions. For further calculation details of the reaction half-lives, please see Appendix B.

Furthermore, the nucleophilic addition of glucuronic acid to the QM intermediate (benzyl ester formation) demonstrates the highest activation barrier, 61 kJ/mol, in the gas phase. Although this activation energy is still surmountable under physiological conditions, it demonstrates a several orders of magnitude increase in the reaction half-life, indicating that it

would not be a kinetically competitive reaction compared to benzyl ether formation or nucleophilic addition of water. This is a result of the carbonyl oxygen in the carboxylic acid group of glucuronic acid pulling the electron density away from the carbon that is shared with the -OH group (carbon 6, refer to Figure 2.3). As such, the -OH group participating in benzyl ester formation is less nucleophilic, compared to the nucleophilic oxygens present in other sugars and water. Therefore, benzyl ester formation is unlikely to be a competitive reaction compared to the other reaction pathways, given its relative reaction free energy and activation free energy. As such, the formation of benzyl ester linkages is unlikely to occur frequently *in vivo*. This result refutes previous speculation that glucuronic acid cross links are responsible for the recalcitrant nature of biomass [32, 159] since the relative frequency is questionable given the present data.

Although the assumption of a first order reaction is a crude approximation for the step-wise mechanism presently considered, the conclusion of facile kinetics for the nucleophilic addition of water and benzyl ether linkage formation and the restricted kinetics of benzyl ester formation remain, regardless of the assumed order. Additionally, the conclusions described above are further reinforced in an implicit solvent, *i.e.*, the activation energy for benzyl ether formation decreases while the activation energy for benzyl ester formation increases.

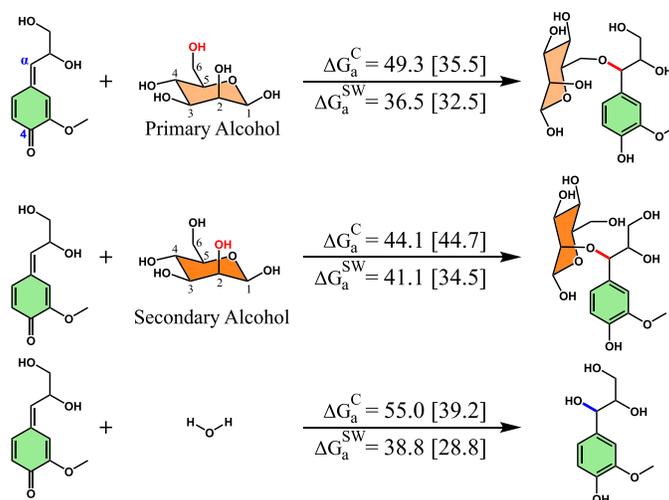


Figure 2.6: Activation free energies (ΔG_a) for the step-wise (SW) and concerted (C) mechanisms for the nucleophilic addition of mannose to the QM intermediate. Un-bracketed numbers indicate gas phase calculations and bracketed numbers indicate implicit solvent calculations.

2.3.6 Discussion

The present study demonstrates that chemical interactions between lignin and cellulose / hemicellulose are thermodynamically favorable across a wide range of compositions in the

cell wall. This has several implications. Synthetic dehydrogenation polymers (DHPs), *i.e.*, synthetic lignin, when synthesized with no external controls, show a reduced quantity of β -O-4 linkages when compared to ‘native’ lignin, suggesting that additional factors need to be considered. An experimental *in vitro* investigation with DHPs demonstrated that the presence of xylan (hemicellulose) increases the amount of DHP generated and the frequency of the β -O-4 linkages, similar to that of ‘native’ lignin [57]. The present investigation provides an explanation for such an experimental observation, since the favored formation of LCC linkages indicates that other biopolymers likely have a strong organizational role in the lignin polymerization process.

The concept of LCC linkages is typically discussed in reference to lignin linked with hemicellulose; however, the idea of lignin being linked to cellulose has also been proposed [160] (and references therein). Jin *et al.* used a carboxymethylation technique followed by NMR spectroscopy to report the existence of LCC linkages with cellulose [160]; however, satisfactory corroboration of such findings has remained scarce. In the present investigation, the nucleophilic addition of cellulose to the QM intermediate shows a more favorable reaction free energy and activation energy compared to the nucleophilic addition of water in both the gas phase and implicit solvent. Therefore, the formation of LCC linkages between lignin and cellulose is a possibility. However, the relative abundance of LCC linkages with cellulose would be anticipated to be dictated by the coverage of the cellulose fibers with hemicellulose and the extent monolignols and cellulose interact during lignin polymerization. As a result, the formation of such LCC linkages would likely be controlled by accessibility rather than relative reaction energetics.

Furthermore, the nucleophilic addition of mannose and xylose to the QM intermediate shows similar thermodynamic and kinetic feasibility. It is worthwhile to note that xylose forms the backbone of xylan-based hemicelluloses whereas mannose forms the backbone for galactoglucomannan based hemicelluloses [32]. Although various pretreatment strategies are employed depending on the biomass characteristics, no one class of feedstock has been identified as substantially ‘better’ in terms of the ease of deconstruction [18, 121]. The comparable reaction free energies and activation energies for mannose and xylose could help in explaining why no substantial difference in deconstruction for softwood vs. hardwood is readily reported since the LCC linkages between the hemicellulose backbone and lignin would be comparable. Additionally, the similar reaction energetics and kinetics of mannose and xylose are an interesting result since xylose does not possess any primary alcohols for LCC linkage formation. This also raises the question whether a benzyl ether LCC linkage to a primary alcohol compared to a secondary alcohol are chemically equivalent; especially in the context of pretreating, where the deconstruction agent needs to ‘access’ the interface between the

two biopolymers in order to separate them.

The present computational data provides evidence that the formation of benzyl ether LCC linkages exhibits facile kinetics and is thermodynamically favored. From a deconstruction perspective, the reverse reaction of benzyl ether linkages would need to be considered, which requires overcoming the positive reaction free energy and higher activation free energy barriers. Deconstructing benzyl ether linkages would need between 99 and 130 kJ/mol in the gas phase and between 76 and 101 kJ/mol in the implicit water solvent. Given the likely abundance of benzyl ether LCC linkages in biomass, this suggests that they would contribute strongly to the recalcitrant nature of lignocellulosic biomass and are essential to consider from a deconstructing perspective, as suggested previously [85, 32].

2.4 Conclusions

The lignin macrostructure is primarily composed of β -O-4 linkages, and requires the re-aromatization of a QM intermediate via nucleophilic addition. Traditionally, the nucleophile was assumed to be water, which adds a hydroxyl group to the growing lignin polymer and results in physical interactions between lignin and cellulose/hemicellulose in the secondary cell wall. However, if a reaction site on cellulose/hemicellulose acts as a nucleophile, then this leads to chemical interactions between the growing lignin structure and cellulose/hemicellulose, *i.e.*, LCCs. The present work uses all electron DFT to compute the thermochemistry and kinetics of the nucleophilic addition of model cellulose/hemicellulose compounds to the QM intermediate. The re-aromatization of the QM intermediate was determined to proceed via a step-wise mechanism, where the first protonation step was followed by nucleophilic addition. Furthermore, the reaction free energy for the nucleophilic addition of model cellulose/hemicellulose compounds, *i.e.*, mannose, glucose, galactose, arabinose, xylose and glucuronic acid, showed to be favored (50 to 91 kJ/mol) over nucleophilic addition of water, while remaining kinetically facile (14–82 kJ/mol). Therefore, the pathway forming chemical interactions between lignin and cellulose/hemicellulose (LCC linkages) is preferred over the pathway that leads to physical interactions between these biopolymers. However, formation of the benzyl ester linkage is kinetically less competitive compared to the formation of benzyl ether linkages, suggesting the benzyl ester linkages are not abundant *in vivo*. Furthermore, considering the reverse reaction, *i.e.*, deconstructing benzyl ether LCC linkages, requires surmounting a positive reaction free energy and higher activation free energy barriers (76–128 kJ/mol). Thus, providing computational evidence to support the idea of benzyl ether LCC linkages contributing to biomass recalcitrance.

Chapter 3

Deconstruction of Benzyl Ether LCC Linkages

3.1 Introduction

Despite evidence suggesting that LCCs play a key role in biomass recalcitrance, and that they are a significant impediment in designing a mild and efficient deconstruction technology, the mechanism and pathways available for LCC deconstruction remain largely unknown [42]. Sipponen *et al.* [161] sequentially treated maize stem with mild and severe alkaline conditions before and after endoglucanase treatment. This procedure yielded two distinct lignin carbohydrate fractions, where each fraction retained substantial amounts (>80%) of hydroxycinnamic acids ether-linked to lignins and much smaller amounts (<10%) of hydroxycinnamic acids ester-linked. It was concluded that the ether linkages were resistant to alkaline hydrolysis conditions. Li *et al.* [162] employed an ethanol solvothermal treatment that fractionated 97% of the total lignin and 70% the hemicellulose, with subsequent treatments using alkaline hydrogen peroxide removing the residual lignin and one-third of the remaining hemicellulose. Characterization results demonstrated two primary routes for delignification. First, the LCC ether and phenyl glycosides linkages were cleaved, with the large molecular weight portions of lignin becoming free. Second, cleavage of the carbon-carbon and ether bonds in the free lignin fractions to generate low molecular weight phenolic compounds. Yao *et al.* [163] studied the effects of process severity and pH on LCCs during deconstruction. First, corn stover was treated using hydrothermal deconstruction conditions and analyzed using a variety of experimental characterization techniques. Then a kinetic model was developed by approximating reaction pathways associated with prod-

uct distributions during biomass deconstruction as well as assuming every reaction followed pseudo-first-order kinetics and was free from transport limitations. Fitting the model to the experimental data, the reaction profiles of LCC components were calculated. It was determined that LCC deconstruction was largely pH dependent, with a low pH promoting the hydrolysis of ester linkages, contrary to Sipponen *et al.* [161], and repolymerization of aromatics into insoluble polymers, whereas a high pH promoted aldol condensations from xylose to aromatics and depolymerization of lignin to phenols. The majority of investigations taking into consideration LCC deconstruction have characterized reaction states using NMR, chromatography, spectroscopy, to name a few, and attempted to extrapolate insights into the reaction pathways associated with breaking LCCs. As such, detailed mechanistic insights into the deconstruction pathways of LCCs have not yet been elucidated. Proceeding without adequately quantifying LCC deconstruction forces investigations to continually rely on a time consuming and resource intensive trial-and-error procedures of varying operating conditions, feedstocks, catalysts and solvents while monitoring product yields to assess the efficacy of the methodology. Therefore, the present Chapter uses density functional theory (DFT) calculations to quantify the reaction energetics associated with cleaving benzyl ether LCC linkages as well as the other competing deconstruction pathways, in acidic conditions.

3.2 Computational Methodology

3.2.1 Model Compounds

The lignin moiety was composed of a guaiacyl unit, where the β -O-4 linkage for a dimer structure is simplified to an anisole group. The hemicellulose moieties were selected as a mannose, xylose or arabinose monomer. A combination of primary and secondary benzyl ether LCC linkages were modeled. The choice of guaiacyl as the monolignol is justified since it comprises nearly the entirety of lignin in softwood and is the predominant fraction in most lignocellulosic biomass [36]. Mannose, xylose and arabinose constitute the majority of hemicellulose in a wide range of lignocellulosic biomass [30, 32], and LCC linkages involving these hemicellulose moieties are the strongest benzyl ether linkages as shown in Chapter 2. Given that biomass deconstruction is typically performed under acidic conditions [125], an acid catalyst was modeled by utilizing a Zundel structure [164], *i.e.*, two water molecules and an additional proton. The model LCC compounds as well as the atomic numbering used to distinguish reaction sites are depicted in Figure 3.1. Protonation of the hydroxyl groups resulted in subsequent dehydration reactions, whereas protonation of the benzyl ether groups resulted in breaking one of the oxygen carbon bonds. In all cases, the bond

undergoing the cleavage is explicitly labeled using the notation shown in Figure 3.1. For the primary benzyl ether LCC linkage involving mannose, a total of 3 hydroxyl groups, 1 methoxy ether, 1 phenoxy ether and 1 benzyl ether groups were considered to undergo protonation and subsequently deconstruction, for a total of 8 reactions. For the primary benzyl ether LCC linkage involving arabinose as well as the secondary benzyl ether LCC linkages involving mannose and xylose, 1 benzyl ether group was considered to undergo protonation and subsequent deconstruction, for a total of 2 reactions each.

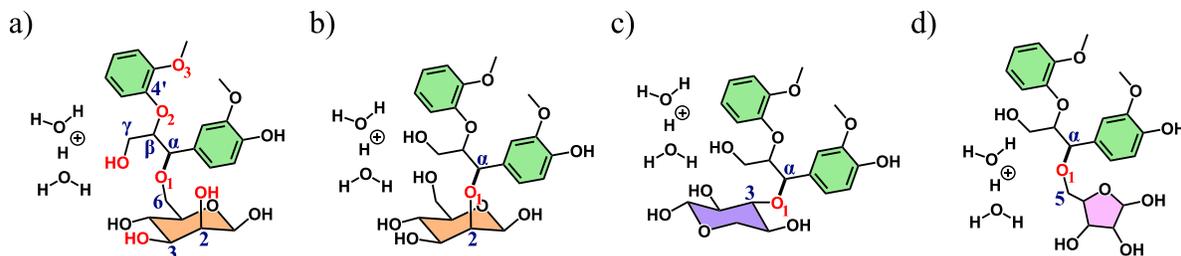


Figure 3.1: Structure of the (a) primary benzyl ether LCC linkage with mannose, (b) secondary benzyl ether LCC linkage with mannose, (c) secondary benzyl ether LCC linkage with xylose and (d) primary benzyl ether LCC linkage with arabinose as well as atomic numbering used to distinguish the reaction pathways, *e.g.*, the LCC linkage would be the $C_{\alpha} - O_1$ bond in every structure.

3.2.2 Computational Procedure

All-electron DFT calculations were performed using the Gaussian 09 code [145]. Activation barriers associated with deconstructing the LCC linkages were computed following a similar computational strategy outlined in Chapter 2. The reaction coordinate, in this case, the distance between oxygen and carbon, is constrained and varied in small increments with the Zundel structure appropriately placed to catalyze the cleavage of the bond. The highest energy structure obtained in the constrained calculations was then used as the input geometry for a full transition state (TS) search utilizing the Berny algorithm [148, 149], with no constraints. Geometry optimizations were followed by frequency calculations to compute the desired thermochemical properties and diagnose the type of stationary point, *i.e.*, assess if the optimized structure possesses a single imaginary frequency, characteristic of a saddle point. Intrinsic reaction coordinate (IRC) calculations were then performed to obtain the corresponding reactant and product structures. The reactant and product compounds obtained from the IRC were subsequently geometry optimized with no constraints and used to compute the free energy of the reaction. The optimized reactant and TS structures were then used to compute the activation free energy. The hybrid meta M06-2X functional [146] is used at every stage since it has demonstrated to be sufficiently accurate for the modeling of organic

compounds, as shown in Chapter 2. All data reported is at the M06-2X/6-311++G(d,p) level of theory. The reaction energies and activation barriers were calculated at 298 K and 1 atm, unless stated otherwise.

Given the variety of solvents commonly employed in biomass deconstruction, different implicit solvents, *i.e.*, water, tetrahydrofuran (THF) and anisole, are incorporated to help discern the local environment during LCC deconstruction. The polarizable continuum model using the integral equation formalism variant was used where a solvent medium is identified. The dielectric constant of water, THF and anisole for the implicit solvent in the continuum model were selected as 78.3553, 7.4257 and 4.2247, respectively. Water was selected as it comprises a primary fraction of solvent-based deconstruction technologies [125]. THF was selected since it has demonstrated to be a promising co-solvent in biomass deconstruction [83, 123]. Anisole was selected to model a predominantly lignin environment, in the case LCC deconstruction occurs in a solvent lean environment. All geometries are re-optimized, with no constraints, in each implicit solvent model.

3.3 Results & Discussion

3.3.1 Kinetics & Thermochemistry of Benzyl Ether LCC Deconstruction

Figure 3.2 shows multiple reaction pathways considered during the acid catalyzed deconstruction of a primary benzyl ether LCC linkage with mannose as well as the associated reaction energetics in the gas phase and three different implicit solvents. Protonation of the C_2 , C_3 and C_γ hydroxyl groups (*cf.* Figure 3.2 for atom labelling) correspond to the $C_2 - OH$, $C_3 - OH$ and $C_\gamma - OH$ bonds breaking, respectively. Protonation of the C_2 hydroxyl group results in the dehydration of the hemicellulose moiety (mannose), accompanied by ring opening and rearrangement to form a hydrated furanic species containing an aldehyde group. Protonation of the C_3 hydroxyl group results in the dehydration of mannose, accompanied by ring opening to form an ester, hydroxyl and carbon-carbon double bond functional groups. Protonation of the C_γ hydroxyl group results in the dehydration of the lignin moiety, followed by the formation of a methylated benzodioxan. Protonation of the O_1 reaction site (benzyl ether LCC group) leads to either the LCC linkage cleavage or the $C_6 - O_1$ bond breaking. If the LCC linkage breaks, mannose is separated from lignin. If the $C_6 - O_1$ bond breaks, mannose is separated from lignin and undergoes ring opening to form a structure containing aldehyde as well as hydroxyl functional groups. Protonation of the O_2 reaction site leads to either the $C_{4'} - O_2$ or the $C_\beta - O_2$ bond breaking. If the $C_{4'} - O_2$ bond

breaks, the anisole group migrates to the benzene ring in lignin. If the $C_\beta - O_2$ bond breaks, the guaiacol moiety is separated and an epoxide structure in lignin is formed. Protonation of the O_3 reaction site leads to a methyl shift, breaking the $O_3 - CH_3$ bond. The activation and reaction free energies of the different deconstruction pathways must be compared to the energetics associated with the cleavage of the LCC linkage, since breaking the LCC linkage is the desired pathway for biomass deconstruction. If any other pathway occurs, chemical degradation of the hemicellulose and/or lignin moieties will result, diminishing the biopolymer value.

As observed in Figure 3.2, the activation free energies of the competing deconstruction pathways (pathways excluding LCC linkage deconstruction) are all greater than 100 kJ/mol, indicating each pathway is kinetically less favorable compared to breaking the LCC linkages, which possesses barriers between 57-79 kJ/mol. Furthermore, the reaction free energies for the chemical degradation of the lignin moiety, *i.e.*, $O_3 - CH_3$, $C_\gamma - OH$, $C_{4'} - O_2$, and $C_\beta - O_2$ bonds breaking, are significantly, >60 kJ/mol, less favorable compared to the reaction free energies associated with breaking the LCC linkage. This is a result of the formation of unstable oxonium ions or carbocations in the product structures leading to unfavorable reaction energies. Therefore, the deconstruction pathways that require protonation of the C_γ hydroxyl group, O_2 , and O_3 sites are kinetically and thermodynamically hindered compared to the other available pathways, suggesting such reaction sites may not be prevalent under biomass deconstruction conditions. Moreover, the reaction free energies for the chemical degradation of the hemicellulose moiety, *i.e.*, $C_2 - OH$, $C_3 - OH$, and $C_6 - O_1$ bonds breaking, are all thermodynamically more favorable compared to breaking the LCC linkage. This is primarily attributed to the water molecules improving the stabilization of the formal charge through a series of hydrogen bonds and can be observed in the molecular structures, depicted in Appendix C.

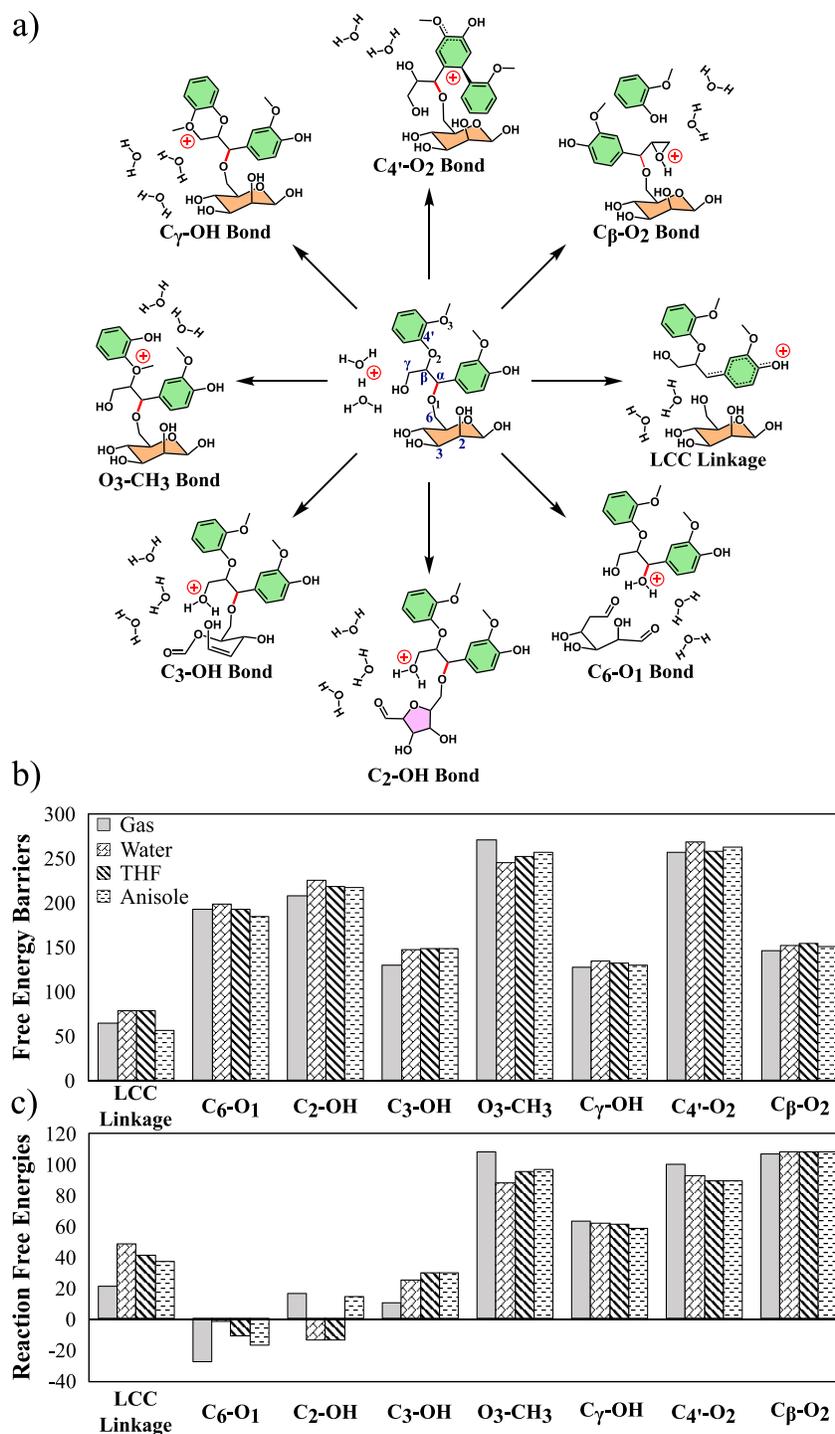


Figure 3.2: (a) The reaction pathways considered in the acid catalyzed deconstruction of the primary benzyl ether LCC linkage with mannose. (b) The activation free energy (ΔG_a , kJ/mol) and (c) reaction free energy (ΔG_r , kJ/mol), in gas phase, water, tetrahydrofuran (THF) and anisole, of the reaction pathways illustrated in (a).

3.3.2 Composition Dependence of Benzyl Ether LCC Deconstruction

Figure 3.3 shows the deconstruction pathways and reaction energetics in the gas phase as well as three different implicit solvents for varying compositions of hemicellulose involved in benzyl ether LCC linkages. Protonation of the O_1 reaction site on each model LCC compound leads to either cleavage of the LCC linkage, or a hemicellulose degradation pathway, *i.e.*, cleavage of the $C_6 - O_1$ for the primary mannose linkage, $C_2 - O_1$ for the secondary mannose linkage, $C_3 - O_1$ for the xylose linkage and $C_5 - O_1$ for the arabinose linkage. If the LCC linkage breaks, the hemicellulose moiety (mannose, xylose and arabinose) is separated from lignin. If the $C_2 - O_1$ bond or the $C_5 - O_1$ bond breaks the hemicellulose moiety (mannose and arabinose) is separated from lignin and undergoes ring opening to form a structure containing an aldehyde, hydroxyl, ether and carbon-carbon double bond functional groups. If the $C_3 - O_1$ bond breaks xylose undergoes ring opening and rearrangement to form a hydrated furanic species containing an aldehyde group.

As shown in Figure 3.3, the reaction free energies for the deconstruction of the primary mannose linkage and the xylose linkage favors the hemicellulose degradation pathways ($C_6 - O_1$ and $C_3 - O_1$), whereas the reaction free energies for the two deconstruction pathways associated with the secondary mannose linkage (LCC linkage and $C_2 - O_1$) are comparable. The reaction free energies for the deconstruction of the arabinose linkage thermodynamically favors breaking the LCC linkage over the hemicellulose degradation pathway. Therefore, the thermodynamic feasibility of deconstructing LCC linkages demonstrates a correlation on the hemicellulose chemical composition.

Furthermore, the activation energies of the hemicellulose degradation pathways, $C_6 - O_1$, $C_2 - O_1$, $C_3 - O_1$ and $C_5 - O_1$, are between 75 – 125 kJ/mol which are greater than the barriers associated with the cleavage of the corresponding LCC linkages, which possess barriers in the range of 57 – 106 kJ/mol. As such, the thermodynamic feasibility of cleaving the LCC linkage, compared to the competing hemicellulose degradation pathways, is dependent on the hemicellulose composition and on the location of the LCC linkage, whereas the preferred kinetic feasibility of cleaving the LCC linkage is independent of such factors.

The relative barriers and reaction free energies presented in Figure 3.2 and Figure 3.3 are similar regardless of the implicit solvent environment considered, *i.e.*, gas phase, water, THF and anisole. However, the presence of explicit solvent molecules may change such conclusions. As a result, the reaction energetics computed in the gas phase for the acid catalyzed deconstruction of benzyl ether LCC linkages are assumed adequate to accurately represent the system across the varying dielectric environments imposed by the implicit solvents studied presently.

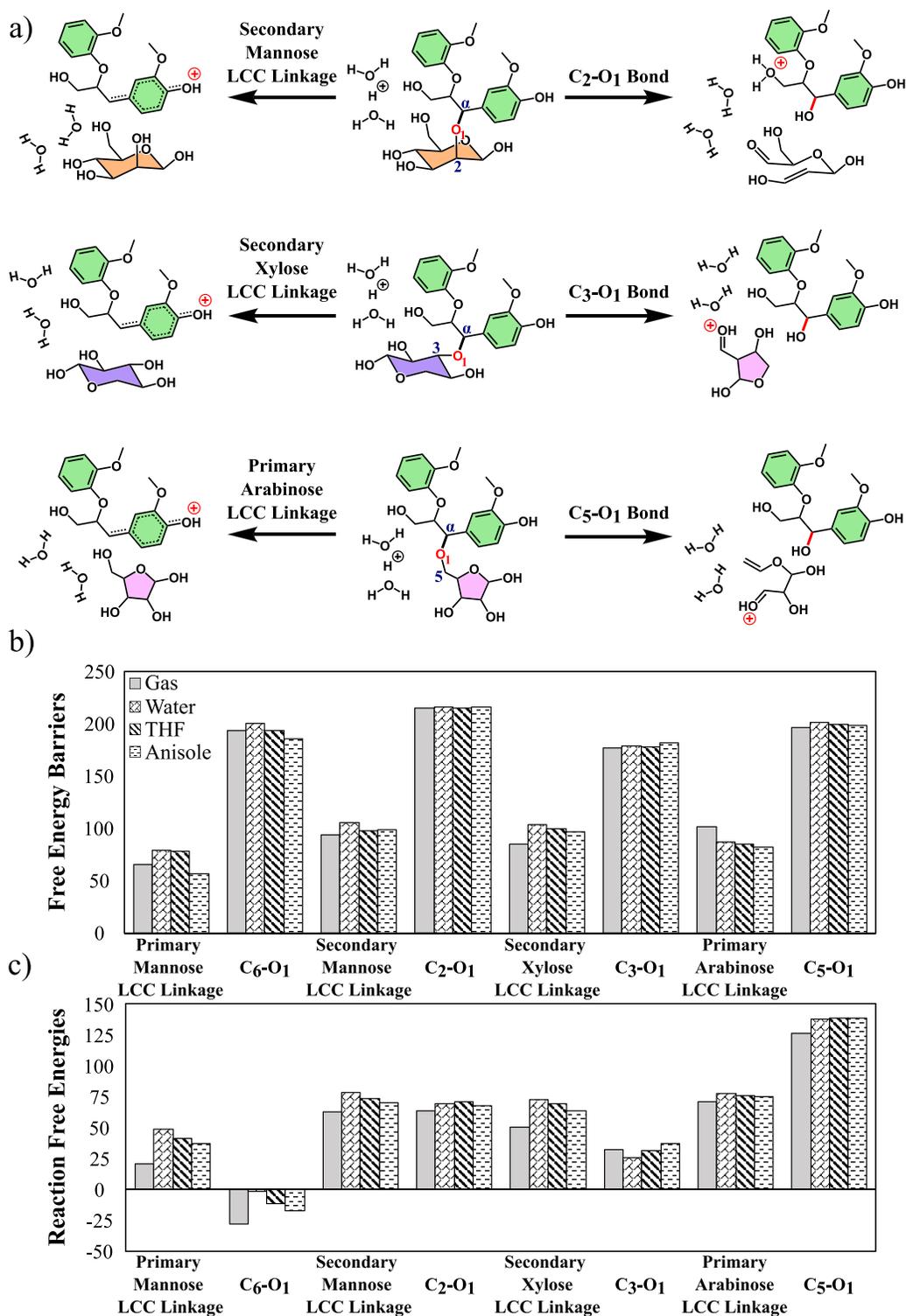


Figure 3.3: (a) The two deconstruction pathways accessible, across varying hemicellulose compositions, when the oxygen participating in LCC linkages, *i.e.*, the O_1 reaction site, is protonated. (b) The activation free energy (ΔG_a , kJ/mol) and (c) reaction free energy (ΔG_r , kJ/mol), in gas phase, water, tetrahydrofuran (THF) and anisole, of the reaction pathways illustrated in (a).

3.3.3 Reaction Energetics Temperature Dependence

Across the variety of biomass deconstruction strategies, elevated temperatures are consistently employed [18, 19], as such, assessing the reaction energetics as a function of temperature is necessary to bring the computational results closer to experimental conditions. Figure 3.4 and Figure 3.5 shows the activation (denoted with subscript “a”) and reaction (denoted with subscript “r”) free energies, respectively, for cleaving the mannose, xylose and arabinose LCC linkages as well as the competing hemicellulose degradation pathways (C_6-O_1 , C_2-O_1 , C_2-OH , C_3-OH , C_3-O_1 and C_5-O_1 bonds breaking), as a function of temperature. The solid lines represent the predicted free energies and were calculated by assuming the enthalpic and entropic contributions were constant at the values computed at 298K. As such, the predicted reaction and activation free energies were determined by simply incorporating the absolute temperature, *i.e.*, $\Delta G = \Delta H - T\Delta S$.

The predicted free energies (solid lines) are all within 7 kJ/mol of the computed free energies, which incorporate the temperature corrected enthalpic and entropic contributions. Furthermore, given the similarity in reaction energetics between the gas phase and implicit solvent environments for each system (Figure 3.2 and 3.3), it is unlikely the free energies dependence on temperature will be drastically affected by the incorporation of a dielectric constant. Therefore, the predicted free energies in the gas phase give a reasonably accurate representation of the activation and reaction free energies dependence on temperature. Note, given the unfavorable reaction energetics as well as minimal entropic changes associated with the pathways degrading the lignin moiety (O_3-CH_3 , $C_\gamma-OH$, $C_{4'}-O_2$ or the $C_\beta-O_2$), the temperature dependence of these pathways was not explored.

As observed in Figure 3.4, the free energy barriers for cleaving the LCC linkages are all less than ~ 100 kJ/mol, with the primary mannose, xylose and arabinose LCC linkages demonstrating an inverse correlation with temperature, *i.e.*, as the temperature is increased the activation energies decrease, because of the larger entropy gain in the TS compared to the reactant. Such a result is anticipated for most dissociation reaction. However, the barrier associated with the secondary mannose LCC linkage demonstrates a temperature independence, a result of the much smaller entropic contribution.

Shown in Figure 3.4, the majority of the hemicellulose degradation pathways, C_6-O_1 , C_2-O_1 , C_2-OH , C_3-O_1 and C_5-O_1 bonds breaking, have free energy barriers greater than ~ 160 kJ/mol, suggesting elevated temperatures or severe conditions will need to be employed before these pathways become kinetically accessible. The free energy barrier associated with breaking the C_3-OH bond shows a barrier between 110-130 kJ/mol. Although such a barrier will still require elevated temperatures and severe conditions (significantly higher than NTP) to overcome, it is the minimum energy pathway associated with hemicellulose

degradation. Furthermore, the free energy barriers associated with all the hemicellulose degradation pathways, $C_6 - O_1$, $C_2 - O_1$, $C_2 - OH$, $C_3 - OH$, $C_3 - O_1$ and $C_5 - O_1$ bonds breaking, all show an inverse correlation with temperature.

As observed in Figure 3.5, the reaction free energy of every pathway shows an inverse correlation with temperature, indicating as the temperature increases, the reaction energies will become more favorable, characteristic of most dissociation reactions. The hemicellulose degradation pathway with the $C_6 - O_1$ bond breaking is exergonic for the entire temperature range explored. The $C_3 - OH$ bond breaking is endergonic until $\sim 430K$, where it becomes exergonic at greater temperatures. A temperature of $430K$ also likely provides sufficient thermal energy to overcome the associated barrier of ~ 125 kJ/mol. Cleaving the primary mannose LCC linkage as well as breaking the $C_2 - OH$ bond are exergonic above $\sim 640K$. At such elevated temperatures it is also likely to overcome the ~ 190 kJ/mol barriers associated with the $C_6 - O_1$ and $C_2 - OH$ bonds breaking. Therefore, simply increasing the temperature to establish an exergonic reaction for the cleavage of the primary mannose LCC linkage will inevitably result in multiple hemicellulose degradation pathways becoming kinetically facile and thermodynamically favored. Moreover, the reaction free energy associated with cleaving the xylose LCC linkage shows a greater change with temperature compared to the competing hemicellulose degradation pathway, $C_3 - O_1$. As such, at $\sim 930K$, cleavage of the xylose LCC linkage becomes exergonic, whereas, breaking the $C_3 - O_1$ bond becomes exergonic at $\sim 990K$. However, $930K$ is well above the temperature range employed during biomass deconstruction, making such an approach unfeasible [18, 19]. Furthermore, cleavage of the secondary mannose LCC linkage as well as the arabinose LCC linkage are endergonic, and the reaction energies have minimal changes as a function of temperature, indicating there will be no reasonable temperature or severity that will establish an exergonic reaction. Therefore, preferential cleavage of benzyl ether LCC linkages requires modifying the reaction environment to establish favorable reaction thermodynamics compared to the competing hemicellulose degradation pathways, irrespective of hemicellulose composition or linkage type.

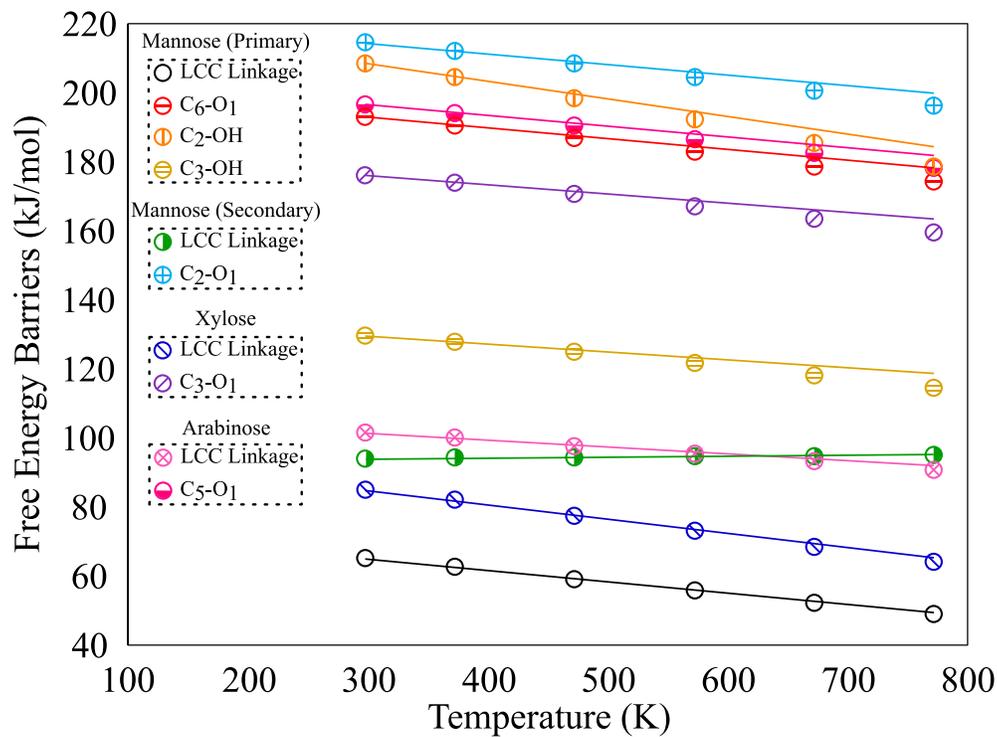


Figure 3.4: The free energy barriers for cleavage of the various LCC linkages as well as the competing hemicellulose degradation pathways as a function of temperature, in the gas phase. The solid lines are calculated assuming the enthalpic (ΔH_a) and entropic (ΔS_a) contributions are constant with the values computed at 298K. Note, since no significant differences were observed in the reaction energetics when an implicit solvent was incorporated, only gas phase data is reported.

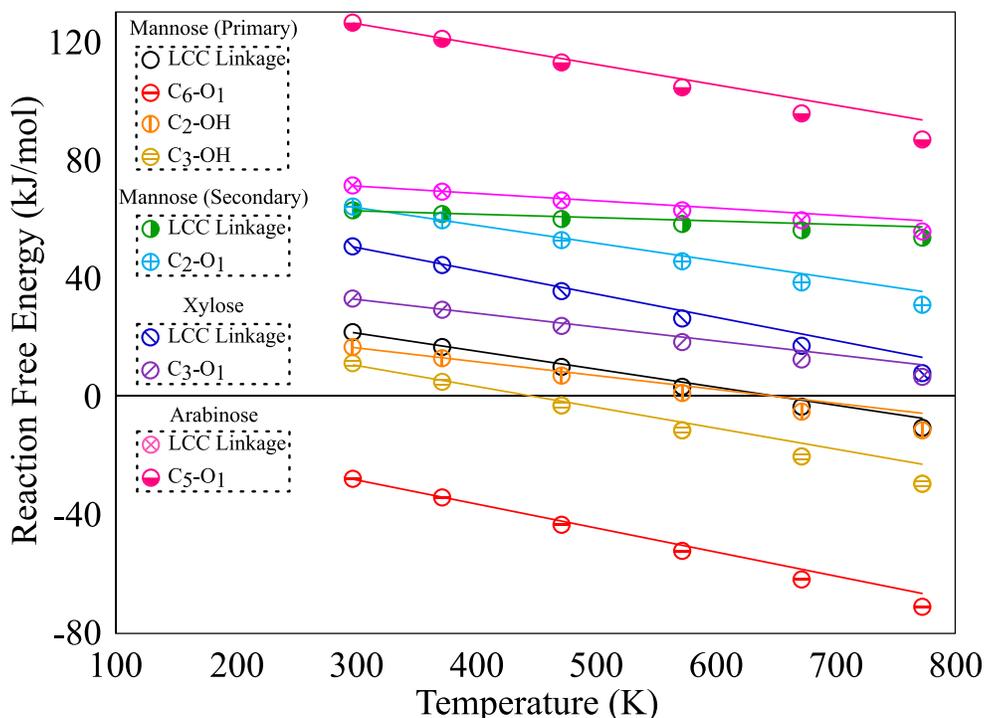


Figure 3.5: The reaction free energies for cleavage of the different LCC linkages as well as the competing hemicellulose degradation pathways as a function of temperature, in the gas phase. The solid lines are calculated assuming the enthalpic (ΔH_r) and entropic (ΔS_r) contributions are constant with the values computed at 298K. Note, since no significant differences were observed in the reaction energetics when an implicit solvent was incorporated, only gas phase data is reported.

3.3.4 Discussion

Biomass deconstruction routinely involves temperatures ranging from 423 – 523 K, pressures on the order of tens of atmospheres and the addition of acidic compounds to assist in the deconstruction [18, 19]. Utilizing such severe conditions implicitly assumes that biomass deconstruction is a kinetically controlled process, *i.e.*, large activation barriers need to be surmounted to extract the desired components. However, cleavage of the benzyl ether LCC linkages possesses low activation barriers, < 100 kJ/mol, indicating such reactions would be kinetically facile, at modest operating conditions. The challenge associated with breaking benzyl ether LCC linkage is establishing conditions to make the reaction exergonic. Comparatively, hemicellulose degradation pathways, such as $C_6 - O_1$ and $C_3 - OH$ bonds breaking, are exergonic at deconstruction temperatures, but possess larger activation barriers, > 130 kJ/mol. Therefore, the present results suggest that biomass deconstruction is associated with a threshold temperature or severity factor, at which either cleaving LCC linkages will become exergonic, or the barriers associated with hemicellulose degradation will begin to

become surmountable.

Considering the reaction energetics' dependence on temperature, as observed in Figure 3.4 and Figure 3.5, it is anticipated that the larger activation barriers, corresponding to hemicellulose degradation, will become surmountable before LCC linkage cleavage becomes exergonic. At which point, thermodynamics would play a key role governing the reaction pathways and favor hemicellulose degradation.

These conclusions are also supported by experimental observations. Nguyen *et al.* [165] treated Douglas fir in acidic conditions at elevated temperatures and used xylose yield as an indication of hemicellulose hydrolysis. At 474 K the xylose and furfural yields were 61.7% and 7.1%, respectively, whereas at 503 K the xylose and furfural yields changed to 12.8% and 10.2%, respectively. Nguyen *et al.* [166] later treated mixed softwood forest thinnings in a two-stage reactor, employing steam and acidic conditions. Treating at 463 K then subsequently at 488 K resulted in 10% of the xylose being converted to furfural and 14.5% of the mannose being converted to hydroxymethylfurfural (HMF). Treating at 473 K then subsequently at 488 K resulted in 27% of the xylose being converted to furfural and 24% of the mannose being converted to HMF. Finally, Nguyen *et al.* [167] established a correlation between the treatment severity factor (temperature and pH) and sugar yield. A temperature and pH of 463 K and 0.95, respectively, resulted in a hemicellulose sugar yield of $\sim 80\%$. However, a temperature and pH of 488 K and 0.49, respectively, resulted in a yield of $\sim 15\%$.

The series of works by Nguyen and coworkers [165, 166, 167] establishes that a certain temperature or severity will result in a significant decrease of the hemicellulose yield, while the conversion of hemicellulose to products such as furfural and HMF increases. In other words, the yield of hemicellulose corresponds to the level of degradation. At a threshold temperature/severity, hemicellulose will begin to degrade, decreasing the yield, and this will lead to an increase in conversion reactions to products such as furfural and HMF. Therefore, the experimental work conducted by Nguyen and coworkers [165, 166, 167] demonstrates the temperature sensitivity hemicellulose degradation possesses, which is in agreement with the present computational results.

During solvated biomass deconstruction treatments, organic acids are generated that have been reported to catalyze the hydrolysis of hemicellulose [18] as well as lead to the depolymerization/repolymerization of lignin [90]. Given the complexity of deconstruction, the reaction conditions can dramatically fluctuate depending on the chemical nature of the degraded components and it has been speculated that non-native LCC linkages could form in such variable conditions [87]. Although severe operating conditions will eventually establish favorable reaction thermodynamics for cleaving the LCC linkage, the reverse reaction, *i.e.*,

LCC linkage formation, possesses minimal barriers and favorable reaction thermodynamics at less severe conditions (Figure 3.4 and Figure 3.5), which inevitably occur during any process.

Comparatively, once a bond associated with hemicellulose degradation breaks ($C_6 - O_1$, $C_2 - OH$, $C_3 - OH$, $C_2 - O_1$, $C_3 - O_1$ and $C_5 - O_1$), the probability of the reverse reaction occurring is negligible, due to the large activation barrier and significant conformational rearrangement necessary for the altered hemicellulose moiety to participate in the reverse reaction. Therefore, the present investigation suggests that biomass recalcitrant is a result of the cleavage of LCC linkages being thermodynamically unfavorable (endergonic). Furthermore, the primary challenge with cleaving LCC linkages is establishing reaction conditions to make the step exergonic, while suppressing their re-formation throughout the reaction conditions present during biomass deconstruction.

The utilization of co-solvents such as tetrahydrofuran (THF), methanol, ethanol, acetone, γ -valerolactone (GVL), *etc.*, have shown to deconstruct biomass in milder acidic conditions, thereby improving the recovery of cellulose, hemicellulose and lignin [18, 21, 125, 83]. The role of solvents in biomass conversion can be categorized as (1) affecting the solubility (physical interactions) of the biomass components, *i.e.*, cellulose, hemicellulose and lignin, as well as (2) affecting the reaction thermodynamics [125]. Most of the literature investigating biomass deconstruction primarily focuses on characterizing the solubility of biomass components, *i.e.*, the physical interactions between the solvent and polymer systems [24, 132, 130, 129, 131]. Comparatively, the effect of solvents on the reaction thermodynamics has garnered far less attention, despite investigations demonstrating the importance in selective conversion within the biomass context.

For example, cellobiose hydrolysis as well as glucose and xylose dehydration demonstrated similar activation barriers between 130-140 kJ/mol in the aqueous phase [168]. However, incorporating a GVL/water solvent system reduced the cellobiose hydrolysis barrier to ~ 80 kJ/mol while leaving the barriers associated with dehydrating glucose and xylose unchanged [168]. Additionally, many researchers have demonstrated that the yields of HMF from fructose dehydration reactions are greater in a polar aprotic solvents such as dimethyl sulfoxide (DMSO), THF or GVL compared to the yields in pure water [125]. Tsilomelekis *et al.* [169] showed that DMSO preferentially solvates the HMF carbonyl group in DMSO/water and increases HMFs lowest unoccupied molecular orbital (LUMO) energy, which decreased its susceptibility to nucleophilic attack, thus improving the stabilization of HMF and explaining the improved yields in selective solvents.

In the present investigation, the deconstruction pathways for benzyl ether LCC linkages are elucidated. Cleavage of the LCC linkages are endergonic at deconstruction conditions,

whereas the pathways associated with hemicellulose degradation are exergonic, despite relatively larger activation energies. It is predicted that the improved deconstruction of biomass when employing specific solvents is a result of the solvent environment promoting the reaction thermodynamics associated with the cleavage of the LCC linkage, compared to the various hemicellulose degradation pathways. Since the implicit solvent model presently implemented in this work does not provide discernible differences in the reaction energetics of the various solvent environments, it is anticipated complex solvent effects (reactant/TS/product stabilization, confinement, inhibition, solvation dynamics to name a few) [170] are at the heart of the improved biomass deconstruction in selected solvents. As such, the implementation of an explicit solvent is necessary to computationally characterize the solvent effects during biomass deconstruction. Overall, this investigation provides evidence to suggest that the thermodynamics of LCC linkages needs to be considered when designing new biomass deconstruction technologies to recover chemically intact cellulose, hemicellulose and lignin.

3.4 Conclusions

The recalcitrant nature of biomass has remained a fundamental impediment in designing mild deconstruction technologies that preserve the chemical integrity and atom-efficient recovery of cellulose, hemicellulose and lignin. Although covalent bonds between lignin and carbohydrates, *i.e.*, LCCs, have been recognized to play a key role in this recalcitrance, the nature of their role in deconstruction is largely unknown and implicitly assumed to be one of kinetic limitations. This Chapter employed computational methods to evaluate thermodynamics and kinetics associated with the deconstruction of benzyl ether LCC linkages with varying hemicellulose compositions, as well as associated competing pathways, in acidic conditions. The protonation and subsequent deconstruction of the LCC linkages, lignin moiety and hemicellulose moieties possessed free energy barriers between 57-105 kJ/mol, 128-271 kJ/mol and 130-214 kJ/mol, respectively. The degradation of the lignin moiety was highly endergonic with large associated barriers, suggesting such pathways may not be prevalent during deconstruction. Cleaving the LCC linkages was moderately endergonic whereas degrading the hemicellulose moieties was exergonic or mildly endergonic. Increasing the temperature improves the thermodynamic favorability of all the pathways, however, increasing the temperature also increases the likelihood of overcoming higher free energy barriers. Computing the reaction energetics as a function of temperature demonstrated that increasing the temperature of the system would likely favor overcoming the large barriers associated with hemicellulose degradation before an exergonic regime could be established for cleaving the

LCC linkages. As a result, increasing the severity of deconstruction conditions to cleave LCC linkages will inevitably lead to overcoming the large barriers associated with the degradation of hemicellulose. At which point, thermodynamics would play a primary role governing the reaction pathways and favor hemicellulose degradation, which is observed experimentally. The present results suggest that the challenge of overcoming biomass recalcitrance, *i.e.*, breaking LCC linkages, is in creating a reaction environment that promotes favorable reaction thermodynamics for cleaving the LCC linkages.

Chapter 4

Formation & Deconstruction of Phenyl Glycoside LCC Linkages

4.1 Introduction

Of the various types of LCC linkages, PGs remain one of the most common linkages reported in the analysis of biomass structure [86, 89, 171, 172, 173, 174, 175, 176, 177]. Despite the reported relevance of PG LCC linkages, their formation mechanism is rarely discussed and remains largely unknown [86, 87, 103]. Presently, there are three proposed mechanisms for PG formation: hemi-acetal formation, transglycosylation and the incorporation of monoglucosides. All mechanisms as well as the carbon numbering used are depicted in Figure 4.1 [87, 100, 178].

Hemi-acetal formation requires an acid catalyst to open the hemicellulose ring, creating an aldehyde group at the reducing end (C_1). The phenolic hydroxyl group of lignin then attaches to the aldehyde carbon (C_1), and the aldehyde oxygen is protonated forming a hydroxyl group. The newly formed hydroxyl group at C_1 is protonated again and undergoes dehydration which allows the hemicellulose ring to close and complete the formation of the PG linkage. Transglycosylation is assumed to be catalyzed by an enzyme possessing acidic and basic sites and involves two bimolecular nucleophilic substitution (S_N2) reactions, as shown in Fig. 1b. First, the acidic enzymatic site donates a proton to the hydroxyl group at the reducing end of the hemicellulose moiety (C_1) making it a facile leaving group. The first S_N2 reaction occurs with the basic enzymatic site acting as a nucleophile to bond to C_1 , while the protonated hydroxyl group detaches (dehydration) to form a water molecule. As a result of donating a proton, the previously acidic site on the enzyme becomes basic and begins

abstracting the hydrogen on the phenolic hydroxyl group of lignin, making it a stronger nucleophile. The second S_N2 reaction occurs with the phenolic hydroxyl group of lignin acting as a nucleophile to bond to C_1 , forming a PG linkage, while the attached enzymatic group detaches as a leaving group. Although hemi-acetal formation and transglycosylation have been proposed to be potentially feasible mechanisms for PG formation, neither of them have been evaluated for their kinetic feasibility. Furthermore, it has been suggested that monolignol glucosides, *i.e.*, a carbohydrate already containing a PG linkage, could be potential lignin precursors [178, 100]. Miyagawa *et al.* [100] demonstrated that monolignol glucosides could be incorporated into the lignin macrostructure, leading to PG linkages, by comparing NMR data of synthetic lignin made in the presence of monolignol glucosides to native biomass sources. However, only in the spectra from the LCC concentrated fraction of the native biomass sources, they observed low levels of agreement between the signals arising from PG linkages in native biomass and the PG linkages in the synthetic lignin. As such, whether monolignol glucosides alone could account for the fraction of PG linkages observed in different biomass sources remains to be confirmed, especially since the significance of monolignol glucosides in specific plant species have been shown to be minimal [179]. Therefore, further definitive evidence of PG formation via the incorporation of monolignol glucosides is required.

Understanding the formation mechanism of PG linkages sheds further light on lignin polymerization, a critical phenomenon that largely dictates the biopolymers' organization and interactions within biomass. Moreover, given that PG linkages are found in significant fractions across different biomass sources, they could also potentially contribute to biomass recalcitrance. Therefore, quantifying the formation and deconstruction of PG linkages is not only of fundamental interest, but also practically important for designing mild deconstruction technologies as well as tailoring valorization pathways for bio-chemicals. As such, the present Chapter uses density functional theory (DFT) calculations to quantify the reaction energetics associated with PG formation, via hemi-acetal formation and transglycosylation. Additionally, a novel mechanism is proposed and investigated, *i.e.*, electrophilic addition of hemicellulose to a lignin quinone methide (QM) intermediate. Furthermore, the first deconstruction step of PG LCC linkages in acidic conditions is elucidated and quantified.

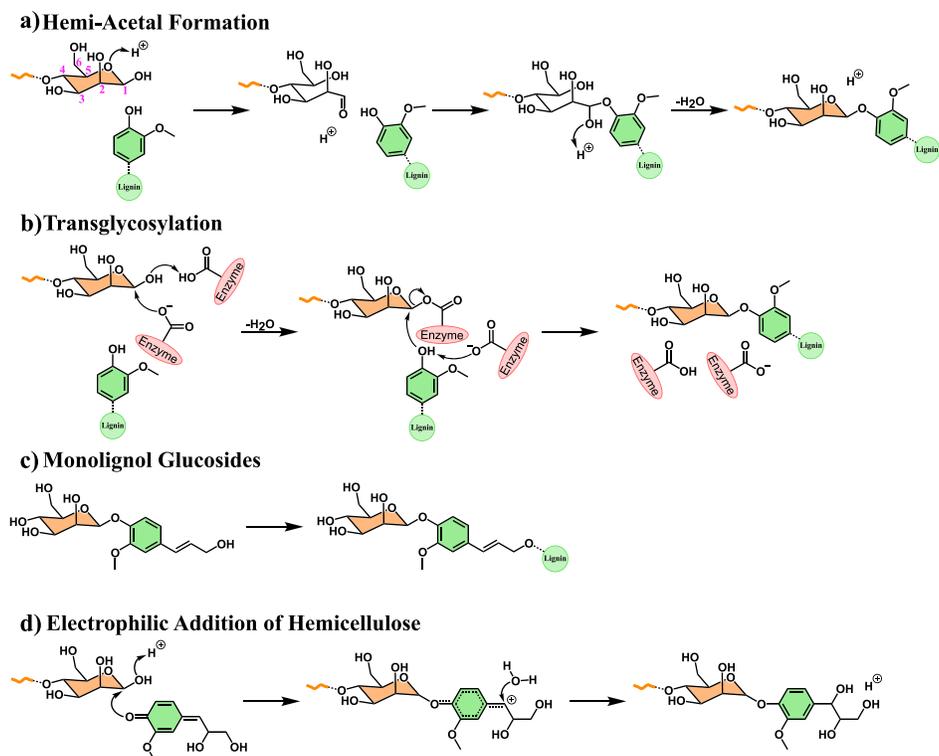


Figure 4.1: Mechanisms proposed in the literature for the formation of phenyl glycoside (PG) linkages in biomass (a) hemi-acetal formation (b) transglycosylation (c) monolignol glucosides, and (d) a newly proposed mechanism of electrophilic addition of hemicellulose to the lignin quinone methide (QM) intermediate. The carbon numbering of the hemicellulose moiety (mannose) is shown in the hemi-acetal mechanism (a).

4.2 Computational Methodology

4.2.1 Model Compounds

Lignin was modeled using guaiacyl units since guaiacyl comprises the majority of lignin in softwood and a predominant fraction in most lignocellulosic biomass [36]. The aliphatic portion (the “tail”) of guaiacyl is modeled when it directly participates in the reactions, otherwise only the aromatic portion of the molecule is retained for the reaction pathways considered. The hemicellulose moieties were selected as mannose and xylose since each represent the primary monomer in mannans and xylans, respectively, making these monomers representative of the majority of hemicellulose in lignocellulosic biomass [31, 30, 32]. As lignin polymerization occurs in a mildly acidic environment [157, 54, 180] an acid catalyst was modeled for the formation of PG linkages. For hemi-acetal formation and electrophilic addition of hemicellulose to a QM intermediate, a Zundel structure [164], *i.e.*, two water molecules and an additional proton, is utilized. For transglycosylation, an enzyme con-

taining acid/base characteristics is speculated to be present, therefore, acetic acid and an acetate ion are selected to represent the acidic and basic moieties of the enzyme, respectively. Three unique reaction pathways are considered for the formation of PG linkages, hemi-acetal formation, transglycosylation and electrophilic addition hemicellulose to the lignin QM intermediate. However, since any carbon attached to a hydroxyl group can act as an electrophile on hemicellulose, the electrophilic addition of hemicellulose to the lignin QM intermediate was considered for all available carbons, *i.e.*, five pathways are considered for mannose and four pathways are considered for xylose. Moreover, the deconstruction pathways of a PG linkage are considered and are modeled with the aromatic portion of a guaiacyl unit bonded with either mannose or xylose. Given that biomass deconstruction is typically performed under acidic conditions [125], an acid catalyst was modeled using a Zundel structure [164]. Four deconstruction pathways are considered for the PG linkage with mannose as well as with xylose. Two pathways protonate and subsequently dehydrate the hydroxyl groups on hemicellulose, whereas the other two pathways protonate the ether in the PG linkage and then subsequently deconstruct the LCC moiety.

4.2.2 Computational Procedure

All-electron DFT calculations were performed using the Gaussian 09 code [145]. The transition states (TS) associated with each reaction pathway were computed following a similar computational strategy described in Chapter 2. The reaction coordinate, typically the distance between two nuclei, is constrained and varied in small increments with the acid or base complex appropriately placed to catalyze the formation or cleavage of a bond. The highest energy structure obtained in the constrained calculations was then used as the input geometry for a full TS search utilizing the Berny algorithm [148, 149], with no constraints. Geometry optimizations were followed by frequency calculations to compute the thermochemical properties as well as to diagnose the type of stationary point, *i.e.*, determine if the optimized structure possessed a single imaginary frequency, distinctive of a saddle point. TS structure identification was followed by intrinsic reaction coordinate (IRC) calculations to obtain the corresponding reactant and product structures. The reactant and product structures obtained from the IRC calculation were subsequently optimized with no constraints and used to compute the free energy of the reaction. The optimized reactant and TS structures were then used to compute the activation free energy. For multi-step reactions, this procedure was repeated, where the product of the subsequent reaction connected to the reactant of the next step via a conformational energy change. The hybrid meta M06-2X functional [146] is used at every stage since it has demonstrated to be suitably accurate for

the modeling of the organic molecules in biomass, as shown in Chapter 2. All reported data employs the M06-2X functional and a 6-311++G(d,p) basis set. The thermochemical data, therefore, the reaction energies and activation barriers, were calculated at 298 K and 1 atm.

4.3 Results & Discussion

4.3.1 Hemi-Acetal & Transglycosylation Formation

Figure 4.2 shows the free energy pathway of phenyl glycoside (PG) linkage formation for mannose and xylose, via the hemi-acetal mechanism. As observed in Figure 4.2, PG linkage formation is exergonic for both mannose and xylose, with reaction free energies of -54 and -33 kJ/mol, respectively. The activation free energies associated with the first two reaction steps, *i.e.*, protonation followed by hemicellulose ring opening (33 and 56 kJ/mol) as well as the formation of the PG linkage to the linear hemicellulose moiety (5 and 4 kJ/mol), are likely surmountable in physiological conditions (~ 298 K and 1 atm). However, the final reaction step, *i.e.*, dehydration followed by ring formation of the hemicellulose moiety, possesses higher activation free energies, > 100 kJ/mol. Therefore, although the formation of PG linkages is thermodynamically feasible, they are kinetically hindered via the hemi-acetal mechanism. As such, PG linkage formation in biomass may not occur via the hemi-acetal mechanism. Note the activation free energy between states M2/X2 and M3/X3 (see Figure 4.2) increases to greater than 150 kJ/mol without an acid catalyst present.

Furthermore, Figure 4.3 shows the free energy pathway of the PG linkage formation for mannose and xylose, via the transglycosylation mechanism. As observed in Figure 4.3, PG linkage formation via the transglycosylation mechanism displays very similar reaction energetics for both mannose and xylose until the final step, where the overall reaction is exergonic for mannose (-10 kJ/mol) but endergonic for xylose (6 kJ/mol). The decrease in thermodynamic feasibility compared with the hemi-acetal mechanism is a result of the change in the local environment modeled, *i.e.*, a Zundel structure for the hemi-acetal mechanism compared to acetic acid and an acetate ion for the transglycosylation mechanism. Acetic acid and an acetate ion provide less opportunities for hydrogen bonding with the reacting species, decreasing the overall thermodynamic feasibility. As such, the difference in the thermodynamics of the two reactions should not be a focal point of comparison. Activation free energies for both reaction steps, *i.e.*, the hemicellulose moiety bonding with acetate (> 130 kJ/mol) as well as the phenolic hydroxyl group of lignin bonding to the hemicellulose moiety forming the PG linkage (> 140 kJ/mol), are substantial, suggesting that the transglycosylation mechanism is also kinetically restricted. However, the transglycosylation

mechanism is assumed to be catalyzed by an enzyme, whereas only acidic and basic functional groups of an enzyme are presently modeled. Therefore, it is likely that the activation energies associated with the transglycosylation mechanism would be significantly reduced in the presence of a real enzyme [181]. However, the existence of such an enzyme for PG linkage formation remains speculative. Given the large free energy barriers and the lack of evidence for the existence of a transglycosylation enzyme specific to PG formation, further investigation is required to corroborate the relevance of the transglycosylation mechanism. As a result, the present data suggests PG formation via the transglycosylation mechanism is highly dependent on the enzyme and its associated catalytic activity.

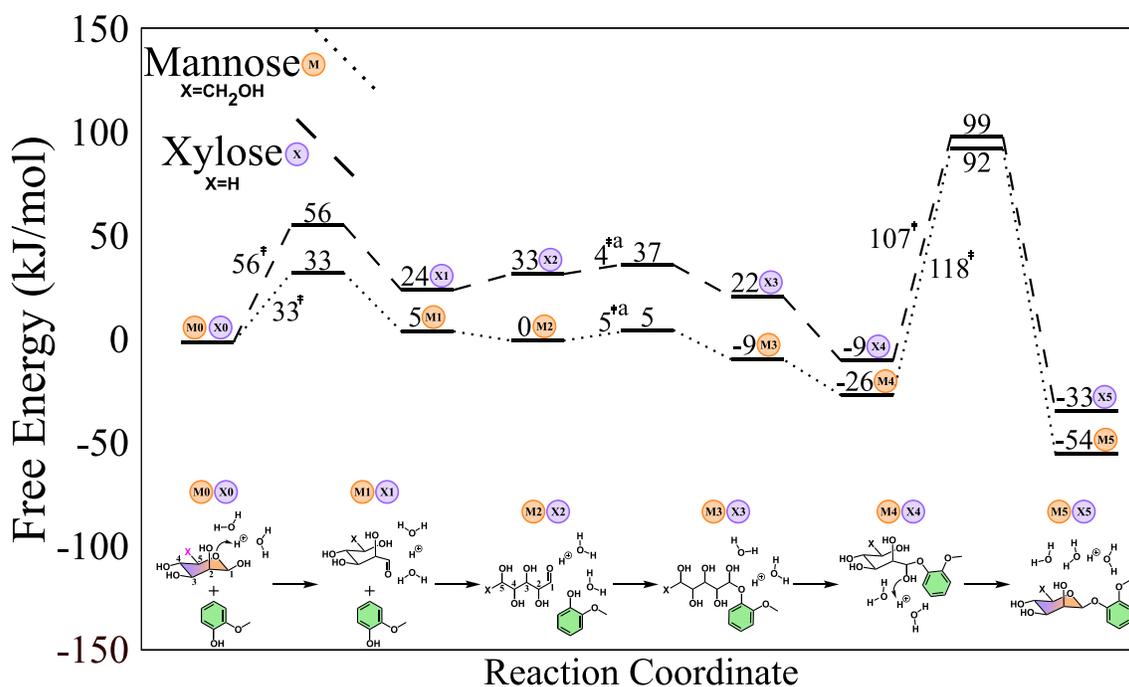


Figure 4.2: The free energy pathway (in kJ/mol) of phenyl glycoside (PG) linkage formation with mannose and xylose via the hemi-acetal mechanism. Note (a): Activation free energy increases to greater than 150 kJ/mol without an acid catalyst present.

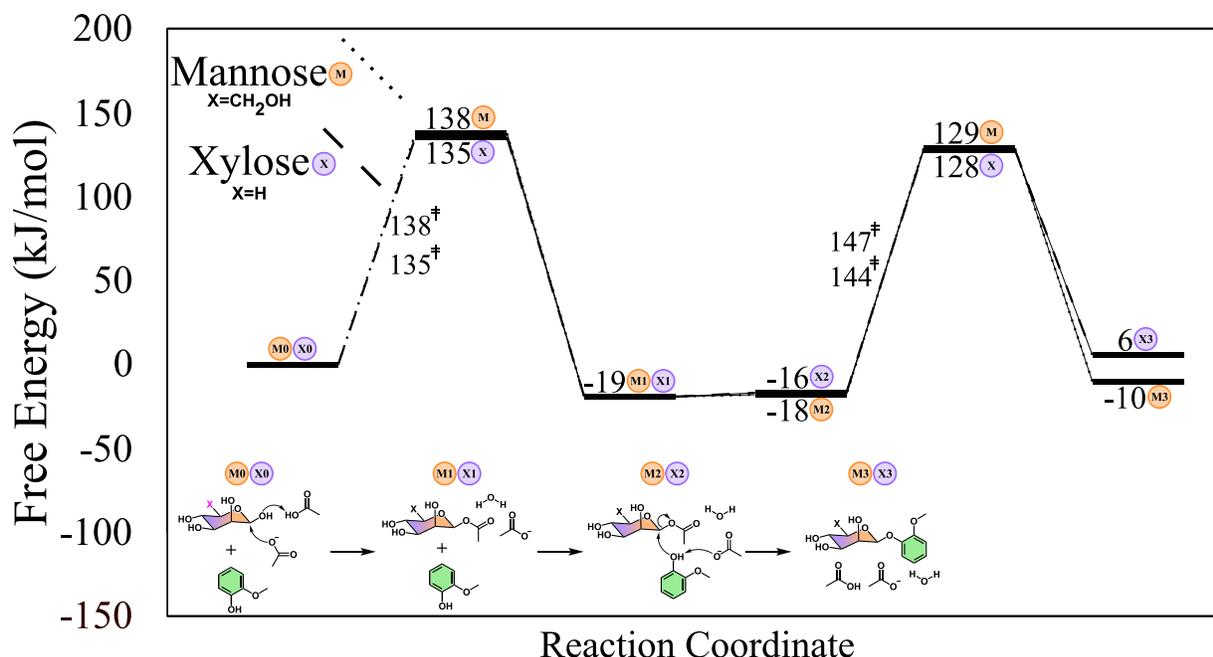


Figure 4.3: The free energy pathway (in kJ/mol) of phenyl glycoside (PG) linkage formation with mannose and xylose via the transglycosylation mechanism.

4.3.2 Electrophilic Addition of Hemicellulose

Lignin polymerization begins by the enzymatic oxidation of monolignols to generate radicals, where the oxidized monolignols undergo radical coupling to create the lignin macrostructure [45]. Due to the resonance stabilization of the radicals, a distribution of lignin linkages can form, with the β -O-4 linkage comprising the dominant linkage in all of lignin, independent of the biomass source [36, 37]. During the formation of the β -O-4 linkage, a quinone methide (QM) intermediate is formed, which requires re-aromatization via electrophilic addition at the carbonyl group (C₄ position on lignin), followed by nucleophilic addition to the α carbon, completing the synthesis of the β -O-4 lignin linkage. This mechanism as well as the carbon numbering is depicted in Figure 4.4. The electrophile is assumed to occur exclusively with a proton [32], however, theoretically any electron accepting group could act as an electrophile. The nucleophile is typically assumed to be a water molecule, adding a hydroxyl group to the α carbon. However, in Chapter 2 it was demonstrated that the hydroxyl groups on hemicellulose were kinetically and thermodynamically preferred nucleophiles [26]. In such cases, a benzyl ether LCC linkage is formed. The present work challenges the assumption of the electrophile exclusively being a free proton and explores hemicellulose as an alternative electrophile. If electrophilic addition of hemicellulose is accompanied by the dehydration of hemicellulose, a PG linkage forms. Then nucleophilic addition of water completes the re-

aromatization of the QM intermediate. Note, although electrophilic and nucleophilic addition to the QM intermediate could occur in a step-wise or concerted mechanism, in Chapter 2 it was shown that the step-wise mechanism possesses lower activation barriers compared to the concerted mechanism. Therefore, in the present Chapter, the step-wise mechanism is assumed to proceed.

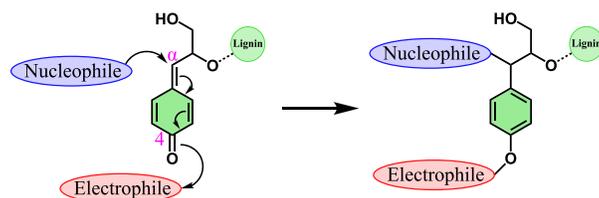


Figure 4.4: Re-aromatization of the quinone methide (QM) lignin intermediate via electrophilic addition at the carbonyl group (C_4 position) followed by nucleophilic addition at the α carbon.

Figures 4.5 and 4.6 show reaction energetics of the re-aromatization of the QM intermediate via the electrophilic addition of mannose and xylose, respectively, forming a PG linkage, followed by nucleophilic addition of water. As observed in Figures 4.5 and 4.6, the electrophilic addition to form a PG linkage is exergonic at locations C_1 and C_3 for mannose and the C_1 location for xylose. For carbon numbering, please refer to Figure 4.1, Figure 4.5 or Figure 4.6. However, the overall reaction, *i.e.*, electrophilic addition followed by the nucleophilic addition, is exergonic at locations C_1 , C_3 , C_4 and C_6 for mannose and at carbon locations C_1 , C_3 and C_4 for xylose, suggesting that PG linkage formation is thermodynamically feasible at these carbon locations.

Activation energies associated with the electrophilic addition of mannose at locations C_2 , C_3 , C_4 , and C_6 are greater than 120 kJ/mol, suggesting such pathways are kinetically restricted. However, the activation energy for the electrophilic addition of mannose at the C_1 location is minimal, 35 kJ/mol, suggesting that PG linkage formation at C_1 is kinetically facile. Furthermore, after the electrophilic addition of the C_1 location on mannose, the activation barrier for the nucleophilic addition of water to the α carbon on the lignin intermediate is also minimal, 20 kJ/mol, indicating the entire pathway is kinetically accessible at plant physiological conditions [158]. Therefore, formation of the PG linkage at the C_1 location on mannose is kinetically and thermodynamically feasible (exergonic) via re-aromatization of the QM intermediate through electrophilic addition, followed by nucleophilic addition of water.

Activation free energies associated with the electrophilic addition of xylose at locations C_2 , C_3 and C_4 are greater than 100 kJ/mol, suggesting such pathways are kinetically restricted.

Moreover, the electrophilic addition of mannose at every carbon location as well as xylose at locations C₂, C₃ and C₄ occur in a concerted mechanism, *i.e.*, the electrophilic addition of hemicellulose is accompanied by dehydration. However, the electrophilic addition of xylose at the C₁ location occurs in a step-wise mechanism, *i.e.*, dehydration of xylose occurs, forming an oxocarbenium ion, and then electrophilic addition of xylose follows, forming a PG linkage. The activation barrier for dehydrating xylose at the C₁ is 21 kJ/mol, followed by a barrier of 3 kJ/mol for the electrophilic addition of xylose and finally an activation barrier of 19 kJ/mol for the nucleophilic addition of water to the α carbon on the lignin intermediate. As a result, the entire pathway is kinetically facile. Therefore, similarly to electrophilic addition of mannose at the C₁ location, formation of the PG linkage at the C₁ location of xylose is kinetically facile and thermodynamically feasible (exergonic) via electrophilic addition followed by nucleophilic addition to the QM intermediate.

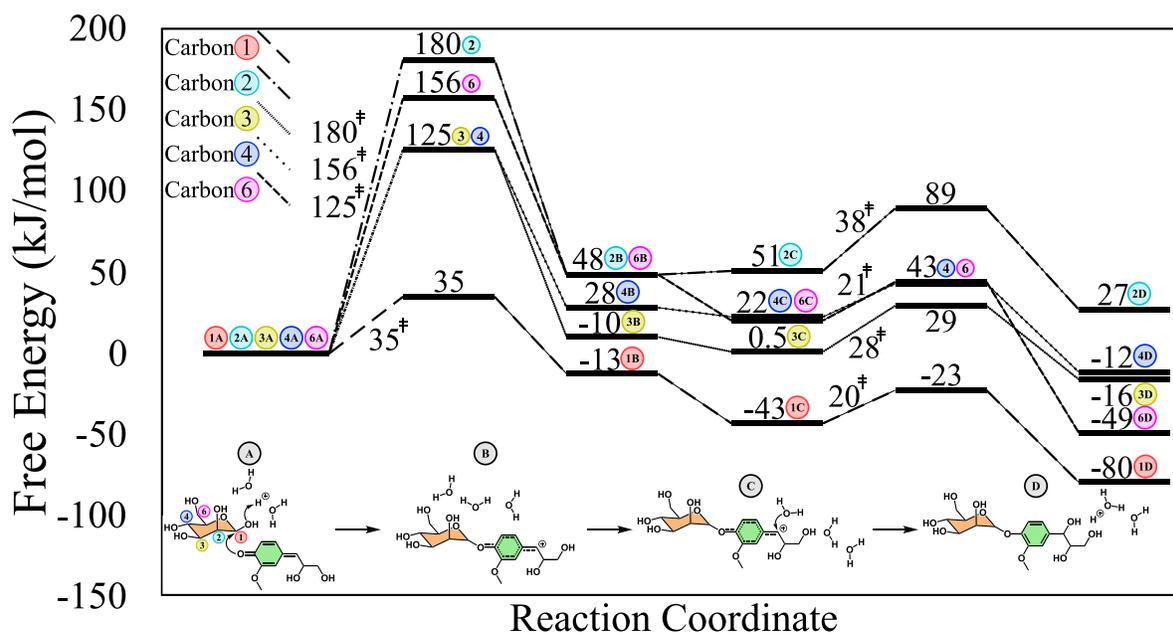


Figure 4.5: Free energy pathway (in kJ/mol) of the re-aromatization of the quinone methide (QM) intermediate via electrophilic addition of mannose at the different carbon locations, forming a phenyl glycoside (PG) linkage, followed by nucleophilic addition of water to the α carbon on the lignin intermediate.

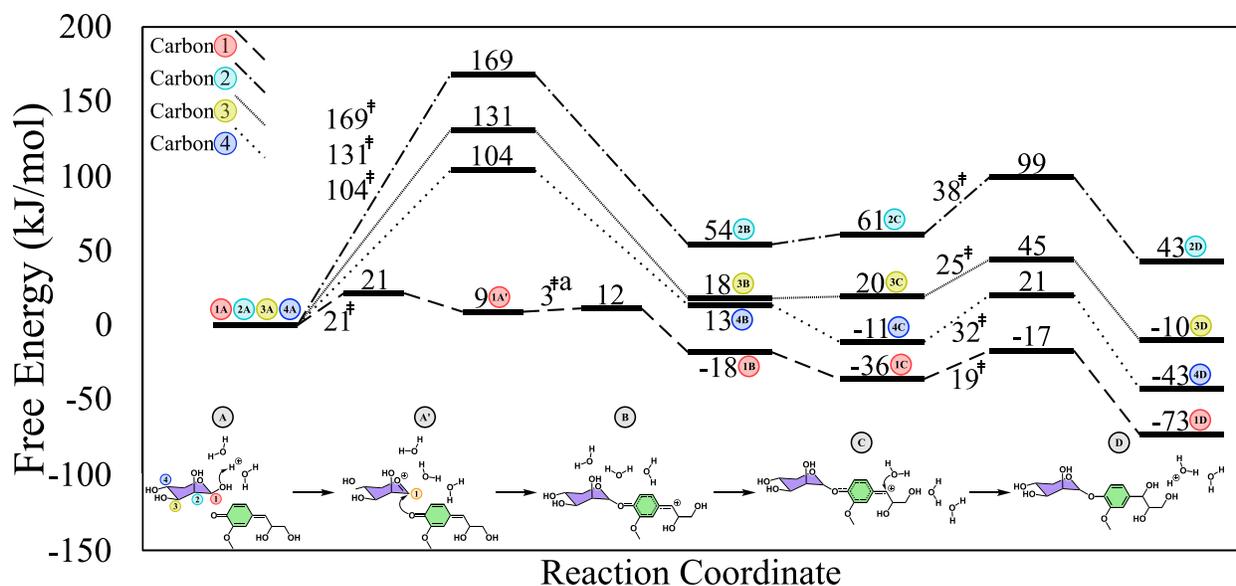


Figure 4.6: Free energy pathway (in kJ/mol) of the re-aromatization of the quinone methide (QM) intermediate via electrophilic addition of xylose at the different carbon locations, forming a phenyl glycoside (PG) linkage, followed by nucleophilic addition of water to the α carbon on the lignin intermediate. Note (a): The barrier was determined via scanning the distance between the C₁ location on xylose and the oxygen at the C₄ location on lignin. For further computational details on the calculation of this barrier please see Appendix D.

4.3.3 Competing LCC Linkages

The two previously speculated mechanisms for PG linkage formation, transglycosylation and hemi-acetal [87], are anticipated to result in a PG linkage at the C₁ location, or in other words, the reducing end of hemicellulose. Furthermore, the only kinetically facile and thermodynamically feasible location for PG linkage formation via electrophilic addition of hemicellulose followed by nucleophilic addition of water to the QM intermediate, is also at the C₁ position of hemicellulose. Therefore, to assess the relative propensity of PG linkage formation at the C₁ location on hemicellulose, other competing reactions must be evaluated. The primary competing reaction would be benzyl ether LCC formation. Benzyl ether LCC linkage formation occurs via the electrophilic addition of a proton, followed by nucleophilic addition of a hydroxyl group on hemicellulose, and in this case, the hydroxyl group at the C₁ location. If a benzyl ether linkage forms at the C₁ location of hemicellulose, then there is no recognizably feasible pathway for a PG linkage to form.

Figures 4.7 and 4.8 show a comparison of the reaction energetics of the formation of a PG LCC linkage and a benzyl ether LCC linkage with mannose and xylose, respectively, via the electrophilic and nucleophilic addition to the QM intermediate. As observed in Figures 4.7 and 4.8, benzyl ether formation begins by the barrierless electrophilic addition

of a proton, followed by a conformational change and then the nucleophilic addition of hemicellulose. Formation of a benzyl ether linkage at the C₁ location with mannose and xylose are exergonic, both with reaction free energies of -27 kJ/mol. However, PG linkage formation with mannose or xylose is thermodynamically more favorable, by ~ 50 kJ/mol, compared to benzyl ether formation with mannose or xylose. Furthermore, nucleophilic additions of mannose and xylose possess activation barriers of ~ 20 kJ/mol, indicating the formation of a benzyl ether linkage is kinetically facile, independent of the hemicellulose source. The activation energy associated with benzyl ether linkage formation with mannose (20 kJ/mol) is slightly lower compared PG linkage formation (≤ 35 kJ/mol). However, the activation energies associated with benzyl ether and PG linkage formation with xylose are nearly equivalent, ~ 20 kJ/mol. Therefore, when the hemicellulose component is mannose, kinetics are also expected to play a role in determining the reaction pathway and favor benzyl ether formation. Comparatively, when the hemicellulose component is xylose, given the similarity in activation energies; thermodynamics is expected to play the dominant role determining the reaction pathway and favor PG linkage formation. Note, the conclusions drawn regarding the kinetics of the reaction are based on the intrinsic reaction kinetics, and the actual kinetics could be governed by several other factors, *e.g.*, transport of chemical species, not considered in the present investigation.

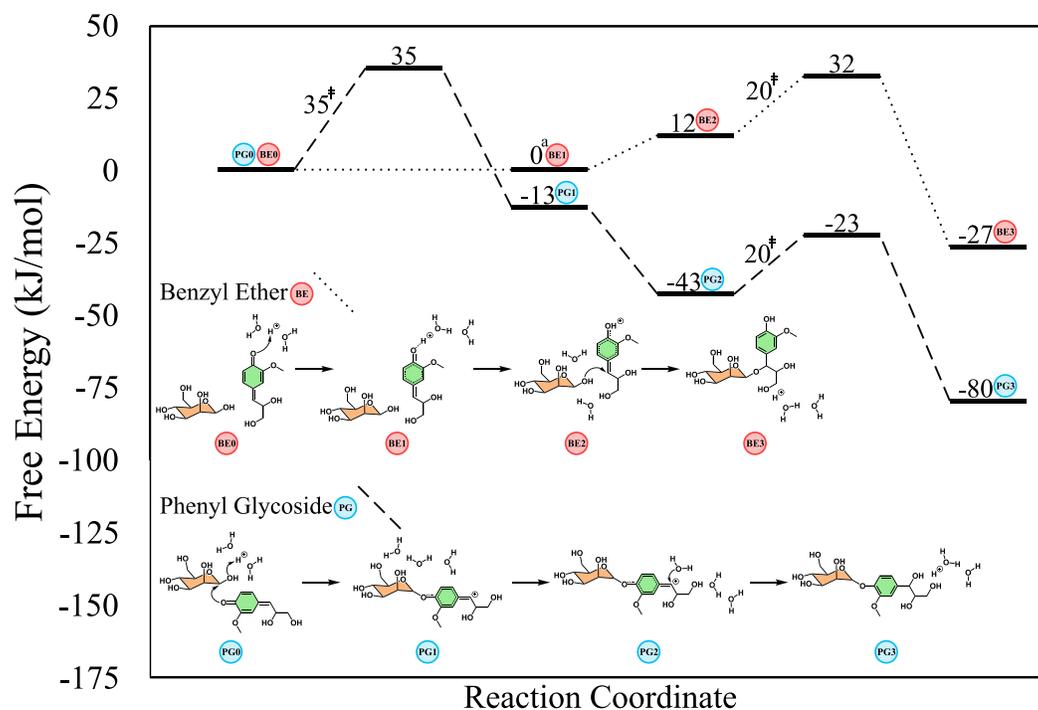


Figure 4.7: Free energy pathway (in kJ/mol) for the formation of a PG LCC linkage via electrophilic addition of mannose at the C₁ location followed by nucleophilic addition of water to the lignin QM intermediate, compared to the reaction energetics for the formation of a benzyl ether LCC linkage via electrophilic addition of a proton followed by nucleophilic addition of mannose at the C₁ location to the lignin QM intermediate. Note (a): The electrophilic addition of a free proton is a barrierless step and occurs spontaneously, therefore, the structures denoted by BE0 and BE1 are equivalent and only used for illustration purposes.

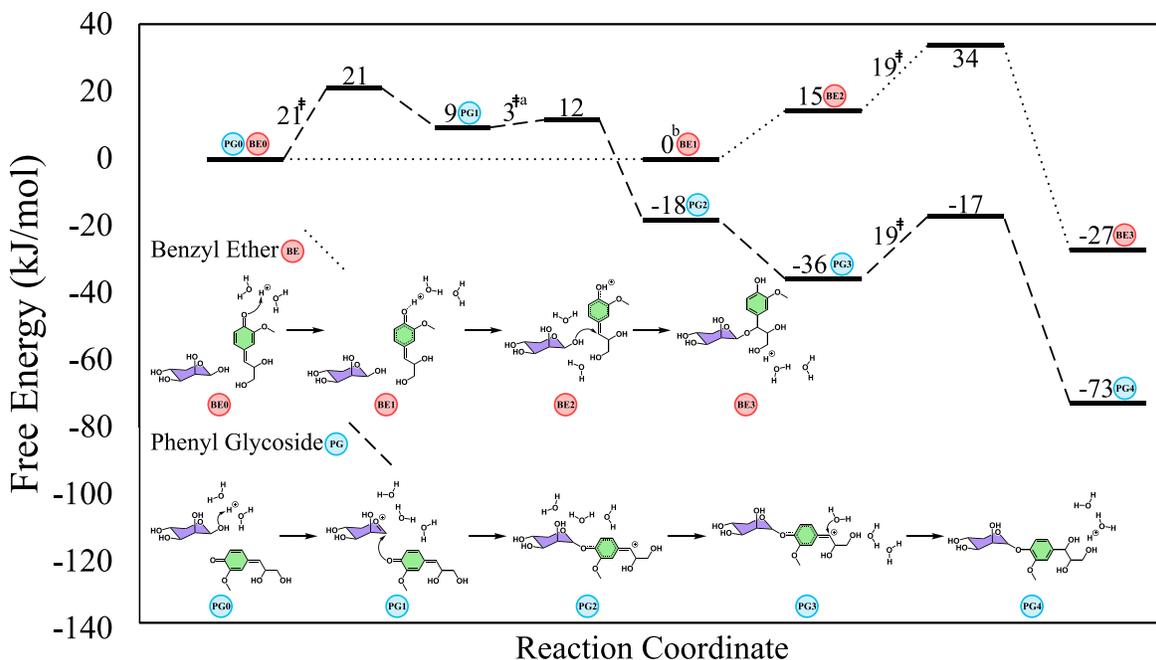


Figure 4.8: The free energy pathway (in kJ/mol) for the formation of a PG LCC linkage via electrophilic addition of xylose at the C₁ location followed by nucleophilic addition of water to the lignin QM intermediate, compared to the reaction energetics for the formation of a benzyl ether LCC linkage via electrophilic addition of a proton followed by nucleophilic addition of xylose at the C₁ location to the lignin QM intermediate. Note (a): The barrier was determined via scanning the distance between the C₁ location on xylose and the oxygen at the C₄ location on lignin. For further computational details on the calculation of this barrier please see the SI. Note (b): The electrophilic addition of a proton is a barrierless step and occurs spontaneously, therefore, the structures denoted by BE0 and BE1 are equivalent and only used for illustration purposes.

4.3.4 Deconstruction of PG LCC Linkages

Characterizing the formation of PG linkages is necessary to categorize LCC linkages based on biomass sources. However, elucidating the deconstruction mechanisms of LCC linkages is equally important to analyze their contribution to biomass recalcitrance. Figure 4.9 shows multiple reaction pathways considered during the acid catalyzed deconstruction of a PG LCC linkage moiety with mannose and xylose as well as the associated reaction energetics. Protonation of the O₁ reaction site leads to either the O₁ – C₁ bond breaking or the O₁ – C_{4'}. If the O₁ – C₁ bond breaks, mannose is separated from lignin and generates a (hydroxymethylated) 1,5-Anhydropent-2-ulose compound. If the O₁ – C_{4'} bond breaks, mannose is separated from lignin. Protonation of the C₂ and C₃ hydroxyl groups corresponds to the C₂ – OH and the C₃ – OH bonds breaking, respectively. Protonation of the C₂ hydroxyl group results in the dehydration of the hemicellulose moiety (mannose/xylose), accompanied

by ring opening and rearrangement to form a furanic species containing hydroxyl group(s) and an aldehyde group. Protonation of the C₃ hydroxyl group results in the dehydration of the hemicellulose moiety (mannose/xylose), followed by a hydride shift and subsequently deprotonation to form a carbonyl group. Please refer to Figure 4.9a for atom labelling. The activation and reaction free energies of the different deconstruction pathways must be compared to the energetics associated with breaking the PG LCC linkage, since this is the desired pathway for biomass deconstruction. Breaking the O₁ – C₁ bond corresponds to breaking the PG LCC linkage, albeit, minor chemical modifications to hemicellulose ensues. However, an acidic environment will likely protonate the generated carbonyl group on hemicellulose converting it into a hydroxyl group, resulting in a dehydrated hemicellulose compound and unaltered lignin to be recovered. Breaking the O₁ – C_{4'} bond also results in the PG linkage breaking and allows for hemicellulose as well as lignin to be recovered in chemically unaltered forms. Therefore, breaking the O₁ – C₁ or the O₁ – C_{4'} bonds are adequate in recovering the biopolymers, without diminishing the value. If the other pathways occurs, *i.e.*, breaking the C₂ – OH or the C₃ – OH bonds, chemical modifications to hemicellulose will result and the PG linkage remains intact, requiring further biopolymer degradation, thereby decreasing the biopolymer values.

As observed in Figure 4.9, the reaction free energy for cleaving the O₁ – C₁ bond and the C₃ – OH is exergonic whereas cleaving the O₁ – C_{4'} bond is endergonic. Cleaving the C₂ – OH bond is exergonic when the hemicellulose moiety is xylose, but endergonic when it is mannose. Therefore, cleaving the PG linkage is thermodynamically favorable (exergonic) via cleavage of the O₁ – C₁ bond. Hemicellulose degradation is thermodynamically favorable via cleavage of the C₃ – OH bond and, when xylose is present, cleavage of the C₂ – OH bond as well.

Furthermore, the activation free energies of the hemicellulose degradation pathways, cleaving the C₂ – OH and C₃ – OH bonds, are >150 kJ/mol, suggesting such pathways are kinetically limited and will require elevated temperatures to overcome such barriers. Cleaving the O₁ – C_{4'} bond possesses activation free energies that are substantial, > 280 kJ/mol, indicating such a pathway is thermodynamically and kinetically restricted. However, cleaving the O₁ – C₁ bond demonstrates a moderate activation free energy of ~ 95 kJ/mol. Therefore, cleaving the O₁ – C₁ bond is kinetically the most facile, while remaining exergonic. Severe biomass deconstruction techniques are often employed, *i.e.*, temperatures above 473 K, pressures greater than 1 MPa, and/or addition of acid components to assist in the deconstruction [18, 19]. In such cases, it is likely that the barriers associated with cleaving the exergonic hemicellulose degradation pathways become surmountable and thermodynamics will favor these deconstruction pathways. Therefore, breaking the PG LCC linkage is kinetically and

thermodynamically favored in mild acid catalyzed deconstruction. For the temperature dependence of the activation barriers as well as the reaction free energies for each pathway, please refer to Appendix D.

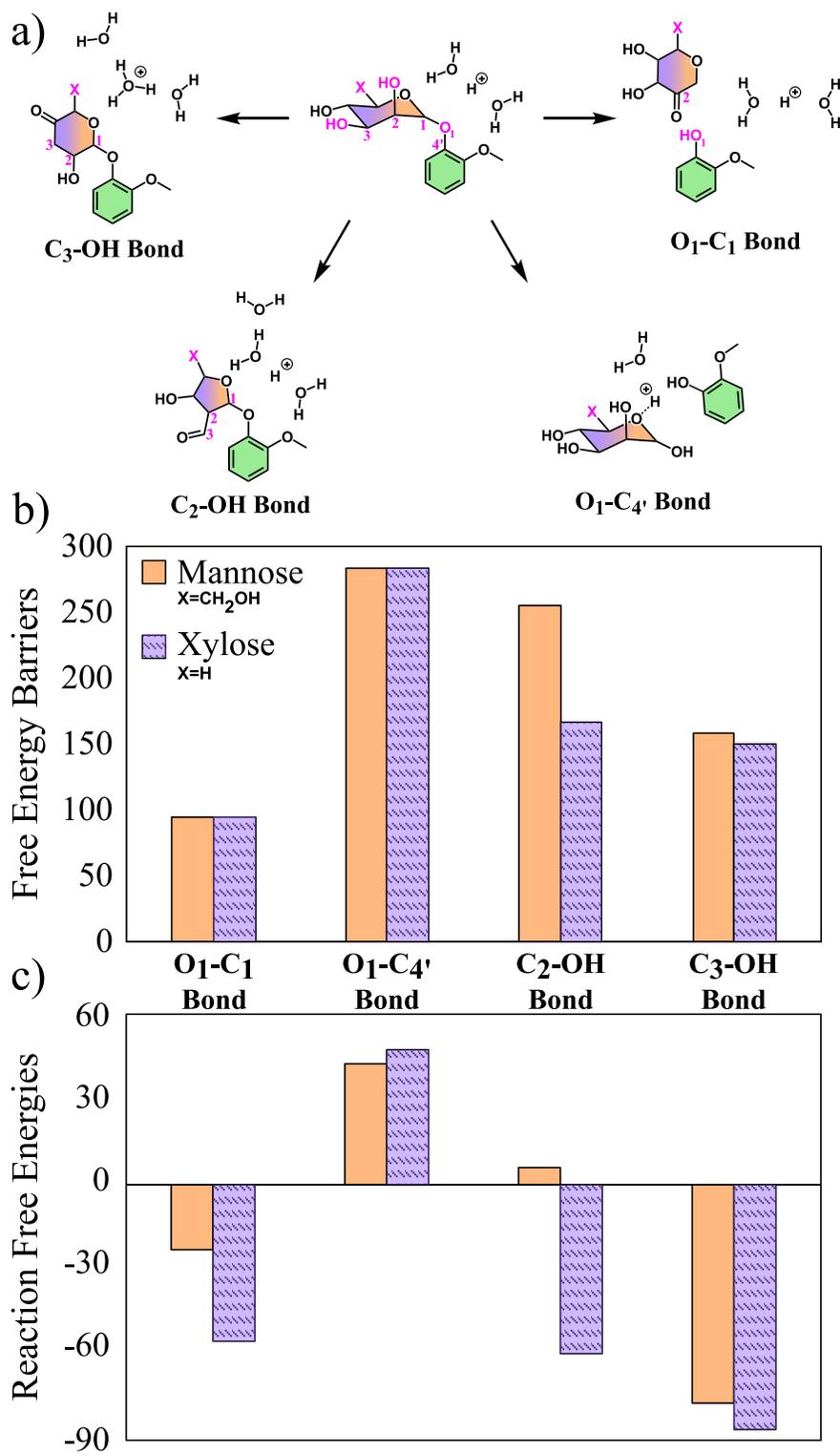


Figure 4.9: (a) The reaction pathways considered in the acid catalyzed deconstruction of the PG LCC linkage with mannose and xylose. (b) Activation free energies (in kJ/mol) and (c) reaction free energies (in kJ/mol) of the reaction pathways illustrated in (a).

4.3.5 Discussion

Although the presence of phenyl glycoside (PG) LCC linkages throughout various biomass sources has garnered increased evidence [171, 89, 172, 173, 174, 175, 86, 176, 177], their formation mechanism remains largely unknown [86, 87, 103]. The present investigation elucidates mechanistic details of the reaction and energetics associated with different pathways of PG formation. It was determined that the two previously proposed mechanisms, hemiacetal formation and transglycosylation, are both associated with substantial free energy barriers, suggesting such mechanisms are unlikely to be responsible for the significant fraction of PG linkages observed. The presently proposed mechanism, electrophilic addition of hemicellulose to a lignin quinone methide (QM) intermediate, possesses facile kinetics and is exergonic at the C₁ position, *i.e.*, at the reducing end of hemicellulose. Although, benzyl ether LCC formation demonstrates a competing reaction at the C₁ position, PG linkage formation is favored when the hemicellulose moiety is xylose whereas benzyl ether formation is favored when the moiety is mannose. As a result, it is anticipated that hemicellulose composed of xylans will have higher fractions of PG LCC linkages whereas hemicellulose composed of mannans will have higher fractions of benzyl ether LCC linkages. These results are also supported by experimental observations.

Balakshin *et al.* [171] isolated a concentrated fraction of LCC linkages from softwood (loblolly pine) and was able to detect PG linkages at the C₁ position position of the carbohydrates using 2D ¹H – ¹³C correlation nuclear magnetic resonance (NMR) techniques. Later they [89] quantified LCC linkages in hardwoods (white birch) and softwoods (loblolly pine) by implementing a combination of quantitative 2D heteronuclear single quantum coherence (HSQC) and ¹³C NMR spectroscopic techniques. A group of signals in the HSQC spectra were attributed to PG linkages at the C₁ position of the carbohydrates, and it was determined that birch wood contained higher fractions of PG linkages compared to that of pinewood. Yuan *et al.* [172] characterized LCC linkages within hardwood (poplar) using a combination of quantitative ¹³C and 2D HSQC NMR techniques and the results showed that xylan was the main carbohydrate associated with lignin and PG linkages were identified at the C₁ position of the carbohydrates. You *et al.* [173] isolated LCC fractions from *Arundo donax*. Fourier-transform infrared spectroscopy (FT-IR) revealed that the LCC fractions were xylan rich and the 2D HSQC NMR spectroscopy quantified a substantial fraction of PG linkages at the C₁ position of the carbohydrates. Yao *et al.* [174] injected ¹³C isotope-labeled xylose into living wheat straw and proceeded to extract LCCs and characterize them using FT-IR, sugar composition, molecular weight analysis, ¹³C-NMR and HSQC. The results showed that PG linkages were connected to the C₁ position of the carbohydrates. Furthermore, the isotope-labeled LCC fraction (¹³C-LCC) contained three times more PG linkages compared

to the non-labelled LCC fractions indicating that part of lignin and xylan in wheat straw were connected by PG linkages. Zhang *et al.* [175] harvested bamboo culms at 2, 4 and 6 months of development to assess the chemical variations in LCCs during plant growth. 2D-HSQC NMR revealed PG linkages at the C₁ position of the carbohydrates and an increase in PG linkages with maturation of the bamboo. Furthermore, sugar composition revealed a dramatic increase of xylose in hemicellulose with the maturation of the bamboo. As a result, it was suggested that PG linkages were primarily between lignin and xylan. Giummarella and Lawoko [86] applied a mild protocol for isolating LCCs from hardwoods and conducted a structural analysis using 2D HSQC NMR spectroscopy, ³¹P NMR spectroscopy, as well as thioacidolysis in conjunction with gas chromatography mass spectrometry (GC-MS) and GC with flame ionization detection. Evidence suggested that PG linkages occurred at the reducing end of xylan, and acetylation on xylan regulates the frequency of PG linkage. Del Rio *et al.* [176] used 2D HSQC NMR to identify PG linkages between lignin and xylan in herbaceous plants (sisal and abaca). Lv *et al.* [177] analyzed bamboo using FT-IR, 2D HSQC NMR, as well as ³¹P NMR techniques and revealed that PG linkages between lignin and the C₁ position of xylan were the predominant type of LCC linkage.

The previous literature has provided overwhelming evidence to indicate that the C₁ position on carbohydrates is the site of the PG linkage location, and xylose-based hemicellulose is predominantly participating in PG linkages with lignin, in excellent agreement with the present computational results. It has been proposed that benzyl ether linkages are dominant in softwood whereas PG linkages are dominant in hardwood [44], however, the reasoning behind such a phenomena has never been explained before. Given that softwoods are predominately composed of mannans whereas hardwoods are composed of xylans [31, 32], the present computational results provide the rationale behind such an observation. The reaction energetics of LCC formation at the C₁ position thermodynamically favours PG formation when hemicellulose is composed of xylose whereas it kinetically favors benzyl ether formation when hemicellulose is composed of mannose. Therefore, the reaction energetics from the presently proposed mechanism for PG linkage formation, *i.e.*, electrophilic addition of hemicellulose to a lignin QM intermediate, fits exceptionally well into our current understanding of PG linkages and suggests it is a likely mechanism responsible for the formation of PG linkages observed.

Moreover, the deconstruction reaction energetics of PG linkages suggests that cleaving PG linkages is kinetically facile with free energy barrier of ~ 95 kJ/mol, in the acid catalyzed deconstruction. In comparison, the free energy barriers associated with hemicellulose degradation deconstruction pathways are > 150 kJ/mol, while remaining exergonic. Comparatively, cleavage of the benzyl ether LCC linkages are endergonic at deconstruction conditions,

whereas the pathways associated with the competing hemicellulose degradation reaction are exergonic, despite relatively larger activation energies, > 130 kJ/mol, as shown in Chapter 3. The reaction free energies of the benzyl ether deconstruction pathways (LCC cleavage and hemicellulose degradation) displayed an inverse correlation with temperature, indicating as the temperature increases, the reaction energies will become more favorable, characteristic of most dissociation reactions. It was anticipated that increasing the temperature of such a system would likely favor overcoming the large barriers associated with hemicellulose degradation before cleaving benzyl ether LCC linkages become thermodynamically facile. This suggests that PG linkages are unlikely to be significantly contributing to biomass recalcitrance, rather benzyl ether linkages are likely the LCC linkages contributing to a much greater extent. Zhang *et al.* [182] investigated the effects of dilute acid flowthrough treatment on poplar (hardwood) and pine (softwood) at temperatures ranging from 473 to 543 K. Under the same treatment conditions 20-40% more hardwood lignin was removed compared to softwood lignin. Lehto *et al.* [183] performed dilute sulfuric acid treatments on birch (hardwood) and pine (softwood) chips in varying treatment conditions, *i.e.*, temperatures of 403 K or 423 K and treatment times ranging from 30 min to 120 min. At 403 K less than 10% of the wood components were dissolved for both samples. However, at 423 K the hardwood sample showed increased dissolution relative to the softwood sample at each of the treatment times, with a total of $\sim 19\%$ hardwood dissolved compared to only $\sim 14\%$ softwood at 120 min. Kundu *et al.* [184] used peracetic acid to treat victorian ash (hardwood) and pine (softwood) species at a temperature of 363 K and treatment times ranging from 30 min to 5 hours. The hardwood demonstrated much faster delignification rates compared to softwood. Since benzyl ether linkages are dominant in softwood whereas PG linkages are dominant in hardwood [44], the present computational data suggests that the increased delignification and dissolution of the hardwood samples during acidic treatment is a result of the larger fraction of labile PG linkages in hardwood, as compared to softwood.

4.4 Conclusions

Developing a mild deconstruction technology that can overcome biomass recalcitrance while maintaining an atom efficient recovery of cellulose, hemicellulose and lignin remains a central challenge for the economical production of biobased chemicals. Although LCCs have been recognized to play a key role in the refractory nature of biomass, experimental characterization of LCCs is limited, leading to a slow development in chemical/biochemical deconstruction methods. As such, the present work investigates mechanistic details for the formation

and deconstruction of the predominant PG LCC linkage in lignocellulosic biomass. Two mechanisms for PG linkage formation, hemiacetal formation and transglycosylation, were determined to be associated with substantial activation barriers indicating both mechanisms are kinetically hindered. However, the third mechanism of electrophilic addition of hemicellulose to a lignin QM intermediate, possesses facile kinetics. Furthermore, the reducing end is the only location on hemicellulose that the electrophilic addition to a lignin QM intermediate possesses favorable reaction energetics, in agreement with the previous experimental literature that has established that PG linkages are located at the reducing end. However, benzyl ether LCC linkage formation demonstrates a competing reaction at the reducing end of hemicellulose. When the hemicellulose moiety is composed of mannose, benzyl ether formation is expected to be kinetically favored at the reducing end. Whereas, for xylose, PG linkage is expected to be thermodynamically favored at the reducing end. As a result, it is anticipated that hemicellulose composed of xylans will have higher fractions of PG LCC linkages whereas hemicellulose composed of mannans will have higher fractions of benzyl ether LCC linkages. Such a result provides an explanation for the experimental literature/characterization observation that benzyl ether linkages are dominant in softwood, which have mannan based hemicellulose, whereas PG linkages are dominant in hardwood, which have xylan based hemicellulose. Additionally, this work quantifies the reaction mechanisms, kinetics and thermodynamics associated with the deconstruction of the PG LCC linkages in biomass under acidic conditions. The primary competing reaction pathways were the chemical degradation of the hemicellulose moieties, which were also investigated. The deconstruction energetics demonstrate that breaking PG linkages is kinetically and thermodynamically favored in acid catalyzed deconstruction, suggesting that PG linkages are unlikely to be significantly contributing to biomass recalcitrance.

Chapter 5

Branching in the Lignin Structure

5.1 Introduction

Despite lignin's importance and relevance in atom-efficient utilization of lignocellulosic biomass as a sustainable carbon source [5, 185], lignin chemistry has faced a fundamental impediment: quantifying its structure. The structure of lignin remains elusive, largely due to its complex synthesis in native biomass [36]. As a result, discrepancies pertaining to the lignin structure still exist throughout the literature, with one of the most polarizing debates being the extent of branching in the structure [35, 33, 34, 186]. Resolving the structural discrepancies requires quantifying the types and frequency of the linkages that give rise to the macroscopic structural features observed.

First-principles based computational studies have been used to compute the energetics of the various radical coupling pathways available in lignin [62, 63], while Monte Carlo based algorithms have been developed to model lignin polymerization *in silico* [64, 65]. However, majority of computational investigations [62, 63, 64, 65] have confined themselves to exploring well accepted reaction pathways established for lignin synthesis. Although this approach is necessary to validate conclusions and solidify fundamental knowledge, it is clear from the elusive lignin structure that exploring beyond direct experimental evidence is essential to characterize lignin linkages and resolve the remaining structural controversies.

Recently, the formation of lignin carbohydrate complexes (LCCs) have garnered attention in playing a key role in the recalcitrant nature of biomass [85, 42]. LCCs arise because of a subtle side reaction during the formation of the predominant β -O-4 linkage [36]. When forming the β -O-4 linkage in lignin, a quinone methide (QM) intermediate is formed, which

requires re-aromatization via nucleophilic addition at the α carbon. Traditionally, the nucleophile was implicitly assumed to be water [36, 63, 64], however, the nucleophilic addition of cellulose/hemicellulose leads to benzyl ether LCCs, as shown in Chapter 2. However, formation of LCCs only represent a small fraction of the reaction pathways available for the QM intermediate. The reaction site of the QM intermediate would primarily be exposed to other monolignols, which would give rise to unrecognized lignin linkages and structurally complex polymers.

As such, the present Chapter uses density functional theory (DFT) calculations to quantify the reaction energetics associated with the nucleophilic addition of the alcohol groups in monolignols, the γ -OH and the 4-OH, to the QM intermediate. All references to the different linkage sites will follow the numbering used in Fig. 5.1.¹

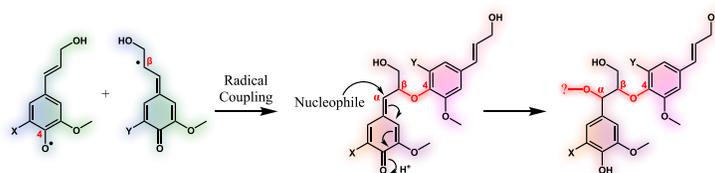


Figure 5.1: The reaction mechanism to synthesize the β -O-4 linkage as well as the re-aromatization of the QM intermediate that leads to various α -O-H/ α -O-4/ α -O- γ based structures. Note, “X” and “Y” either represent a H group or an O – CH₃ group.

5.2 Computational Methodology

5.2.1 Model Compounds

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% [38].

¹Chapter 5 employs a new color scheme to easily distinguish different molecular structures.

5.2.2 Computational Procedure

All-electron DFT calculations were performed using the Gaussian 09 code [145]. All data reported is at the M06-2X/6-311++G(d,p) level of theory [146] as it has demonstrated accurate modeling of the carbohydrates present in lignocellulosic biomass, as shown in Chapter 2. 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. In Chapter 2 the nucleophilic addition to the QM intermediate was investigated to proceed via a step-wise and concerted mechanism and it was concluded that a step-wise mechanism was kinetically favoured. Therefore, a step-wise mechanism is considered presently. The step-wise mechanism assumes the phenolic oxygen at the 4 position in the QM intermediate is protonated in a barrier-less step [108] and the α carbon proceeds to undergo nucleophilic addition. The Berny algorithm was used for the transition state (TS) search [148, 149]. 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 [108]. 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 using the integral equation formalism variant 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 ($\epsilon=78.3553$) since lignin polymerization is understood to occur in an aqueous medium [36]. 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 5.2. 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.

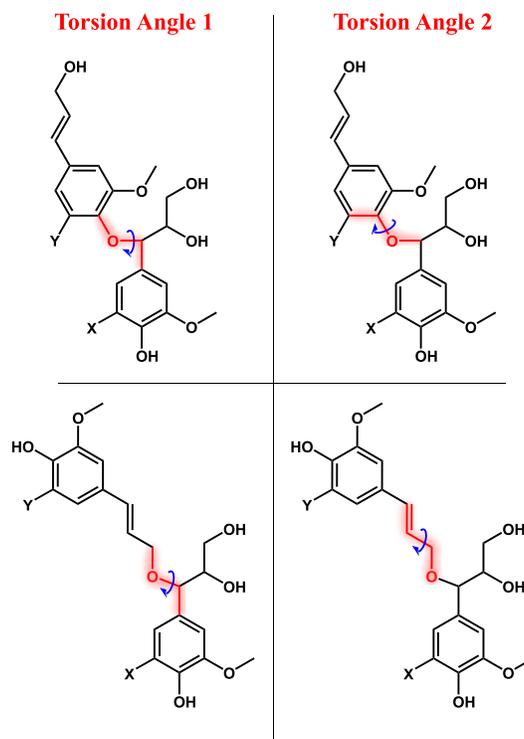


Figure 5.2: The schematic of the torsion angles systematically rotated to generate an approximate potential energy landscape of the product compounds.

5.3 Results & Discussion

5.3.1 Thermochemistry of α -O-4 & α -O- γ Lignin Linkages

Fig. 5.3a) shows the nucleophilic addition of coniferyl alcohol, sinapyl alcohol and water to a guaiacyl and syringyl QM intermediate, as well as the reaction free energies and activation free energy barriers associated with each reaction. Fig. 5.3b) shows the reaction free energies of forming an α -O-4 and α -O- γ lignin linkages, relative to the nucleophilic addition of water. Additionally, the most favorable reaction free energy (among all cellulose/hemicellulose model compounds) reported for the formation of LCC linkages via nucleophilic addition to the QM intermediate calculated in Chapter 2, is also provided for comparison. As lignin polymerization occurs in a mildly acidic, aqueous environment, water has been assumed to exclusively react with the QM intermediate [36, 157, 54]. Therefore, the reaction energetics of forming an α -O-4 and α -O- γ lignin linkages must be compared to the nucleophilic addition of water, forming an α -OH group, to establish the relative feasibility of the two lignin linkages. As lignin is a hydrophobic polymer [187], the interplay between the monolignols' transport and reaction rates would determine if the reaction happens in a water-rich or water-lean environment. To assist in discerning the local reaction environment, an implicit solvent is

incorporated.

The formation of an α -O- γ lignin linkage is 3.3–29.8 kJ/mol more exergonic than the nucleophilic addition of water to form an α -OH. The formation of an α -O-4 lignin linkage has relative reaction free energies within ~ 7 kJ/mol to the formation of an α -OH group. The formation of an α -O- γ lignin linkage is thermodynamically favoured compared to the formation of an α -O-4 linkage, an α -OH group and LCC linkages. Whereas, forming an α -O-4 lignin linkage demonstrates either a neutral or unfavorable thermodynamic feasibility, depending on the monolignol composition, compared to the formation of an α -OH group and of LCC linkages. The present results demonstrate that the formation of an α -O- γ lignin linkage is thermodynamically preferred over the formation of an α -O-4 linkage and an α -OH group, independent of the local environment and monolignol composition. 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 $\pi - \pi$ interactions, reducing the thermodynamic stability, compared to what is observed with the α -O- γ lignin linkages. 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. Moreover, 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 $\pi - \pi$ interactions, providing stability to the product compounds and more favorable reaction free energies.

5.3.2 Kinetics of α -O-4 & α -O- γ Lignin Linkages

Fig. 5.3a) shows the activation barriers for the nucleophilic addition at the α position of the QM intermediate. The activation energies for the α -O- γ and α -O-4 lignin linkage formation range from 11.3 to 36.5 kJ/mol, while the activation energies for the nucleophilic addition of water, resulting in an α -OH group, range from 28.8 to 43.5 kJ/mol. With all activation energies below 45 kJ/mol, such barriers are easily surmountable at temperatures corresponding to plant growth [158], indicating facile kinetics for all reactions. However, such conclusions are based on intrinsic reaction kinetics, and actual kinetics may be governed by other factors.

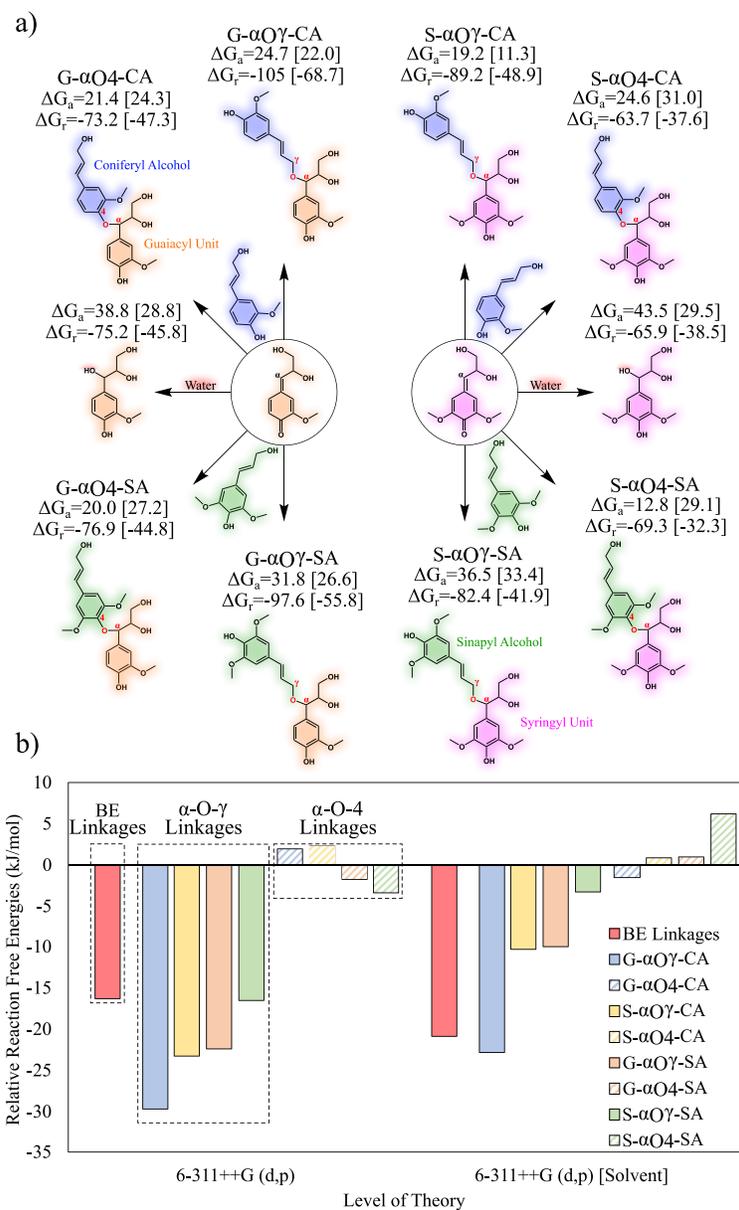


Figure 5.3: The nucleophilic addition of coniferyl alcohol (CA), sinapyl alcohol (SA) and water to a guaiacyl (G) and syringyl (S) QM intermediate. The nucleophilic addition of the γ -OH as well as the 4-OH of coniferyl and sinapyl alcohol are considered. (a) Each reaction is labeled with its reaction free energy (ΔG_r , kJ/mol) and activation free energy (ΔG_a , kJ/mol) with the labels corresponding to the legend in (b). Un-bracketed numbers indicate gas phase calculations and bracketed numbers indicate implicit solvent calculations. (b) Reaction free energies leading to $\alpha\text{-O-}4$ and $\alpha\text{-O-}\gamma$ lignin linkages, relative to the nucleophilic addition of water to each respective monolignol intermediate. The reaction free energies reported for the formation of benzyl ether (BE) LCC linkages as well as the nucleophilic addition of water to the guaiacyl QM intermediate, is taken from Chapter 2. Although the reaction free energy of BE LCC linkage formation depends on the composition of nucleophile, the most favored BE linkage site is selected.

5.3.3 Discussion

The α -O-4 linkage is associated with controversy. Some investigations indicate its presence [35, 186], while others suggest it is absent in plants [38]. The mechanistic evidence for such a linkage has not been well established, therefore, the formation of the α -O-4 is investigated via nucleophilic addition at the 4 position of a monolignol to the α position in a QM monolignol intermediate (*cf.* Figure 5.1). Although the α -O-4 linkage formation is associated with a negative reaction free energy, its formation is thermodynamically less favored compared to the other nucleophiles that are capable of causing the re-aromatization of the QM intermediate, *i.e.*, water, cellulose/hemicellulose [32] and the γ -OH position of monolignols. Therefore, the present work suggests the α -O-4 lignin linkage is less likely to form via nucleophilic addition to the QM intermediate and there remains no substantial evidence, from the lignin polymerization point of view, corroborating the existence of such a linkage. The α -O- γ linkage represents a step beyond our current understanding of the lignin structure. The formation of the α -O- γ linkage possesses the greatest thermodynamic favorability (lowest reaction free energy) compared to any other linkage identified presently or previously, via nucleophilic addition to the QM monolignol intermediate. The favored reaction free energy accompanied with the minimal activation barrier indicates that the α -O- γ linkage could be present in non-negligible amounts in the native lignin structure. Furthermore, the α -O- γ linkage has the added implication that it demonstrates a branching point in the lignin structure, as it would readily undergo dehydrogenation during lignin polymerization.

Capanema *et al.* [188] used 2D heteronuclear multiple quantum coherence (HMQC) nuclear magnetic resonance (NMR) and quantitative NMR to fully characterize spruce wood lignin. Significant fractions of γ -O-alkyl and α -O-alkyl were identified from the NMR data; however, the exact structure of the ether moieties was not known. Balakshin *et al.* [35] reviewed the different methods used for the quantification of lignin branching and contested the assertion that lignin is linear [33, 34]. Although 4 independent experimental methods (combinations of wet chemistry and NMR) confirmed that the two previously recognized branching points, etherified 5-5' and etherified 4-O-5' lignin linkages were present, these linkages could not explain the entirety of branching/crosslinking in the lignin structure. It is concluded that the additional branching must come from the 20% of unidentified side chain structures. Quantitative NMR analyses suggest that branches involve different aliphatic ether types at the α and γ positions, with an intact β -O-4 linkage. However, the difficulty in assigning these structures arise as a result of the small fraction of linkages present (potentially due to the labile nature in ball milling [186, 189, 87]), high structural diversity of the linkages and signal overlap in (2D) NMR spectrums [35]. Therefore, though the experimental evidence corroborating the existence of the α -O- γ lignin linkage has been established, the challenge

was to rationalize the experimental results and it required taking a step outside our current understanding of lignin polymerization and consequently, the lignin structure. Thus, the present investigation provides evidence that the predominant β -O-4 lignin linkage can act as a branching point via the formation of an α -O- γ linkage.

5.4 Conclusions

Despite multiple attempts to characterize lignin linkages, our understanding of the macroscopic structural implications of such linkages is still developing. Recently, the extent of branching in the lignin structure has been a subject of debate, with some advocating for its linearity while others for its branched features. The present work explores alternative reaction pathways during lignin polymerization, which can lead to α -O- γ and α -O-4 linkages, to help clarify such structural discrepancies. The formation of the predominant β -O-4 linkage in lignin requires the re-aromatization of a reactive QM intermediate via nucleophilic addition. Of the available nucleophiles to re-aromatize the QM intermediate, the thermodynamically and kinetically favored pathway is the nucleophilic addition of monolignols at the γ hydroxyl group, forming a α -O- γ lignin linkage. The α -O- γ linkage has the added implication that it acts as a branching point in the lignin structure, providing supporting evidence for the branched nature of lignin.

Chapter 6

Conclusions

Efficient extraction of cellulose, hemicellulose and lignin within lignocellulosic biomass is central to the economics of its transformation to chemicals and materials. However, there is very limited understanding of the physico-chemical interactions within these biopolymers, that give rise to the recalcitrant nature of biomass. Hence, current biomass deconstruction techniques resort to implementing harsh chemicals and high temperature and pressure, resulting in annihilation of a portion of the feedstock and causing significant carbon loss. Covalent linkages between lignin and cellulose/hemicellulose, commonly referred to as LCCs, have been identified to directly correlate with biomass recalcitrance, however, the existence and chemical nature of these LCC linkages remain controversial and speculative, primarily due to the harsh experimental techniques implemented to characterize them. As a result, the formation mechanisms of LCCs as well as their potential deconstruction pathways and chemistry remain largely unknown. Since experimental techniques can be cumbersome in elucidating the molecular level details of the interactions between biopolymers, computational methods can be used to complement and accelerate our current understanding. Hence, in this thesis, the kinetics and thermodynamics of the reaction pathways resulting in the formation of the various types of LCC linkages (benzyl ether, benzyl ester and PG) are computed using first-principles quantum mechanical methods. For the stable LCC linkages that are anticipated to contribute to the biomass' recalcitrance the most, reaction mechanisms and energetics of the various deconstruction pathways, including the competing reactions, are also computed. Independent of the source of biomass, the predominate lignin linkage is the β -O-4 linkage, which requires the re-aromatization of a QM intermediate via electrophilic addition (reaction site one) followed by nucleophilic addition (reaction site two). The electrophile is commonly assumed to be a proton from the acidic conditions of the plant cell wall and the nucleophile is assumed to be water. In such a case, both reaction sites would be terminated, and lignin would only interact physically with cellulose/hemicellulose. The present work explores the electrophilic and nucleophilic addition of hemicellulose, both of which would lead to LCC

linkages as well as the nucleophilic addition of other monoglinols, leading to novel lignin linkages and the nucleophilic addition of water, to evaluate the relative feasibility of the reactions.

The thermodynamics and kinetics of forming benzyl ether and benzyl ester linkages via nucleophilic addition of hemicellulose to the QM intermediate was quantified. The formation of benzyl ether and benzyl ester LCC linkages showed thermodynamic favorability, while remaining kinetically facile, compared to the traditionally assumed mechanism of the QM intermediate reacting with water. However, formation of the benzyl ester linkage is kinetically less competitive compared to the formation of benzyl ether linkages, suggesting the benzyl ester linkages are not abundant *in vivo*. The favorable reaction energetics of benzyl ether linkages across the variety of hemicellulose compositions suggest that these LCC linkages are likely abundant. Furthermore, the computed reaction energetics indicate that lignin reacting with the primary alcohol group in mannose is one of the most thermodynamically favorable reaction sites. This is in agreement with previous experimental work that has provided direct evidence that the primary alcohol group in mannose is the reaction site that forms the benzyl ether LCC linkage.

As a result of the favorable reaction energetics for benzyl ether linkage formation, the reaction mechanisms, kinetics and thermodynamics associated with the deconstruction of the benzyl ether LCC linkages in biomass under acidic conditions is computed. The competing reactions are also computed to ensure the desired deconstruction pathway, *i.e.*, cleaving the LCC linkage, can be compared to all deconstruction pathways energetically accessible. As such, the possible reaction pathways identified include degrading the lignin structure, degrading the hemicellulose structure, or cleaving the LCC linkage. The degradation of the lignin moiety was highly endergonic with large kinetic barriers, suggesting such pathways may not be prevalent during deconstruction. Cleaving the LCC linkages was moderately endergonic and possessed the lowest kinetic barriers. Degradation of the hemicellulose moiety was exergonic or mildly endergonic and possessed large kinetic barriers. Computing the reaction energetics as a function of temperature demonstrated that increasing the temperature of the system would likely favor overcoming the large barriers associated with hemicellulose degradation before an exergonic regime could be established for cleaving the LCC linkages. Therefore, increasing the severity of deconstruction conditions to cleave LCC linkages will inevitably lead to overcoming the large barriers associated with the degradation of hemicellulose. Previous experimental work has established that at a threshold temperature/severity, hemicellulose will begin to degrade, decreasing the yield. Such results demonstrates the temperature sensitivity hemicellulose degradation possesses, which is in agreement with the present computational results.

The reaction mechanisms, kinetics and thermodynamics associated with the formation and deconstruction of the PG LCC linkage are computed. The two previously proposed mechanisms for the formation of PG linkages, hemiacetal and transglycosylation, are associated with significant activation barriers, suggesting these pathways are kinetically limited. A new mechanism is proposed, the electrophilic addition of hemicellulose to a lignin QM intermediate, that possesses facile kinetics and is exergonic, suggesting it could be the pathway responsible for the significant fraction of PG linkages observed in native biomass. Moreover, PG formation showed a composition dependence, suggesting that xylans will have higher fractions of PG linkages compared to mannans, explaining why softwoods and hardwoods have different reported types of LCCs. Additionally, the reaction mechanisms, kinetics and thermodynamics associated with the deconstruction of the PG LCC linkages in biomass under acidic conditions are investigated. The chemical degradation of the hemicellulose moieties is the primary competing reaction; however, the deconstruction energetics demonstrate that breaking PG linkages is kinetically and thermodynamically favored in acid catalyzed deconstruction. This indicates that PG linkages may not significantly contribute to the biomass recalcitrance. Previous experimental work has demonstrated that during acidic treatment hardwood displays faster delignification rates compared to softwood. The present computational results suggests that the increased delignification and dissolution of the hardwood samples is a result of the larger fraction of labile PG linkages in hardwood, as compared to softwood.

The thermodynamics and kinetics of forming α -O- γ and α -O-4 lignin linkages via nucleophilic addition of monolignols to the QM intermediate was computed. Although the α -O-4 linkage formation is associated with a negative reaction free energy, its formation is not thermodynamically competitive with the other nucleophiles that are capable of causing the re-aromatization of the QM intermediate, *i.e.*, water and cellulose/hemicellulose. The formation of the α -O- γ linkage possesses the greatest thermodynamic favorability (lowest reaction free energy) compared to any other linkage identified presently, via nucleophilic addition to the QM monolignol intermediate. Furthermore, the α -O- γ linkage has the added implication that it demonstrates a branching point in the lignin structure, as it would readily undergo dehydrogenation during lignin polymerization. Therefore, the present results provides evidence that the predominant β -O-4 lignin linkage can act as a branching point via the formation of an α -O- γ linkage. Previous experimental work has demonstrated that the known branching linkages in lignin could not explain the entirety of branching/crosslinking in the lignin structure. As such, the present results provide an explanation for the increased amounts of branching/crosslinking in the lignin structure.

Although evidence suggested that chemical interactions, *i.e.*, LCCs, between lignin and cel-

lulose/hemicellulose were responsible for biomass recalcitrance, which type of linkage, as well as the corresponding formation and deconstruction mechanisms remained largely unknown. The present work computed the reaction energetics associated with the formation of the benzyl ether, benzyl ester and PG LCC linkages and determined that benzyl ether and PG linkages possessed facile kinetics and exergonic reaction energies, suggesting they were likely present in native biomass. Elucidating the reaction mechanisms, kinetics and thermodynamics associated with the deconstruction of benzyl ether and PG linkages revealed that benzyl ether linkages were likely responsible for biomass recalcitrance. Furthermore, cleaving benzyl ether LCC linkages to obtain lignin and carbohydrates in their chemically intact form was demonstrated to be a thermodynamically controlled process. Therefore, separating cellulose, hemicellulose and lignin in their chemically intact form requires creating a reaction environment that alters the thermodynamics of the deconstruction pathways to favor cleaving benzyl ether LCC linkages. As such, this investigation provides deeper insight into the molecular origin of biomass recalcitrance and provides a new perspective for developing milder biomass deconstruction technologies

Previous literature provided strong evidence to suggest that branching does occur in the lignin structure, however, an unrecognized branching linkage in lignin remained to be identified. The present work quantified a novel reaction pathway that could lead to branching in the lignin structure. In summary, the present thesis provides deeper insight in the convoluted nature of lignin polymerization, that will play a key role in developing novel deconstruction technology to recover chemically intact cellulose, hemicellulose and lignin.

Future Work

The research outlined in the present thesis has been restricted to relatively isolated systems, i.e., limited environmental considerations. This work can be built upon significantly by expanding the computational systems and incorporating local environmental considerations, such as explicit solvent environments and other biopolymers. The formation reactions of LCC linkages as well as the majority of computational work completed for lignin polymerization, is in the absence of a solvent system. However, in reality these reactions would be in the presence of water and other biopolymers. Whether LCC formation and lignin polymerization proceed in a completely solvated environment or a solvent lean environment as a result of favorable interactions between lignin monomers and other biopolymers remains to be determined.

Furthermore, computing the deconstruction of LCC linkages as well as the competing re-

actions in various solvent environments is a particularly important area of work. Although solvents are capable of drastically altering the reaction energetics during biomass deconstruction, the molecular origin of many complex solvent effects remain to be determined. Identifying the complex solvent effects capable of promoting the cleavage of LCC linkages while preserving the chemical integrity of the biopolymers, provides a foundation for systematically designing/identifying novel solvents for improved biomass deconstruction. Once complex solvent effects in biomass deconstruction are well categorized, solvent technology can be further developed to cleave LCC linkages (chemical interactions) as well as dissociate the biopolymers (physical interactions), providing an opportunity to design deconstruction technology with an excellent atom economy.

Bibliography

- [1] P. McKendry, "Energy production from biomass (part 1): overview of biomass," *Biore-source Technology*, vol. 83, no. 1, pp. 37–46, 2002.
- [2] S. Kim and B. E. Dale, "Global potential bioethanol production from wasted crops and crop residues," *Biomass and Bioenergy*, vol. 26, no. 4, pp. 361–375, 2004.
- [3] D. Tilman, R. Socolow, J. A. Foley, J. Hill, E. Larson, L. Lynd, S. Pacala, J. Reilly, T. Searchinger, C. Somerville *et al.*, "Beneficial biofuels—the food, energy, and environment trilemma," *Science*, vol. 325, no. 5938, pp. 270–271, 2009.
- [4] S. N. Naik, V. V. Goud, P. K. Rout, and A. K. Dalai, "Production of first and second generation biofuels: a comprehensive review," *Renewable and Sustainable Energy Reviews*, vol. 14, no. 2, pp. 578–597, 2010.
- [5] N. Brun, P. Hesemann, and D. Esposito, "Expanding the biomass derived chemical space," *Chemical Science*, vol. 8, no. 7, pp. 4724–4738, 2017.
- [6] X. Li, E. Mupondwa, S. Panigrahi, L. Tabil, S. Sokhansanj, and M. Stumborg, "A review of agricultural crop residue supply in canada for cellulosic ethanol production," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 5, pp. 2954–2965, 2012.
- [7] M. McCoy, "Succinic acid, once a biobased chemical star, is barely being made: the fermentation-derived molecule, conceived in an era of high oil prices, became a victim of the shale revolution," *Chem Eng News*, vol. 97, p. 12, 2019.
- [8] M. McCoy, "Succinic acid maker bioamber is bankrupt," *Chemical & Engineering News*, vol. 96, no. 20, pp. 14–14, 2018.
- [9] "Natural resources canada, "ecoenergy for biofuels program," canada.ca. <https://www.nrcan.gc.ca/energy-efficiency/transportation-alternative-fuels/alternative-fuels/biofuels/ecoenergy-for-biofuels-program/12358> (accessed aug. 15, 2022)."
- [10] "Natural resources canada, "next-generation biofuels fund," canada.ca. <https://www.nrcan.gc.ca/next-generation-biofuels-fund/3643> (accessed aug. 15, 2022)."
- [11] V. S. Inc., "Report card on the alberta bioenergy producer credit program," Viresco Solutions Inc., Tech. Rep., 2015.

- [12] A. C. C. Office, "Bioenergy producer program outline," Government of Alberta, Tech. Rep., 2017.
- [13] "Agriculture and agri-food canada, "the government of canada invests in innovation to help grow canada's bioeconomy: News release," canada.ca. <https://www.canada.ca/en/agriculture-agri-food/news/2019/02/the-government-of-canada-invests-in-innovation-to-help-grow-canadas-bioeconomy.html> (accessed aug. 15, 2022)."
- [14] "Natural resources canada, "canada investing in first-in-canada forest industry technologies in sarnia: News release," canada.ca. <https://www.canada.ca/en/natural-resources-canada/news/2019/08/canada-investing-in-first-in-canada-forest-industry-technologies-in-sarnia.html> (accessed aug. 15, 2022)."
- [15] "Government of canada, "production and conversion of biorefinery cellulose to advanced fuels, biochemicals and biomaterials." <https://www.nrcan.gc.ca/energy/funding/current-funding-programs/eii/16051> (accessed aug. 15, 2022)."
- [16] "Natural resources canada, "lignin-to-drop-in biojetfuels and chemicals," canada.ca. <https://www.nrcan.gc.ca/science-and-data/funding-partnerships/funding-opportunities/current-investments/lignin-drop-biojetfuels-and-chemicals/16049> (accessed aug. 15, 2022)."
- [17] "Natural resources canada, "biomass-rich waste conversion into drop-in fuels," canada.ca. <https://www.nrcan.gc.ca/energy/funding/current-funding-programs/eii/16080> (accessed aug. 15, 2022)."
- [18] V. B. Agbor, N. Cicek, R. Sparling, A. Berlin, and D. B. Levin, "Biomass pretreatment: fundamentals toward application," *Biotechnology advances*, vol. 29, no. 6, pp. 675–685, 2011.
- [19] N. Mosier, C. Wyman, B. Dale, R. Elander, Y. Lee, M. Holtzappple, and M. Ladisch, "Features of promising technologies for pretreatment of lignocellulosic biomass," *Biore-source technology*, vol. 96, no. 6, pp. 673–686, 2005.
- [20] B. Yang and C. E. Wyman, "Pretreatment: the key to unlocking low-cost cellulosic ethanol," *Biofuels, Bioproducts and Biorefining: Innovation for a sustainable economy*, vol. 2, no. 1, pp. 26–40, 2008.
- [21] Y.-H. P. Zhang, S.-Y. Ding, J. R. Mielenz, J.-B. Cui, R. T. Elander, M. Laser, M. E. Himmel, J. R. McMillan, and L. R. Lynd, "Fractionating recalcitrant lignocellulose at modest reaction conditions," *Biotechnology and bioengineering*, vol. 97, no. 2, pp. 214–223, 2007.
- [22] R. E. Davis, N. J. Grundl, L. Tao, M. J. Bidy, E. C. Tan, G. T. Beckham, D. Humbird, D. N. Thompson, and M. S. Roni, "Process design and economics for the conversion of lignocellulosic biomass to hydrocarbon fuels and coproducts: 2018 biochemical design

- case update; biochemical deconstruction and conversion of biomass to fuels and products via integrated biorefinery pathways,” National Renewable Energy Lab.(NREL), Golden, CO (United States), Tech. Rep., 2018.
- [23] S. P. Chundawat, G. T. Beckham, M. E. Himmel, and B. E. Dale, “Deconstruction of lignocellulosic biomass to fuels and chemicals,” *Annual review of chemical and biomolecular engineering*, vol. 2, pp. 121–145, 2011.
- [24] L. Petridis and J. C. Smith, “Molecular-level driving forces in lignocellulosic biomass deconstruction for bioenergy,” *Nature Reviews Chemistry*, vol. 2, no. 11, pp. 382–389, 2018.
- [25] A. Anukam and J. Berghel, *Biomass pretreatment and characterization: A review*. IntechOpen London, UK, 2020, vol. 1.
- [26] S. V. Vassilev, D. Baxter, L. K. Andersen, and C. G. Vassileva, “An overview of the chemical composition of biomass,” *Fuel*, vol. 89, no. 5, pp. 913–933, 2010.
- [27] F. Xu, Y.-C. Shi, and D. Wang, “X-ray scattering studies of lignocellulosic biomass: a review,” *Carbohydrate polymers*, vol. 94, no. 2, pp. 904–917, 2013.
- [28] Y. Nishiyama, “Structure and properties of the cellulose microfibril,” *Journal of wood science*, vol. 55, no. 4, pp. 241–249, 2009.
- [29] R. A. Burton, M. J. Gidley, and G. B. Fincher, “Heterogeneity in the chemistry, structure and function of plant cell walls,” *Nature chemical biology*, vol. 6, no. 10, pp. 724–732, 2010.
- [30] H. V. Scheller and P. Ulvskov, “Hemicelluloses,” *Annual review of plant biology*, vol. 61, pp. 263–289, 2010.
- [31] B. C. Saha, “Hemicellulose bioconversion,” *Journal of industrial microbiology and biotechnology*, vol. 30, no. 5, pp. 279–291, 2003.
- [32] O. M. Terrett and P. Dupree, “Covalent interactions between lignin and hemicelluloses in plant secondary cell walls,” *Current Opinion in Biotechnology*, vol. 56, pp. 97–104, 2019.
- [33] C. Crestini, F. Melone, M. Sette, and R. Saladino, “Milled wood lignin: a linear oligomer,” *Biomacromolecules*, vol. 12, no. 11, pp. 3928–3935, 2011.
- [34] F. Yue, F. Lu, S. Ralph, and J. Ralph, “Identification of 4-o-5 units in softwood lignins via definitive lignin models and nmr,” *Biomacromolecules*, vol. 17, no. 6, pp. 1909–1920, 2016.
- [35] M. Balakshin, E. A. Capanema, X. Zhu, I. Sulaeva, A. Potthast, T. Rosenau, and O. J. Rojas, “Spruce milled wood lignin: linear, branched or cross-linked?” *Green Chemistry*, vol. 22, no. 13, pp. 3985–4001, 2020.

- [36] W. Boerjan, J. Ralph, and M. Baucher, "Lignin biosynthesis," *Annual review of plant biology*, vol. 54, no. 1, pp. 519–546, 2003.
- [37] J. Ralph, K. Lundquist, G. Brunow, F. Lu, H. Kim, P. F. Schatz, J. M. Marita, R. D. Hatfield, S. A. Ralph, J. H. Christensen *et al.*, "Lignins: natural polymers from oxidative coupling of 4-hydroxyphenyl-propanoids," *Phytochemistry reviews*, vol. 3, no. 1, pp. 29–60, 2004.
- [38] J. Ralph, C. Lapierre, and W. Boerjan, "Lignin structure and its engineering," *Current opinion in biotechnology*, vol. 56, pp. 240–249, 2019.
- [39] K. Ruel, F. Barnoud, and D. Goring, "Lamellation in the s2 layer of softwood tracheids as demonstrated by scanning transmission electron microscopy," *Wood Science and Technology*, vol. 12, no. 4, pp. 287–291, 1978.
- [40] J. Bidlack, "Molecular structure and component integration of secondary cell walls in plants," in *Proceedings of the Oklahoma Academy of Science*, vol. 72, 1992, pp. 51–56.
- [41] N. Terashima, K. Kitano, M. Kojima, M. Yoshida, H. Yamamoto, and U. Westermark, "Nanostructural assembly of cellulose, hemicellulose, and lignin in the middle layer of secondary wall of ginkgo tracheid," *Journal of wood science*, vol. 55, no. 6, pp. 409–416, 2009.
- [42] Y. Zhao, U. Shakeel, M. S. U. Rehman, H. Li, X. Xu, and J. Xu, "Lignin-carbohydrate complexes (lccs) and its role in biorefinery," *Journal of Cleaner Production*, vol. 253, p. 120076, 2020.
- [43] M. Schuetz, R. Smith, and B. Ellis, "Xylem tissue specification, patterning, and differentiation mechanisms," *Journal of experimental botany*, vol. 64, no. 1, pp. 11–31, 2013.
- [44] D. Tarasov, M. Leitch, and P. Fatehi, "Lignin-carbohydrate complexes: properties, applications, analyses, and methods of extraction: a review," *Biotechnology for biofuels*, vol. 11, no. 1, pp. 1–28, 2018.
- [45] C.-J. Liu, "Deciphering the enigma of lignification: precursor transport, oxidation, and the topochemistry of lignin assembly," *Molecular Plant*, vol. 5, no. 2, pp. 304–317, 2012.
- [46] J. C. Del Río, J. Rencoret, P. Prinsen, A. T. Martínez, J. Ralph, and A. Gutierrez, "Structural characterization of wheat straw lignin as revealed by analytical pyrolysis, 2d-nmr, and reductive cleavage methods," *Journal of Agricultural and Food Chemistry*, vol. 60, no. 23, pp. 5922–5935, 2012.
- [47] L. B. Davin and N. G. Lewis, "Dirigent proteins and dirigent sites explain the mystery of specificity of radical precursor coupling in lignan and lignin biosynthesis," *Plant physiology*, vol. 123, no. 2, pp. 453–462, 2000.

- [48] J. Ralph, G. Brunow, P. J. Harris, R. A. Dixon, P. F. Schatz, and W. Boerjan, “Lignification: are lignins biosynthesized via simple combinatorial chemistry or via proteinaceous control and template replication,” *Recent advances in polyphenol research*, vol. 1, pp. 36–66, 2008.
- [49] R. Hatfield and W. Vermerris, “Lignin formation in plants. the dilemma of linkage specificity,” *Plant physiology*, vol. 126, no. 4, pp. 1351–1357, 2001.
- [50] M. Chabannes, K. Ruel, A. Yoshinaga, B. Chabbert, A. Jauneau, J.-P. Joseleau, and A.-M. Boudet, “In situ analysis of lignins in transgenic tobacco reveals a differential impact of individual transformations on the spatial patterns of lignin deposition at the cellular and subcellular levels,” *The Plant Journal*, vol. 28, no. 3, pp. 271–282, 2001.
- [51] N. Demont-Caulet, C. Lapierre, L. Jouanin, S. Baumberger, and V. Méchin, “Arabidopsis peroxidase-catalyzed copolymerization of coniferyl and sinapyl alcohols: kinetics of an endwise process,” *Phytochemistry*, vol. 71, no. 14-15, pp. 1673–1683, 2010.
- [52] D. Fournand, B. Cathala, and C. Lapierre, “Initial steps of the peroxidase-catalyzed polymerization of coniferyl alcohol and/or sinapyl aldehyde: capillary zone electrophoresis study of ph effect,” *Phytochemistry*, vol. 62, no. 2, pp. 139–146, 2003.
- [53] Y. Tobimatsu, T. Takano, H. Kamitakahara, and F. Nakatsubo, “Reactivity of syringyl quinone methide intermediates in dehydrogenative polymerization. part 2: ph effect in horseradish peroxidase-catalyzed polymerization of sinapyl alcohol,” *Holzforshung*, vol. 64, pp. 183–192, 2010.
- [54] S. Aminzadeh, L. Zhang, and G. Henriksson, “A possible explanation for the structural inhomogeneity of lignin in lcc networks,” *Wood Science and Technology*, vol. 51, no. 6, pp. 1365–1376, 2017.
- [55] H. Hwang, S.-J. Moon, K. Won, Y. H. Kim, and J. W. Choi, “Parameters affecting in vitro monolignol couplings during dehydrogenative polymerization in the presence of peroxidase and h₂o₂,” *Journal of Industrial and Engineering Chemistry*, vol. 26, pp. 390–395, 2015.
- [56] N. Terashima, R. Atalla, S. Ralph, L. L. Landucci, C. Lapierre, and B. Monties, “New preparations of lignin polymer models under conditions that approximate cell wall lignification. i. synthesis of novel lignin polymer models and their structural characterization by ¹³c nmr,” *Holzforshung*, vol. 49, no. 6, pp. 521–527, 1995.
- [57] Q. Li, K. Koda, A. Yoshinaga, K. Takabe, M. Shimomura, Y. Hirai, Y. Tamai, and Y. Uraki, “Dehydrogenative polymerization of coniferyl alcohol in artificial polysaccharides matrices: effects of xylan on the polymerization,” *Journal of agricultural and food chemistry*, vol. 63, no. 18, pp. 4613–4620, 2015.
- [58] T. Warinowski, S. Koutaniemi, A. Kärkönen, I. Sundberg, M. Toikka, L. K. Simola, I. Kilpeläinen, and T. H. Teeri, “Peroxidases bound to the growing lignin polymer produce natural like extracellular lignin in a cell culture of norway spruce,” *Frontiers in plant science*, vol. 7, p. 1523, 2016.

- [59] Y. Tobimatsu and M. Schuetz, "Lignin polymerization: how do plants manage the chemistry so well?" *Current Opinion in Biotechnology*, vol. 56, pp. 75–81, 2019.
- [60] U. Takahama, T. Oniki, and H. Shimokawa, "A possible mechanism for the oxidation of sinapyl alcohol by peroxidase-dependent reactions in the apoplast: enhancement of the oxidation by hydroxycinnamic acids and components of the apoplast," *Plant and cell physiology*, vol. 37, no. 4, pp. 499–504, 1996.
- [61] H. Önnerud, L. Zhang, G. Gellerstedt, and G. Henriksson, "Polymerization of monolignols by redox shuttle-mediated enzymatic oxidation: a new model in lignin biosynthesis i," *The Plant Cell*, vol. 14, no. 8, pp. 1953–1962, 2002.
- [62] A. K. Sangha, J. M. Parks, R. F. Standaert, A. Ziebell, M. Davis, and J. C. Smith, "Radical coupling reactions in lignin synthesis: a density functional theory study," *The Journal of Physical Chemistry B*, vol. 116, no. 16, pp. 4760–4768, 2012.
- [63] T. Z. Gani, M. J. Orella, E. M. Anderson, M. L. Stone, F. R. Brushett, G. T. Beckham, and Y. Román-Leshkov, "Computational evidence for kinetically controlled radical coupling during lignification," *ACS Sustainable Chemistry & Engineering*, vol. 7, no. 15, pp. 13 270–13 277, 2019.
- [64] F. R. van Parijs, K. Morreel, J. Ralph, W. Boerjan, and R. M. Merks, "Modeling lignin polymerization. i. simulation model of dehydrogenation polymers," *Plant physiology*, vol. 153, no. 3, pp. 1332–1344, 2010.
- [65] M. J. Orella, T. Z. Gani, J. V. Vermaas, M. L. Stone, E. M. Anderson, G. T. Beckham, F. R. Brushett, and Y. Román-Leshkov, "Lignin-kmc: A toolkit for simulating lignin biosynthesis," *ACS Sustainable Chemistry & Engineering*, vol. 7, no. 22, pp. 18 313–18 322, 2019.
- [66] S. Levy, W. S. York, R. Stuike-Prill, B. Meyer, and L. A. Staehelin, "Simulations of the static and dynamic molecular conformations of xyloglucan. the role of the fucosylated sidechain in surface-specific sidechain folding," *The Plant Journal*, vol. 1, no. 2, pp. 195–215, 1991.
- [67] V. Finkenstadt, T. Hendrixson, and R. Millane, "Models of xyloglucan binding to cellulose microfibrils1," *Carbohydrate Chemistry*, vol. 14, no. 4-5, pp. 601–611, 1995.
- [68] S. Levy, G. Maclachlan, and L. A. Staehelin, "Xyloglucan sidechains modulate binding to cellulose during in vitro binding assays as predicted by conformational dynamics simulations," *The Plant Journal*, vol. 11, no. 3, pp. 373–386, 1997.
- [69] J. Hanus and K. Mazeau, "The xyloglucan–cellulose assembly at the atomic scale," *Biopolymers: Original Research on Biomolecules*, vol. 82, no. 1, pp. 59–73, 2006.
- [70] Q. Zhang, H. Brumer, H. Ågren, and Y. Tu, "The adsorption of xyloglucan on cellulose: effects of explicit water and side chain variation," *Carbohydrate research*, vol. 346, no. 16, pp. 2595–2602, 2011.

- [71] K. Mazeau and L. Charlier, “The molecular basis of the adsorption of xylans on cellulose surface,” *Cellulose*, vol. 19, no. 2, pp. 337–349, 2012.
- [72] R. L. Silveira, S. R. Stoyanov, S. Gusarov, M. S. Skaf, and A. Kovalenko, “Plant biomass recalcitrance: effect of hemicellulose composition on nanoscale forces that control cell wall strength,” *Journal of the American Chemical Society*, vol. 135, no. 51, pp. 19 048–19 051, 2013.
- [73] C. S. Pereira, R. L. Silveira, P. Dupree, and M. S. Skaf, “Effects of xylan side-chain substitutions on xylan–cellulose interactions and implications for thermal pretreatment of cellulosic biomass,” *Biomacromolecules*, vol. 18, no. 4, pp. 1311–1321, 2017.
- [74] N. Zhang, S. Li, L. Xiong, Y. Hong, and Y. Chen, “Cellulose-hemicellulose interaction in wood secondary cell-wall,” *Modelling and Simulation in Materials Science and Engineering*, vol. 23, no. 8, p. 085010, 2015.
- [75] A. Khodayari, W. Thielemans, U. Hirn, A. W. Van Vuure, and D. Seveno, “Cellulose-hemicellulose interactions-a nanoscale view,” *Carbohydrate Polymers*, vol. 270, p. 118364, 2021.
- [76] R. Kumar, S. Bhagia, M. D. Smith, L. Petridis, R. G. Ong, C. M. Cai, A. Mittal, M. H. Himmel, V. Balan, B. E. Dale *et al.*, “Cellulose–hemicellulose interactions at elevated temperatures increase cellulose recalcitrance to biological conversion,” *Green Chemistry*, vol. 20, no. 4, pp. 921–934, 2018.
- [77] C. J. Houtman and R. H. Atalla, “Cellulose-lignin interactions (a computational study),” *Plant physiology*, vol. 107, no. 3, pp. 977–984, 1995.
- [78] S. M. Shevchenko and G. W. Bailey, “The mystery of the lignin-carbohydrate complex: a computational approach,” *Journal of Molecular Structure: THEOCHEM*, vol. 364, no. 2-3, pp. 197–208, 1996.
- [79] S. Besombes and K. Mazeau, “The cellulose/lignin assembly assessed by molecular modeling. part 1: adsorption of a threo guaiacyl β -o-4 dimer onto a i β cellulose whisker,” *Plant Physiology and Biochemistry*, vol. 43, no. 3, pp. 299–308, 2005.
- [80] —, “The cellulose/lignin assembly assessed by molecular modeling. part 2: seeking for evidence of organization of lignin molecules at the interface with cellulose,” *Plant Physiology and Biochemistry*, vol. 43, no. 3, pp. 277–286, 2005.
- [81] B. Lindner, L. Petridis, R. Schulz, and J. C. Smith, “Solvent-driven preferential association of lignin with regions of crystalline cellulose in molecular dynamics simulation,” *Biomacromolecules*, vol. 14, no. 10, pp. 3390–3398, 2013.
- [82] J. V. Vermaas, M. F. Crowley, and G. T. Beckham, “A quantitative molecular atlas for interactions between lignin and cellulose,” *ACS Sustainable Chemistry & Engineering*, vol. 7, no. 24, pp. 19 570–19 583, 2019.

- [83] A. S. Patri, B. Mostofian, Y. Pu, N. Ciaffone, M. Soliman, M. D. Smith, R. Kumar, X. Cheng, C. E. Wyman, L. Tetard *et al.*, “A multifunctional cosolvent pair reveals molecular principles of biomass deconstruction,” *Journal of the American Chemical Society*, vol. 141, no. 32, pp. 12545–12557, 2019.
- [84] R. L. Silveira, S. R. Stoyanov, S. Gusarov, M. S. Skaf, and A. Kovalenko, “Supramolecular interactions in secondary plant cell walls: effect of lignin chemical composition revealed with the molecular theory of solvation,” *The journal of physical chemistry letters*, vol. 6, no. 1, pp. 206–211, 2015.
- [85] D.-y. Min, C. Yang, V. Chiang, H. Jameel, and H.-m. Chang, “The influence of lignin–carbohydrate complexes on the cellulase-mediated saccharification ii: Transgenic hybrid poplars (*populus nigra* l. and *populus maximowiczii* a.),” *Fuel*, vol. 116, pp. 56–62, 2014.
- [86] N. Giummarella and M. Lawoko, “Structural basis for the formation and regulation of lignin–xylan bonds in birch,” *ACS Sustainable Chemistry & Engineering*, vol. 4, no. 10, pp. 5319–5326, 2016.
- [87] N. Giummarella, Y. Pu, A. J. Ragauskas, and M. Lawoko, “A critical review on the analysis of lignin carbohydrate bonds,” *Green Chemistry*, vol. 21, no. 7, pp. 1573–1595, 2019.
- [88] K. M. Holtman, H.-m. Chang, H. Jameel, and J. F. Kadla, “Quantitative ¹³c nmr characterization of milled wood lignins isolated by different milling techniques,” *Journal of Wood Chemistry and Technology*, vol. 26, no. 1, pp. 21–34, 2006.
- [89] M. Balakshin, E. Capanema, H. Gracz, H.-m. Chang, and H. Jameel, “Quantification of lignin–carbohydrate linkages with high-resolution nmr spectroscopy,” *Planta*, vol. 233, no. 6, pp. 1097–1110, 2011.
- [90] J. Li, G. Henriksson, and G. Gellerstedt, “Lignin depolymerization/repolymerization and its critical role for delignification of aspen wood by steam explosion,” *Bioresource technology*, vol. 98, no. 16, pp. 3061–3068, 2007.
- [91] R. Hatfield and R. S. Fukushima, “Can lignin be accurately measured?” *Crop science*, vol. 45, no. 3, pp. 832–839, 2005.
- [92] J. Erdmann, “Ueber die concretionen in den birnen,” *Justus Liebigs Annalen der Chemie*, vol. 138, no. 1, pp. 1–19, 1866.
- [93] J. Ralph, J. H. Grabber, and R. D. Hatfield, “Lignin-ferulate cross-links in grasses: active incorporation of ferulate polysaccharide esters into ryegrass lignins,” *Carbohydrate research*, vol. 275, no. 1, pp. 167–178, 1995.
- [94] J. H. Grabber, J. Ralph, and R. D. Hatfield, “Model studies of ferulate- coniferyl alcohol cross-product formation in primary maize walls: implications for lignification in grasses,” *Journal of Agricultural and Food Chemistry*, vol. 50, no. 21, pp. 6008–6016, 2002.

- [95] J. Ralph, "Hydroxycinnamates in lignification," *Phytochemistry Reviews*, vol. 9, no. 1, pp. 65–83, 2010.
- [96] H. Nishimura, A. Kamiya, T. Nagata, M. Katahira, and T. Watanabe, "Direct evidence for α ether linkage between lignin and carbohydrates in wood cell walls," *Scientific reports*, vol. 8, no. 1, pp. 1–11, 2018.
- [97] T. Watanabe, S. Kaizu, and T. Koshijima, "Binding sites of carbohydrate moieties toward lignin in "lignin-carbohydrate complex" from pinus densiflora wood," *Chemistry Letters*, vol. 15, no. 11, pp. 1871–1874, 1986.
- [98] J. Zeng, G. L. Helms, X. Gao, and S. Chen, "Quantification of wheat straw lignin structure by comprehensive nmr analysis," *Journal of Agricultural and Food Chemistry*, vol. 61, no. 46, pp. 10 848–10 857, 2013.
- [99] T. Watanabe and T. Koshijima, "Evidence for an ester linkage between lignin and glucuronic acid in lignin-carbohydrate complexes by ddq-oxidation," *Agricultural and biological chemistry*, vol. 52, no. 11, pp. 2953–2955, 1988.
- [100] Y. Miyagawa, Y. Tobimatsu, P. Y. Lam, T. Mizukami, S. Sakurai, H. Kamitakahara, and T. Takano, "Possible mechanisms for the generation of phenyl glycoside-type lignin-carbohydrate linkages in lignification with monolignol glucosides," *The Plant Journal*, vol. 104, no. 1, pp. 156–170, 2020.
- [101] Y. Xie, S. Yasuda, H. Wu, and H. Liu, "Analysis of the structure of lignin-carbohydrate complexes by the specific ^{13}C tracer method," *Journal of wood science*, vol. 46, no. 2, pp. 130–136, 2000.
- [102] H. I. Bolker, "A lignin carbohydrate bond as revealed by infra-red spectroscopy," *Nature*, vol. 197, no. 4866, pp. 489–490, 1963.
- [103] Y. Miyagawa, T. Mizukami, H. Kamitakahara, and T. Takano, "Synthesis and fundamental hsqc nmr data of monolignol β -glycosides, dihydromonolignol β -glycosides and p-hydroxybenzaldehyde derivative β -glycosides for the analysis of phenyl glycoside type lignin-carbohydrate complexes (lccs)," *Holzforschung*, vol. 68, no. 7, pp. 747–760, 2014.
- [104] G. Ward, Y. Hadar, I. Bilkis, L. Konstantinovskiy, and C. G. Dosoretz, "Initial steps of ferulic acid polymerization by lignin peroxidase," *Journal of Biological Chemistry*, vol. 276, no. 22, pp. 18 734–18 741, 2001.
- [105] F. Carunchio, C. Crescenzi, A. M. Girelli, A. Messina, and A. M. Tarola, "Oxidation of ferulic acid by laccase: identification of the products and inhibitory effects of some dipeptides," *Talanta*, vol. 55, no. 1, pp. 189–200, 2001.
- [106] M. Remko, "Molecular orbital investigations on lignin model compounds. xx. conformational analysis of benzyl ether and benzyl ester linkages in lignin-carbohydrate complexes," *Holzforschung*, vol. 40, no. 4, pp. 205–209, 1986.

- [107] S. Diner, J. Malrieu, F. Jordan, and M. Gilbert, “Localized bond orbitals and the correlation problem-iii. energy up to the third-order in the zero-differential overlap approximation. application to σ -electron systems,” *Theoretica Chimica Acta*, vol. 15, no. 2, pp. 100–110, 1969.
- [108] B. Durbeej and L. A. Eriksson, “Formation of β -o-4 lignin models-a theoretical study,” *Holzforschung*, vol. 57, no. 5, pp. 466–478, 2003.
- [109] X. Zhang, W. Yang, and W. Blasiak, “Modeling study of woody biomass: interactions of cellulose, hemicellulose, and lignin,” *Energy & Fuels*, vol. 25, no. 10, pp. 4786–4795, 2011.
- [110] A. D. Becke, “Density-functional exchange-energy approximation with correct asymptotic behavior,” *Physical review A*, vol. 38, no. 6, p. 3098, 1988.
- [111] C. Lee, W. Yang, and R. G. Parr, “Development of the colle-salvetti correlation-energy formula into a functional of the electron density,” *Physical review B*, vol. 37, no. 2, p. 785, 1988.
- [112] A. D. Becke, “Density-functional thermochemistry. iv. a new dynamical correlation functional and implications for exact-exchange mixing,” *The Journal of chemical physics*, vol. 104, no. 3, pp. 1040–1046, 1996.
- [113] L. Charlier and K. Mazeau, “Molecular modeling of the structural and dynamical properties of secondary plant cell walls: influence of lignin chemistry,” *The Journal of Physical Chemistry B*, vol. 116, no. 14, pp. 4163–4174, 2012.
- [114] N. Jacquet, G. Maniet, C. Vanderghem, F. Delvigne, and A. Richel, “Application of steam explosion as pretreatment on lignocellulosic material: a review,” *Industrial & Engineering Chemistry Research*, vol. 54, no. 10, pp. 2593–2598, 2015.
- [115] P. Langan, L. Petridis, H. M. O’Neill, S. V. Pingali, M. Foston, Y. Nishiyama, R. Schulz, B. Lindner, B. L. Hanson, S. Harton *et al.*, “Common processes drive the thermochemical pretreatment of lignocellulosic biomass,” *Green Chemistry*, vol. 16, no. 1, pp. 63–68, 2014.
- [116] W.-H. Chen, S. Nižetić, R. Sirohi, Z. Huang, R. Luque, A. M. Papadopoulos, R. Sakthivel, X. P. Nguyen, and A. T. Hoang, “Liquid hot water as sustainable biomass pretreatment technique for bioenergy production: A review,” *Bioresource technology*, vol. 344, p. 126207, 2022.
- [117] E. Shirkevand, S. Baroutian, D. J. Gapes, and B. R. Young, “Combination of fungal and physicochemical processes for lignocellulosic biomass pretreatment—a review,” *Renewable and Sustainable Energy Reviews*, vol. 54, pp. 217–234, 2016.
- [118] S. P. Chundawat, R. K. Pal, C. Zhao, T. Campbell, F. Teymouri, J. Videto, C. Nielson, B. Wieferrich, L. Sousa, B. E. Dale *et al.*, “Ammonia fiber expansion (afex) pretreatment of lignocellulosic biomass,” *JoVE (Journal of Visualized Experiments)*, no. 158, p. e57488, 2020.

- [119] R. R. Gonzales, P. Sivagurunathan, and S.-H. Kim, “Effect of severity on dilute acid pretreatment of lignocellulosic biomass and the following hydrogen fermentation,” *International Journal of Hydrogen Energy*, vol. 41, no. 46, pp. 21 678–21 684, 2016.
- [120] X. Zhao, K. Cheng, and D. Liu, “Organosolv pretreatment of lignocellulosic biomass for enzymatic hydrolysis,” *Applied microbiology and biotechnology*, vol. 82, no. 5, pp. 815–827, 2009.
- [121] K. Zhang, Z. Pei, and D. Wang, “Organic solvent pretreatment of lignocellulosic biomass for biofuels and biochemicals: a review,” *Bioresource technology*, vol. 199, pp. 21–33, 2016.
- [122] J. S. Luterbacher, J. M. Rand, D. M. Alonso, J. Han, J. T. Youngquist, C. T. Maravelias, B. F. Pfleger, and J. A. Dumesic, “Nonenzymatic sugar production from biomass using biomass-derived γ -valerolactone,” *Science*, vol. 343, no. 6168, pp. 277–280, 2014.
- [123] A. S. Patri, R. Mohan, Y. Pu, C. G. Yoo, A. J. Ragauskas, R. Kumar, D. Kisailus, C. M. Cai, and C. E. Wyman, “Thf co-solvent pretreatment prevents lignin redeposition from interfering with enzymes yielding prolonged cellulase activity,” *Biotechnology for biofuels*, vol. 14, no. 1, pp. 1–13, 2021.
- [124] Y. Sun and J. Cheng, “Hydrolysis of lignocellulosic materials for ethanol production: a review,” *Bioresource technology*, vol. 83, no. 1, pp. 1–11, 2002.
- [125] L. Shuai and J. Luterbacher, “Organic solvent effects in biomass conversion reactions,” *ChemSusChem*, vol. 9, pp. 133–155, 2016.
- [126] Y.-Y. Wang, P. Sengupta, B. Scheidemantle, Y. Pu, C. E. Wyman, C. M. Cai, and A. J. Ragauskas, “Effects of celf pretreatment severity on lignin structure and the lignin-based polyurethane properties,” *Frontiers in Energy Research*, vol. 8, p. 149, 2020.
- [127] X. Meng, S. Bhagia, Y. Wang, Y. Zhou, Y. Pu, J. R. Dunlap, L. Shuai, A. J. Ragauskas, and C. G. Yoo, “Effects of the advanced organosolv pretreatment strategies on structural properties of woody biomass,” *Industrial Crops and Products*, vol. 146, p. 112144, 2020.
- [128] T. Raj, K. Chandrasekhar, R. Banu, J.-J. Yoon, G. Kumar, and S.-H. Kim, “Synthesis of γ -valerolactone (gvl) and their applications for lignocellulosic deconstruction for sustainable green biorefineries,” *Fuel*, vol. 303, p. 121333, 2021.
- [129] M. D. Smith, X. Cheng, L. Petridis, B. Mostofian, and J. C. Smith, “Organosolv-water cosolvent phase separation on cellulose and its influence on the physical deconstruction of cellulose: A molecular dynamics analysis,” *Scientific reports*, vol. 7, no. 1, pp. 1–9, 2017.

- [130] B. Mostofian, C. M. Cai, M. D. Smith, L. Petridis, X. Cheng, C. E. Wyman, and J. C. Smith, "Local phase separation of co-solvents enhances pretreatment of biomass for bioenergy applications," *Journal of the American Chemical Society*, vol. 138, no. 34, pp. 10 869–10 878, 2016.
- [131] M. D. Smith, C. M. Cai, X. Cheng, L. Petridis, and J. C. Smith, "Temperature-dependent phase behaviour of tetrahydrofuran–water alters solubilization of xylan to improve co-production of furfurals from lignocellulosic biomass," *Green Chemistry*, vol. 20, no. 7, pp. 1612–1620, 2018.
- [132] M. D. Smith, L. Petridis, X. Cheng, B. Mostofian, and J. C. Smith, "Enhanced sampling simulation analysis of the structure of lignin in the thf–water miscibility gap," *Physical Chemistry Chemical Physics*, vol. 18, no. 9, pp. 6394–6398, 2016.
- [133] C. D'Agostino, G. L. Brett, P. J. Miedziak, D. W. Knight, G. J. Hutchings, L. F. Gladden, and M. D. Mantle, "Understanding the solvent effect on the catalytic oxidation of 1, 4-butanediol in methanol over au/tio₂ catalyst: Nmr diffusion and relaxation studies," *Chemistry—A European Journal*, vol. 18, no. 45, pp. 14 426–14 433, 2012.
- [134] L. Qi, R. Alamillo, W. A. Elliott, A. Andersen, D. W. Hoyt, E. D. Walter, K. S. Han, N. M. Washton, R. M. Rioux, J. A. Dumesic *et al.*, "Operando solid-state nmr observation of solvent-mediated adsorption-reaction of carbohydrates in zeolites," *ACS Catalysis*, vol. 7, no. 5, pp. 3489–3500, 2017.
- [135] M. A. Mellmer, C. Sener, J. M. R. Gallo, J. S. Luterbacher, D. M. Alonso, and J. A. Dumesic, "Solvent effects in acid-catalyzed biomass conversion reactions," *Angewandte chemie international edition*, vol. 53, no. 44, pp. 11 872–11 875, 2014.
- [136] H. Wan, A. Vitter, R. V. Chaudhari, and B. Subramaniam, "Kinetic investigations of unusual solvent effects during ru/c catalyzed hydrogenation of model oxygenates," *Journal of Catalysis*, vol. 309, pp. 174–184, 2014.
- [137] N. Nikbin, S. Caratzoulas, and D. G. Vlachos, "A first principles-based microkinetic model for the conversion of fructose to 5-hydroxymethylfurfural," *ChemCatChem*, vol. 4, no. 4, pp. 504–511, 2012.
- [138] R. Bermejo-Deval, R. Gounder, and M. E. Davis, "Framework and extraframework tin sites in zeolite beta react glucose differently," *ACS catalysis*, vol. 2, no. 12, pp. 2705–2713, 2012.
- [139] Y. Liu, A. Vjunov, H. Shi, S. Eckstein, D. M. Camaioni, D. Mei, E. Baráth, and J. A. Lercher, "Enhancing the catalytic activity of hydronium ions through constrained environments," *Nature Communications*, vol. 8, no. 1, pp. 1–8, 2017.
- [140] B. N. Zope, D. D. Hibbitts, M. Neurock, and R. J. Davis, "Reactivity of the gold/water interface during selective oxidation catalysis," *Science*, vol. 330, no. 6000, pp. 74–78, 2010.

- [141] S. H. Mushrif, J. J. Varghese, and D. G. Vlachos, "Insights into the cr (iii) catalyzed isomerization mechanism of glucose to fructose in the presence of water using ab initio molecular dynamics," *Physical Chemistry Chemical Physics*, vol. 16, no. 36, pp. 19564–19572, 2014.
- [142] J. C. Velasco Calderón, S. Jiang, and S. H. Mushrif, "Understanding the effect of solvent environment on the interaction of hydronium ion with biomass derived species: A molecular dynamics and metadynamics investigation," *ChemPhysChem*, vol. 22, no. 21, pp. 2222–2230, 2021.
- [143] R. Car and M. Parrinello, "Unified approach for molecular dynamics and density-functional theory," *Physical review letters*, vol. 55, no. 22, p. 2471, 1985.
- [144] R. S. Assary, P. C. Redfern, J. R. Hammond, J. Greeley, and L. A. Curtiss, "Computational studies of the thermochemistry for conversion of glucose to levulinic acid," *The Journal of Physical Chemistry B*, vol. 114, no. 27, pp. 9002–9009, 2010.
- [145] M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. Petersson *et al.*, "Gaussian 09, revision d. 01, 2009, gaussian," *Inc., Wallingford CT*, 2009.
- [146] Y. Zhao and D. G. Truhlar, "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*, vol. 120, no. 1, pp. 215–241, 2008.
- [147] J. S. Arora, K. B. Ansari, J. W. Chew, P. J. Dauenhauer, and S. H. Mushrif, "Unraveling the catalytic influence of naturally occurring salts on biomass pyrolysis chemistry using glucose as a model compound: a combined experimental and dft study," *Catalysis Science & Technology*, vol. 9, no. 13, pp. 3504–3524, 2019.
- [148] H. B. Schlegel, "Optimization of equilibrium geometries and transition structures," *Journal of computational chemistry*, vol. 3, no. 2, pp. 214–218, 1982.
- [149] X. Li and M. J. Frisch, "Energy-represented direct inversion in the iterative subspace within a hybrid geometry optimization method," *Journal of chemical theory and computation*, vol. 2, no. 3, pp. 835–839, 2006.
- [150] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, "General performance of density functionals," *The Journal of Physical Chemistry A*, vol. 111, no. 42, pp. 10439–10452, 2007.
- [151] M. Appell, G. Strati, J. Willett, and F. Momany, "B3lyp/6-311++ g** study of α - and β -d-glucopyranose and 1, 5-anhydro-d-glucitol: 4c1 and 1c4 chairs, 3, ob and b3, o boats, and skew-boat conformations," *Carbohydrate research*, vol. 339, no. 3, pp. 537–551, 2004.

- [152] F. Momany, M. Appell, G. Strati, and J. Willett, “B3lyp/6-311++ g** study of monohydrates of α - and β -d-glucopyranose: hydrogen bonding, stress energies, and effect of hydration on internal coordinates,” *Carbohydrate research*, vol. 339, no. 3, pp. 553–567, 2004.
- [153] A. Kovalenko and F. Hirata, “Self-consistent description of a metal–water interface by the kohn–sham density functional theory and the three-dimensional reference interaction site model,” *The Journal of chemical physics*, vol. 110, no. 20, pp. 10 095–10 112, 1999.
- [154] A. Kovalenko, “Multiscale modeling of solvation in chemical and biological nanosystems and in nanoporous materials,” *Pure and Applied Chemistry*, vol. 85, no. 1, pp. 159–199, 2013.
- [155] L. A. Donaldson, “Lignification and lignin topochemistry-an ultrastructural view,” *Phytochemistry*, vol. 57, no. 6, pp. 859–873, 2001.
- [156] U. Westermark, “Calcium promoted phenolic coupling by superoxide radical-a possible lignification reaction in wood,” *Wood Science and Technology*, vol. 16, no. 1, pp. 71–78, 1982.
- [157] H. Felle, “ph: signal and messenger in plant cells,” *Plant Biology*, vol. 3, no. 6, pp. 577–591, 2001.
- [158] J. L. Hatfield and J. H. Prueger, “Temperature extremes: Effect on plant growth and development,” *Weather and Climate Extremes*, vol. 10, pp. 4–10, 2015.
- [159] J. J. Lyczakowski, K. B. Wicher, O. M. Terrett, N. Faria-Blanc, X. Yu, D. Brown, K. B. Krogh, P. Dupree, and M. Busse-Wicher, “Removal of glucuronic acid from xylan is a strategy to improve the conversion of plant biomass to sugars for bioenergy,” *Biotechnology for biofuels*, vol. 10, no. 1, pp. 1–11, 2017.
- [160] Z. Jin, K. S. Katsumata, T. B. T. Lam, and K. Iiyama, “Covalent linkages between cellulose and lignin in cell walls of coniferous and nonconiferous woods,” *Biopolymers: Original Research on Biomolecules*, vol. 83, no. 2, pp. 103–110, 2006.
- [161] M. H. Sipponen, C. Lapierre, V. Méchin, and S. Baumberger, “Isolation of structurally distinct lignin–carbohydrate fractions from maize stem by sequential alkaline extractions and endoglucanase treatment,” *Bioresource technology*, vol. 133, pp. 522–528, 2013.
- [162] Y. Li, Y. Liu, W. Chen, Q. Wang, Y. Liu, J. Li, and H. Yu, “Facile extraction of cellulose nanocrystals from wood using ethanol and peroxide solvothermal pretreatment followed by ultrasonic nanofibrillation,” *Green Chemistry*, vol. 18, no. 4, pp. 1010–1018, 2016.
- [163] K. Yao, Q. Wu, R. An, W. Meng, M. Ding, B. Li, and Y. Yuan, “Hydrothermal pretreatment for deconstruction of plant cell wall: Part i. effect on lignin-carbohydrate complex,” *AIChE Journal*, vol. 64, no. 6, pp. 1938–1953, 2018.

- [164] G. Zundel, "Hydration structure and intermolecular interaction in polyelectrolytes," *Angewandte Chemie International Edition in English*, vol. 8, no. 7, pp. 499–509, 1969.
- [165] Q. Nguyen, M. Tucker, B. Boynton, F. Keller, and D. Schell, "Dilute acid pretreatment of softwoods: Scientific note," *Applied biochemistry and biotechnology*, vol. 70, pp. 77–87, 1998.
- [166] Q. A. Nguyen, M. P. Tucker, F. A. Keller, D. A. Beaty, K. M. Connors, and F. P. Eddy, "Dilute acid hydrolysis of softwoods: Scientific note," in *Twentieth Symposium on Biotechnology for Fuels and Chemicals: Presented as Volumes 77–79 of Applied Biochemistry and Biotechnology Proceedings of the Twentieth Symposium on Biotechnology for Fuels and Chemicals Held May 3–7, 1998, Gatlinburg, Tennessee*. Springer, 1999, pp. 133–142.
- [167] Q. A. Nguyen, M. P. Tucker, F. A. Keller, and F. P. Eddy, "Two-stage dilute-acid pretreatment of softwoods," in *Twenty-First Symposium on Biotechnology for Fuels and Chemicals: Proceedings of the Twenty-First Symposium on Biotechnology for Fuels and Chemicals Held May 2–6, 1999, in Fort Collins, Colorado*. Springer, 2000, pp. 561–576.
- [168] M. A. Mellmer, D. M. Alonso, J. S. Luterbacher, J. M. R. Gallo, and J. A. Dumesic, "Effects of γ -valerolactone in hydrolysis of lignocellulosic biomass to monosaccharides," *Green Chemistry*, vol. 16, no. 11, pp. 4659–4662, 2014.
- [169] G. Tsilomelekis, T. R. Josephson, V. Nikolakis, and S. Caratzoulas, "Origin of 5-hydroxymethylfurfural stability in water/dimethyl sulfoxide mixtures," *ChemSusChem*, vol. 7, no. 1, pp. 117–126, 2014.
- [170] J. J. Varghese and S. H. Mushrif, "Origins of complex solvent effects on chemical reactivity and computational tools to investigate them: a review," *Reaction Chemistry & Engineering*, vol. 4, no. 2, pp. 165–206, 2019.
- [171] M. Y. Balakshin, E. A. Capanema, and H.-m. Chang, "Mwl fraction with a high concentration of lignin-carbohydrate linkages: Isolation and 2d nmr spectroscopic analysis," *Holzforschung*, 2007.
- [172] T.-Q. Yuan, S.-N. Sun, F. Xu, and R.-C. Sun, "Characterization of lignin structures and lignin-carbohydrate complex (lcc) linkages by quantitative ^{13}C and ^{2}D hsqc nmr spectroscopy," *Journal of Agricultural and Food Chemistry*, vol. 59, no. 19, pp. 10604–10614, 2011.
- [173] T.-T. You, L.-M. Zhang, S.-K. Zhou, and F. Xu, "Structural elucidation of lignin-carbohydrate complex (lcc) preparations and lignin from *arundo donax* linn," *Industrial Crops and Products*, vol. 71, pp. 65–74, 2015.
- [174] L. Yao, C. Chen, X. Zheng, Z. Peng, H. Yang, and Y. Xie, "Determination of lignin-carbohydrate complexes structure of wheat straw using carbon-13 isotope as a tracer," *BioResources*, vol. 11, no. 3, pp. 6692–6707, 2016.

- [175] B. Zhang, G.-Q. Fu, Y.-S. Niu, F. Peng, C.-L. Yao, and R.-C. Sun, “Variations of lignin–lignin and lignin–carbohydrate linkages from young neosinocalamus affinis bamboo culms,” *RSC advances*, vol. 6, no. 19, pp. 15 478–15 484, 2016.
- [176] J. C. Del Río, P. Prinsen, E. M. Cadena, Á. T. Martínez, A. Gutiérrez, and J. Rencoret, “Lignin–carbohydrate complexes from sisal (agave sisalana) and abaca (musa textilis): chemical composition and structural modifications during the isolation process,” *Planta*, vol. 243, pp. 1143–1158, 2016.
- [177] Z. Lv, Z. Bai, L. Su, J. Rao, Y. Hu, R. Tian, S. Jia, Y. Guan, B. Lü, and F. Peng, “Unveiling lignin structures and lignin-carbohydrate complex (lcc) linkages of bamboo (phyllostachys pubescens) fibers and parenchyma cells,” *International Journal of Biological Macromolecules*, vol. 241, p. 124461, 2023.
- [178] N. Terashima, C. Ko, Y. Matsushita, and U. Westermark, “Monolignol glucosides as intermediate compounds in lignin biosynthesis. revisiting the cell wall lignification and new 13c-tracer experiments with ginkgo biloba and magnolia liliiflora,” *Holzforschung*, vol. 70, no. 9, pp. 801–810, 2016.
- [179] M. Perkins, R. A. Smith, and L. Samuels, “The transport of monomers during lignification in plants: anything goes but how?” *Current Opinion in Biotechnology*, vol. 56, pp. 69–74, 2019.
- [180] S. Beck, P. Choi, and S. H. Mushrif, “Origins of covalent linkages within the lignin–carbohydrate network of biomass,” *Physical Chemistry Chemical Physics*, vol. 24, no. 34, pp. 20 480–20 490, 2022.
- [181] A. Warshel, P. K. Sharma, M. Kato, Y. Xiang, H. Liu, and M. H. Olsson, “Electrostatic basis for enzyme catalysis,” *Chemical reviews*, vol. 106, no. 8, pp. 3210–3235, 2006.
- [182] L. Zhang, Y. Pu, J. R. Cort, A. J. Ragauskas, and B. Yang, “Revealing the molecular structural transformation of hardwood and softwood in dilute acid flowthrough pretreatment,” *ACS Sustainable Chemistry & Engineering*, vol. 4, no. 12, pp. 6618–6628, 2016.
- [183] J. Lehto, J. Louhelainen, M. Huttunen, and R. Alén, “Spectroscopic analysis of hot-water-and dilute-acid-extracted hardwood and softwood chips,” *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 184, pp. 184–190, 2017.
- [184] C. Kundu, S. P. Samudrala, M. A. Kibria, and S. Bhattacharya, “One-step peracetic acid pretreatment of hardwood and softwood biomass for platform chemicals production,” *Scientific Reports*, vol. 11, no. 1, p. 11183, 2021.
- [185] E. Paone, T. Tabanelli, and F. Mauriello, “The rise of lignin biorefinery,” *Current Opinion in Green and Sustainable Chemistry*, vol. 24, pp. 1–6, 2020.
- [186] H.-M. Chang and X. Jiang, “Biphenyl structure and its impact on the macromolecular structure of lignin: A critical review,” *Journal of Wood Chemistry and Technology*, vol. 40, no. 2, pp. 81–90, 2020.

- [187] L. Petridis, R. Schulz, and J. C. Smith, "Simulation analysis of the temperature dependence of lignin structure and dynamics," *Journal of the American Chemical Society*, vol. 133, no. 50, pp. 20 277–20 287, 2011.
- [188] E. A. Capanema, M. Y. Balakshin, and J. F. Kadla, "A comprehensive approach for quantitative lignin characterization by nmr spectroscopy," *Journal of Agricultural and Food Chemistry*, vol. 52, no. 7, pp. 1850–1860, 2004.
- [189] A. Fujimoto, Y. Matsumoto, C. Hou-Min, and G. Meshitsuka, "Quantitative evaluation of milling effects on lignin structure during the isolation process of milled wood lignin," *Journal of Wood Science*, vol. 51, no. 1, pp. 89–91, 2005.
- [190] F. Jensen, *Introduction to Computational Chemistry*. John wiley & sons, 2017.
- [191] P. M. Morse, "Diatomic molecules according to the wave mechanics. ii. vibrational levels," *Physical review*, vol. 34, no. 1, p. 57, 1929.
- [192] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. a. Swaminathan, and M. Karplus, "Charmm: a program for macromolecular energy, minimization, and dynamics calculations," *Journal of computational chemistry*, vol. 4, no. 2, pp. 187–217, 1983.
- [193] A. K. Rappé, C. J. Casewit, K. Colwell, W. A. Goddard III, and W. M. Skiff, "Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations," *Journal of the American chemical society*, vol. 114, no. 25, pp. 10 024–10 035, 1992.
- [194] D. Marx and J. Hutter, "Ab initio molecular dynamics: Theory and implementation," *Modern methods and algorithms of quantum chemistry*, vol. 1, no. 301-449, p. 141, 2000.
- [195] A. van Duin, "Reaxff user manual," *California Institute of Technology, Materials and Process Simulation Center*, 2002.
- [196] R. J. Deeth, A. Anastasi, C. Diedrich, and K. Randell, "Molecular modelling for transition metal complexes: Dealing with d-electron effects," *Coordination Chemistry Reviews*, vol. 253, no. 5-6, pp. 795–816, 2009.
- [197] P.-O. Norrby and P. Brandt, "Deriving force field parameters for coordination complexes," *Coordination Chemistry Reviews*, vol. 212, no. 1, pp. 79–109, 2001.
- [198] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," *Physical review*, vol. 136, no. 3B, p. B864, 1964.
- [199] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," *Physical review*, vol. 140, no. 4A, p. A1133, 1965.
- [200] L. H. Thomas, "The calculation of atomic fields," in *Mathematical proceedings of the Cambridge philosophical society*, vol. 23, no. 5. Cambridge University Press, 1927, pp. 542–548.

- [201] E. Fermi, "Un metodo statistico per la determinazione di alcune priorieta dell'atome," *Rend. Accad. Naz. Lincei*, vol. 6, no. 602-607, p. 32, 1927.
- [202] A. Laio and M. Parrinello, "Escaping free-energy minima," *Proceedings of the National Academy of Sciences*, vol. 99, no. 20, pp. 12 562–12 566, 2002.
- [203] M. Iannuzzi, A. Laio, and M. Parrinello, "Efficient exploration of reactive potential energy surfaces using car-parrinello molecular dynamics," *Physical Review Letters*, vol. 90, no. 23, p. 238302, 2003.
- [204] "Cpmd, <http://www.cpmd.org/>, copyright ibm corp 1990-2019, copyright mpi für festkörperforschung stuttgart 1997-2001."
- [205] N. Troullier and J. L. Martins, "Efficient pseudopotentials for plane-wave calculations," *Physical review B*, vol. 43, no. 3, p. 1993, 1991.
- [206] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," *Physical review letters*, vol. 77, no. 18, p. 3865, 1996.
- [207] Y. Zhang and W. Yang, "Comment on "generalized gradient approximation made simple"," *Physical Review Letters*, vol. 80, no. 4, p. 890, 1998.
- [208] M. Zhang, Z. Geng, and Y. Yu, "Density functional theory (dft) study on the dehydration of cellulose," *Energy & fuels*, vol. 25, no. 6, pp. 2664–2670, 2011.
- [209] S. Nosé, "A unified formulation of the constant temperature molecular dynamics methods," *The Journal of chemical physics*, vol. 81, no. 1, pp. 511–519, 1984.
- [210] W. G. Hoover, "Canonical dynamics: Equilibrium phase-space distributions," *Physical review A*, vol. 31, no. 3, p. 1695, 1985.
- [211] W. Humphrey, A. Dalke, and K. Schulten, "Vmd: visual molecular dynamics," *Journal of molecular graphics*, vol. 14, no. 1, pp. 33–38, 1996.
- [212] A. Laio and F. L. Gervasio, "Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science," *Reports on Progress in Physics*, vol. 71, no. 12, p. 126601, 2008.
- [213] B. Ensing, A. Laio, M. Parrinello, and M. L. Klein, "A recipe for the computation of the free energy barrier and the lowest free energy path of concerted reactions," *The journal of physical chemistry B*, vol. 109, no. 14, pp. 6676–6687, 2005.
- [214] S. H. Mushrif, A. D. Rey, and G. H. Peslherbe, "Energetics and dynamics of hydrogen adsorption, desorption and migration on a carbon-supported palladium cluster," *Journal of Materials Chemistry*, vol. 20, no. 46, pp. 10 503–10 510, 2010.
- [215] M. Appell, J. Willett, and F. A. Momany, "Dft study of α - and β -d-mannopyranose at the b3lyp/6-311++ g** level," *Carbohydrate research*, vol. 340, no. 3, pp. 459–468, 2005.

- [216] F. A. Momany, M. Appell, J. Willett, U. Schnupf, and W. B. Bosma, “Dft study of α - and β -d-galactopyranose at the b3lyp/6-311++ g** level of theory,” *Carbohydrate research*, vol. 341, no. 4, pp. 525–537, 2006.

Appendix A

Molecular Modeling Techniques

The underlying principle behind every model is to solve a set of mathematical equations that will in some capacity describe physical reality. Depending on the time vs length scale, computational modeling and the underlying mathematical equations that describe the system vary, depending on the phenomena of interest. Typically, the mathematics associated with the models are so sophisticated an analytical solution is practically impossible. Therefore, numerical methods are used to obtain results, where computers can be employed to perform the iterations necessary with these methods. For the objectives of this thesis, electronic structure and atomistic modeling are the primary tools required, therefore, will be the only models discussed in further depth. The various time vs length scale of electronic structure and atomistic modeling methods are depicted in Figure A.1.

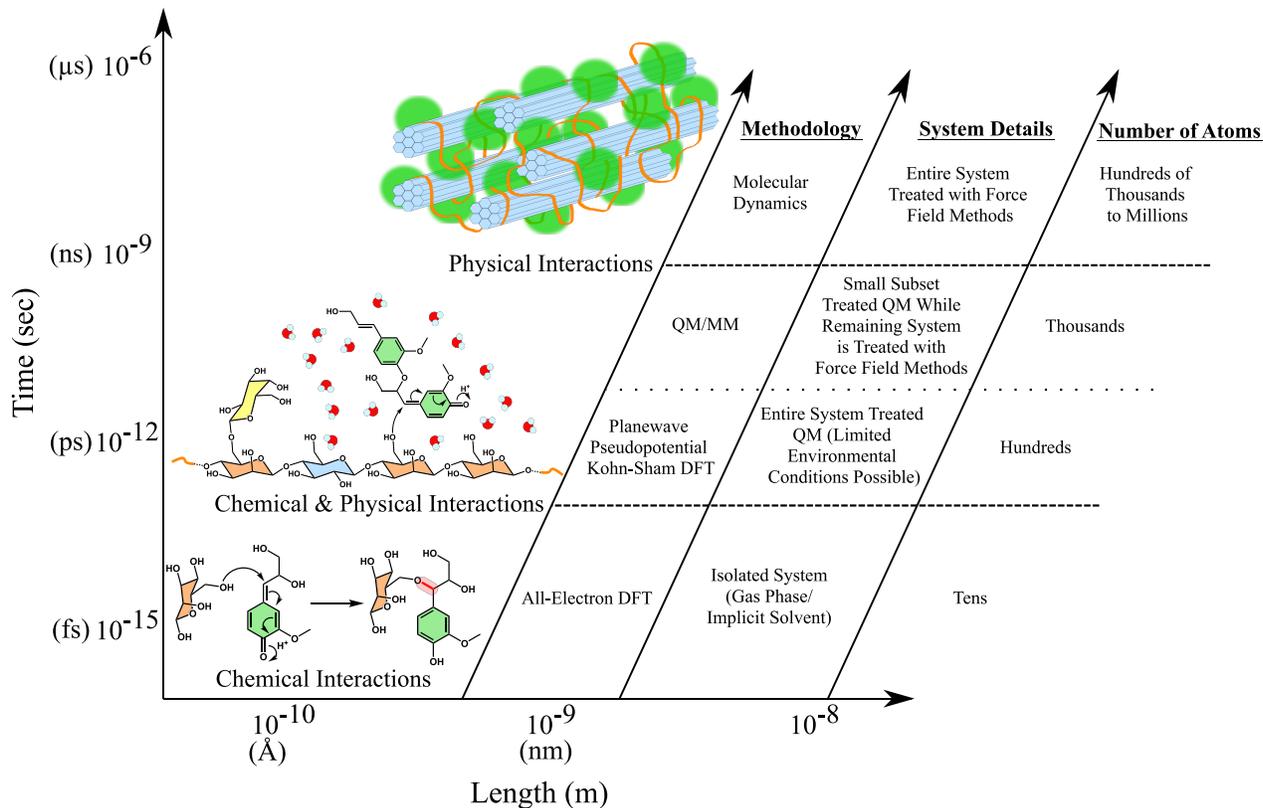


Figure A.1: The multiscale nature of the research problem pertaining to understanding the biomass microstructure and to develop solvent based deconstruction methods. Modeling tools that can be used for different time and length scales are also listed.

A.1 Molecular Mechanics

In atomistic models the most fundamental particles are atoms. The entire system is described as a series of 'balls' (atoms) of different sizes and 'springs' (bonds) of different lengths and stiffness, giving rise to the term molecular mechanics. The potential energy is calculated using a force field, which is simply a set of equations and parameters used to describe the bonded and non-bonded interactions within the system. The dynamics of the system are modeled by integrating the classical Newton's equations of motion.

Bonded interactions include three distinct deformations: stretching, bending and twisting (torsion) of the bonds. Non-bonded interactions include van der Waals as well as electrostatic interactions.

$$E_{MM} = \underbrace{E_{Stretch} + E_{Bend} + E_{Torsion}}_{E_{Bonded}} + \underbrace{E_{vdW} + E_{Electrostatic}}_{E_{Non-Bonded}}$$

Where $E_{Stretch}$ is the energy associated with stretching a bond between two atoms, E_{Bend} is the energy associated with bending an angle formed by three bonded atoms and $E_{Torsion}$ is the energy associated with the twisting of a bond. Additionally, E_{vdW} and $E_{Electrostatic}$ are the energies associated with the van der Waals and the electrostatic interactions between two

atoms, respectively. Figure A.2 visually illustrates the terms considered when calculating the force field energy. Since the energy is a function of the atomic coordinates or conformation, every molecule has an associated 'free energy landscape' as a function of the conformation of the molecule. Minimizing the energy of the system (E_{MM}) will give rise to a stable conformation, however, whether that conformation is a local minimum or global minimum in the free energy landscape requires extensive sampling and/or sophisticated algorithms.

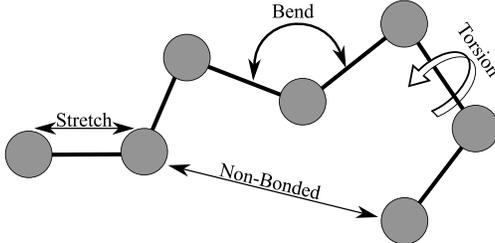


Figure A.2: A visual depiction of the bonded and non-bonded energy terms accounted for in a force field.

The energy associated with stretching two bonded atoms can be written as a Taylor series expansion around the natural bond length. Assuming the stretching energy of interest is between atoms A and B is given as [190],

$$E_{Stretch}^{AB} = E_o + \left. \frac{dE}{dl} \right|_{l^o} (l^{AB} - l^o) + \frac{1}{2!} \left. \frac{d^2E}{dl^2} \right|_{l^o} (l^{AB} - l^o)^2 + \dots \quad (\text{A.1})$$

The derivatives are evaluated at l^o with the first derivative being zero and E_o set to zero since it is the zero point on the energy scale. Terminating the expansion at the second order results in Equation A.1 be simply written as

$$E_{Stretch}^{AB} = \frac{1}{2!} \left. \frac{d^2E}{dl^2} \right|_{l^o} (l^{AB} - l^o)^2 = K_{Stretch} (l^{AB} - l^o)^2 = K_{Stretch} (\Delta l^{AB})^2 \quad (\text{A.2})$$

Where, $K_{Stretch}$ is the force constant. Equation A.2 is in the form of a harmonic oscillator and is sufficient for determining most equilibrium geometries. However, using the harmonic approximation for certain strained and crowded systems demonstrate significant deviation from experimental results, indicating the need for an improved expression. A straight forward improvement would be retaining more terms in the Taylor expansion, however, this adds more parameters. Additionally, the polynomial expression does not have the correct limiting behavior, e.g. if a poor initial guess is used the molecule can "fly apart". A simpler function that appropriately captures this limiting behavior is the Morse potential [191].

$$E_{Stretch}^{AB}(\Delta l^{AB}) = D \left(1 - e^{-\alpha \Delta l^{AB}} \right)^2 \quad (\text{A.3})$$

$$\alpha = \sqrt{\frac{K}{2D}}$$

Where, D is the dissociation energy and α is related to the force constant. The Morse function reproduces experimental behavior quite accurately over a range of distances, especially for the region of interest around the natural bond length.

The energy associated with bending an angle between three atoms, A-B-C, can be expressed as a Taylor series expansion around a natural bond angle. Terminating the expansion at the second order yields the harmonic approximation.

$$E_{Bend}^{ABC} = K^{ABC}(\theta^{ABC} - \theta_o^{ABC})^2 \quad (\text{A.4})$$

Similar to the stretching energy, the harmonic approximation is suitable for most applications, however, when higher accuracy is required, retaining the third order term is the next improvement. Assigning constants beyond the third order can rarely be done confidently due to limited experimental data, however, this is seldom required or done in practice.

The energy associated with twisting around a B-C bond in a four atom sequence, A-B-C-D, (torsion energy) where A-B, B-C and C-D are bonded atoms is fundamentally different from the stretching energy, $E_{Stretch}$, as well as the bending energy, E_{Bend} , in three aspects. Firstly, the rotational barrier has contributions from the non-bonded as well as the bonded interactions. Secondly, the torsion energy is periodic by nature, i.e., after a 360° rotation the energy should return to the same value. Thirdly, the energy associated with rotating a molecule is typically low, therefore, large deviations from the minimum can occur, making a Taylor expansion impractical. To adequately capture the periodic nature of the torsional energy, $E_{Torsion}$, it is written as a Fourier series.

$$E_{Torsion}^{ABCD} = \sum_{n=1} V_n \cos(n\omega) \quad (\text{A.5})$$

Where, ω , is the torsion angle, i.e., if looking down the B-C bond, ω is the angle formed between the the A-B and C-D bond. The $n = 1$ term describes the rotation that is periodic by 360° , $n = 2$ describes the rotation that is periodic by 180° , $n = 3$ describes the rotation that is periodic by 120° and so on. The V_n constant determines the size of barrier for rotation around the bond, which can be zero depending on the system of interest.

The non-bonded energy is split into the van der Waals energy as well as the electrostatic energy. The van der Waals energy describes the attraction or repulsion between non-bonded atoms, and can be thought of as the non-polar part of the interaction, i.e., the interaction not related to charges. The van der Waals energy at large interatomic distances becomes zero and at short interatomic distances it becomes very repulsive due to the overlap of electron clouds. However, at intermediate distances there is a slight attraction between electron clouds as a result of induced dipole-dipole interactions. The van der Waals energy is commonly expressed as the Lennard-Jones potential, which satisfies these physical requirements.

$$E_{vdW}^{AB} = \varepsilon \left[\left(\frac{R_o}{R^{AB}} \right)^{12} - 2 \left(\frac{R_o}{R^{AB}} \right)^6 \right] \quad (\text{A.6})$$

Where R_o is the minimum energy distance, ε is the depth of the minimum and R^{AB} is the interatomic distance. The electrostatic energy is a result of the rearrangement of electrons when individual atoms form a molecule, creating positive and negative parts to the molecule. The electrostatic energy between atoms is commonly given by the Coulomb potential.

$$E_{Electrostatic}^{AB} = \frac{Q^A Q^B}{\kappa R^{AB}} \quad (\text{A.7})$$

Where Q is the atomic charge of the species and κ is the dielectric constant.

Once the desired mathematical form of the energy terms are established, the parameterization of these equations is necessary. This is usually done from a set of reference data,

i.e., data for the properties of interest obtained from experiments and/or electronic structure calculations. A key consideration in the parameterization becomes the balance between accuracy and generalizability. All force fields extend beyond the data used to fit the parameters, however, the extent of generalizability compared to the relative accuracy, whether quantitative or qualitative, begins to vary. For example, the CHARMM force field developed by Professor Martin Karplus [192] is designed specifically for modeling proteins, therefore, one can expect to reproduce accurate quantitative data when modeling proteins. However, if this force field was applied to a different system such as a metallic system then the results would likely be erroneous and even qualitative agreement with reality may not be well represented. Whereas, the Universal Force Field (UFF) developed by Professor William Goddard [193] is designed for all elements, therefore, one could expect relatively good qualitative agreement across many systems, but quantitative agreement would be unlikely for many systems. As a result, the nature of the system, phenomena of interest as well as the desired accuracy are starting points for choosing an appropriate force field. Many force fields have been developed to date and primarily differ in the functional form of the energy terms, the number of additional energy terms, and/or the reference data used to parameterize.

Force field methods are capable of modeling large molecular systems with thousands of atoms such as proteins, DNA, membranes, etc. at a reasonable computational cost. Additionally, it is possible to model solvation effects with explicit solvent molecules in condensed phase systems or even crystal packing can be studied [190]. Many pharmaceutical companies utilize these techniques to predict how drugs will interact with target compounds, accelerating the drug discovery process. However, these techniques are also accompanied by certain limitations:

(1) Inability to model unusual systems [190, 194]. Force fields are designed to perform well for systems when an abundant amount of information is available for the parameterization of similar systems. However, when molecules are “exotic” or there is limited information on the system under investigation then the force field may perform poorly.

(2) Diverse types of molecules [190, 194]. Force fields have a balance between generalizability and accuracy. Including diverse types of molecules in the parameterization process may improve the generality of the force field, however, this may reduce the accuracy in specific applications. Moreover, depending on the functional form chosen additional data may not be beneficial; alternatively modifying the functional form or adding more terms may reduce computational efficiency or in some cases remove the cancellation of error that is present in simpler functional forms. Therefore, force field methods are designed for specific types of molecules and have difficulty in representing diverse systems in one force field.

(3) Inability to model chemical reactions [194, 195]. Force fields require the specification of the atom types, interactions between the atoms, i.e., whether atoms are bonded or non-bonded, as well as the geometry of the system. The type of force field thus the functional form of the energy terms is governed by the atom types and interactions specified. However, during a chemical interaction the atom types and interactions between them change as a result of breaking and forming new bonds. For example, a sp^2 carbon may become sp^3 from an addition to a double bond. Therefore, the energy functions that describe the reactant and product are different and the force field method would fail to model such a phenomena.

(4) Metal systems [190, 196, 197]. Bonding within metal system is drastically more complex than in organic compounds. In pure metallic systems the bonding may vary depending on

the metal cluster size, whereas, in metal-ligand complexes coordination bonds are formed. Therefore, varied coordination numbers and geometries as well as varied oxidation states and electronic structures make implementing force field methods for metal systems a laborious task.

A.2 Electronic Structure Methods

A.2.1 Hartree-Fock

Self Consistent Field (SCF) Procedure

Start with the time independent Schrödinger equation under the Born-Oppenheimer approximation.

$$H_e \Psi_e(r_e, R_n) = E_e \Psi_e(r_e, R_n)$$

The coordinates of the nuclei are inputted and the number of electrons are specified, i.e., charge and multiplicity of the system. Now the molecular orbitals (wavefunction) needs a mathematical form before we can compute the energy. According to postulate 5 of quantum mechanics, the set of eigenfunctions of the operator will form a complete set of linearly independent functions.

$$\Psi = \sum c_j \Psi_j$$

Therefore, the molecular orbitals are expressed as a linear combination of atomic orbitals (LCAO).

$$\chi = c_1 \phi_1 + c_2 \phi_2 + c_3 \phi_3 + \dots + c_n \phi_n$$

Where χ is the molecular orbital, ϕ is an atomic orbital and c_n are coefficients associated with the atomic orbitals. And the molecular orbitals can be expressed as a Slater determinant which ensures the wavefunction follows the antisymmetric constraint.

$$\Psi(1, 2, \dots, N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(1) & \cdots & \chi_N(1) \\ \vdots & \ddots & \vdots \\ \chi_1(N) & \cdots & \chi_N(N) \end{vmatrix} = |\chi_1(1) \chi_2(2) \dots \chi_N(N)|$$

Now, if the atomic orbitals are obtained, there is a clear pathway to obtain the wavefunction for the system. To mathematically represent the atomic orbitals, we need a set of clearly defined equations to approximate them. The equations to approximate the atomic orbitals are referred to as basis functions. Essentially any form of basis function can be used; however, the more accurately it represents the physics the system, the better. The two most typically used are the Slater type orbital (STO) and the Gaussian type orbital (GTO).

$$\varphi_{\zeta, n, l, m}(r, \theta, \phi) = NY_{l, m}(\theta, \phi) r^{n-1} e^{-\zeta r} (STO)$$

$$\varphi_{\zeta, n, l, m}(r, \theta, \phi) = NY_{l, m}(\theta, \phi) r^{2n-2-l} e^{-\zeta r^2} (GTO)$$

STOs more accurately represent the physics of the system, specifically near a nucleus, however, GTOs (even multiple GTOs) are more computationally friendly, i.e., easier to compute. Therefore, we approximate the STOs with multiple GTOs to save computational time. Now, how the basis functions are implemented and combined to represent the atomic orbitals need

to be chosen; or in other words, the basis set needs to be selected. Although a clear mathematical form for the atomic orbitals is achieved there are unknowns associated with the GTOs as well as the coefficients associated with every atomic orbital. However, the variation principle states the calculated energy will always be greater than the true energy. Therefore, if the unknowns associated with the GTOs as well as the coefficients associated with the atomic orbitals are chosen in such a way to minimize the energy than those unknowns and coefficients are the “correct” ones for the system. Typically, the unknowns associated with the GTOs have been well optimized previously for the selected basis set and further considerations for them are not required. Therefore, optimizing these unknowns are not included in the following discussion. Furthermore, an initial guess can be chosen for the atomic orbital coefficients and the wavefunction can be constructed.

The Hamiltonian for the system is given by:

$$H = \underbrace{\sum_i^N -\frac{1}{2}\nabla^2 - \sum_i^N \sum_\alpha \frac{Z_\alpha}{|r_i - r_\alpha|}}_{h_i} + \underbrace{\sum_i^N \sum_j^N \frac{1}{|r_i - r_j|}}_{g_{ij}}$$

Where the first term represents the kinetic energy of N electrons, the second term represents the interaction of N electrons with α nuclei and the third term represent the electron – electron interactions. Then the energy for the system is,

$$E = \langle \Psi | H | \Psi \rangle$$

Substituting the simplified Hamiltonian ($h_i + g_{ij}$) in yields,

$$E = \left\langle \Psi \left| \sum_i^N h_i \right| \Psi \right\rangle + \left\langle \Psi \left| \sum_i^N \sum_j^N g_{ij} \right| \Psi \right\rangle$$

Writing the wavefunction as a Slater determinant and expanding yields,

$$\sum_i^N \left\langle \chi_i \left| -\frac{\nabla_i^2}{2} \right| \chi_i \right\rangle - \sum_i^N \left\langle \chi_i \left| -\frac{Z_\alpha}{r_{i\alpha}} \right| \chi_i \right\rangle + \sum_i^N \sum_j^N (J_{ij} - K_{ij})$$

Where, J_{ij} is the Coulomb integral considering electron repulsion and K_{ij} is the Exchange integral which arises from the antisymmetric constraint.

$$J_{ij} = \left\langle \chi_1(1) \chi_2(2) \left| \frac{1}{r_{12}} \right| \chi_1(1) \chi_2(2) \right\rangle$$

$$K_{ij} = \left\langle \chi_1(1) \chi_2(2) \left| \frac{1}{r_{12}} \right| \chi_2(1) \chi_1(2) \right\rangle$$

Which if the system (molecule) is neutral every orbital is doubly occupied, then $N = 2n$. And through some further rearrangement the energy of the system can be calculated with:

$$E = \sum_i^n 2h_{ii} + \sum_{i,j}^n (2J_{ij} - K_{ij})$$

However, since the coefficients used were initially guessed an algorithm needs to be in place to efficiently optimize the coefficients to determine the lowest energy for the selected nuclear coordinates, i.e., apply the variation principle to obtain the lowest energy for the system

thus yielding the “best” coefficients for the system. To do this the method of Lagrange multipliers is employed. The maximum or minimum of function, $f(x)$, which is subject to a constraint $g(x) = 0$, can be obtained by finding the stationary points of the Lagrangian function $[\mathcal{L}(x, \lambda)]$ given as: $\mathcal{L}(x, \lambda) = f(x) - \lambda g(x)$

In this case, varying the electronic energy (E) as a function of the atomic orbital coefficients (c_1, c_2, \dots, c_n) and equating it to zero. If the energy is a minimum the derivative of the energy with respect to the coefficients will be zero.

$$\begin{aligned} \mathfrak{J}[\chi] &= \sum_i^n 2h_{ii} + \sum_{i,j}^n (2J_{ij} - K_{ij}) - \sum_i^n 2\lambda_{ij} [\langle \chi_i | \chi_j \rangle - \delta_{ij}] \\ \Rightarrow \frac{\delta \mathfrak{J}}{\delta \chi_i} = 0 &\Rightarrow \left[h_i + \sum_j^n (2J_{ij} - K_{ij}) \right] \chi_i = \sum_j^n \lambda_{ij} \chi_j \end{aligned}$$

Where the Lagrange multipliers (λ_{ij}) are introduced because the energy minimization must be performed under the constrain that the molecular orbitals remain orthogonal and normalized. Since the wavefunction is a determinant the Lagrange multipliers can be diagonalized to yield the orbital energies (ε_i) and this gives the Hartree-Fock equations.

$$\left[h_i + \sum_j^n (2J_j - K_j) \right] \chi_i = \mathbb{F}_i \chi_i = \varepsilon_i \chi_i$$

As discussed previous, MOs are simply a LCAO,

$$\chi_i = \sum_v^M c_{iv} \varphi_v$$

$$\mathbb{F}_i \chi_i = \mathbb{F}_i \sum_v^M c_{iv} \varphi_v = \varepsilon_i \sum_v^M c_{iv} \varphi_v$$

Utilizing a specific basis function and integrating yields the Roothan – Hall matrix equations:

$$\mathbb{F}\mathbf{c} = \mathbb{S}\mathbf{c}\varepsilon$$

$$F_{\alpha\beta} = \langle \varphi_\alpha | \mathbb{F} | \varphi_\beta \rangle$$

$$S_{\alpha\beta} = \langle \varphi_\alpha | \varphi_\beta \rangle$$

Where \mathbb{F} is the Fock Matrix and \mathbb{S} is the Overlap Matrix. The Fock Matrix can be solved and diagonalized to yield the new coefficients for the MOs, which then update the initially guessed coefficients. This procedure is repeated until the “guessed coefficients” match the coefficients obtained from the diagonalized Fock Matrix. This means the system is at the minimum energy and the coefficients are the “best” for the particular nuclear coordinates inputted at the start. The overall procedure is called the self-consistent-field (SCF) procedure.

Within the Hartree-Fock (HF) method the electron coulombic correlation is not accurately captured, because of the way the wavefunction is written in the Slater determinant. In other words, within the HF method electrons will not feel the explicit electron – electron interactions; rather they feel an average effect of the electron repulsion which gives rise to electron correlation errors. Post HF methods such as configuration interaction (CIS, CISD(T), Full CI, etc.) , Moller-Plesset Perturbation theory (MP2, MP4, etc.) and Couple Cluster Theory (CCSD, CCSD(T)), etc. all improve the description of electron correlation, however, this increased accuracy results in an increase in computational cost.

A.2.2 Density Functional Theory (DFT)

For every particle moving in a force field, there is an associated wavefunction that entirely describes the particle. For wavefunction theories, Hartree Fock (HF) and post HF methods, for a set of nuclear coordinates the wavefunction needs to be determined for every electron to calculate all the physical observables such as energy, density, etc. In other words, when nuclear coordinates and number of electrons (charge) of the system are specified, an external potential is specified. In this potential the wavefunction is needed to calculate the physical observable, however, neither the wavefunction or physical observable are known. This requires many iterations and the use of the variational principle to “solve” the system. However, the wavefunction and the associated procedure is complicated and if accurate electron-electron interactions are desired (post HF) this procedure becomes computationally expensive to model any significant sized system. Therefore, there was a desire to simplify this procedure. Instead of using a wavefunction to completely describe the system, a tangible physical observable that describes the system would be easier to work with.

In 1964, Pierre Hohenberg and Walter Kohn [198] suggested that electron density (or electron probability) was a physical observable that could describe the system. This electron density can be thought of as the probability of finding an electron in a finite volume element and all other electrons elsewhere. Therefore, when the external potential the electrons see is specified it is possible to determine the electron density of the system via the wavefunction. However, as mentioned previous the wavefunction is complicated and circumventing it would be beneficial, therefore, the reverse is desired, i.e., can the external potential be obtained from the electron density? This is exactly the question that Hohenberg and Kohn addressed in 1964 when they showed a proof by contradiction that a unique potential exists for a given electron density. Additionally, previously, the energy of the system was determined by minimizing the energy with respect to the coefficients within the wavefunction. However, Hohenberg and Kohn also showed minimizing the energy of a multi electron system with respect to the electron density was possible. The shortcoming of the work was that the functional form of electron density was still unknown. It was only proven once that was determined the unique potential of the system could be obtained and the energy of the system could be calculated by minimizing the energy with respect to the electron density.

This shortcoming was (mostly) rectified by Kohn and Sham [199] later on by using the work of Thomas-Fermi in 1927 [200, 201]. Thomas-Fermi took a non-interacting system, i.e., electrons have a kinetic energy and interact with a positive charge, however, the electrons do not interact with each other. Therefore, the energy term has two components the kinetic energy of the electrons and the potential energy of the electrons interacting with a positive charge (nucleus). The potential energy term is readily acquired, however, the kinetic energy had to be derived. To do this Thomas-Fermi took a uniform electron gas, i.e., an infinite number of electrons moving in an infinite space interacting with a uniformly distributed positive charge, and derived the kinetic energy per electron. Therefore, this gave rise to an explicit expression for the total energy which could be written in terms of electron density. Therefore, the work of Thomas-Fermi in 1927 [200, 201] as well as of Hohenberg and Kohn in 1964 [198] likely provided the foundation for the implementation of Kohn-Sham density functional theory (DFT) that is used today. As noted previous, Kohn-Sham mostly rectified the shortcoming of the Hohenberg and Kohn work, this is because the wavefunction is not

completely bypassed, it is however, greatly simplified. To start Kohn-Sham took a fictitious system of non-interacting electrons, where the Hamiltonian could be written simply as the kinetic energy of the electrons (first term in Equation (A.8)) and potential energy of the electron-nucleus interactions (second term in Equation (A.8)).

$$h_{ks} = -\frac{1}{2}\nabla^2 + \nu_{ks}(\vec{r}) \quad (\text{A.8})$$

Now the potential energy term for the electron-nucleus interactions can be written generally as a Kohn-Sham potential (ν_{ks}) which corresponds to the non-interacting system. Since the system is non-interacting it is easy to solve. However, the Kohn-Sham potential is defined in such a way that the electron density that is calculated using this Kohn-Sham potential (using Kohn-Sham orbitals/wavefunction) is the same electron density of an interacting system with a ‘real’ potential. Therefore, the electron density provides a ‘connection’ between the fictitious, non-interacting system and the real system. This ‘connection’ provides the basis for DFT since a non-interacting system is easy to solve, therefore, it is possible to use the non-interacting system to help solve the interacting system, which is described in detail next. Now for the wavefunction of the real system ψ the total energy can be calculated by taking the expectation value of the Hamiltonian which is composed of three terms: the kinetic energy operator for the real system, the real potential corresponding to electron-nucleus interactions and the operator for electron-electron interactions for the real system.

$$E = \langle \psi | T + \nu + V_{ee} | \psi \rangle \quad (\text{A.9})$$

Which can be written as

$$E = \underbrace{\langle \psi | T | \psi \rangle}_{T_K} + \int d^3r \rho(\vec{r}) \nu(\vec{r}) + \langle \psi | V_{ee} | \psi \rangle \quad (\text{A.10})$$

Now the real kinetic energy T_K is not known, however, we can easily obtain the kinetic energy of the fictitious non-interacting system, T_s , therefore, we simply rewrite Equation (A.10) by adding and subtracting T_s . This splits the kinetic energy into two parts, one that can be easily solved for the non-interacting system and one that is a correction term for the difference in kinetic energy between the real system and the non-interacting system.

$$E = T_s + (T_K - T_s) + \int d^3r \rho(\vec{r}) \nu(\vec{r}) + \langle \psi | V_{ee} | \psi \rangle \quad (\text{A.11})$$

Now let’s assume the electron-electron interactions can be approximated to a certain extent by repulsion of a classical charge distribution, i.e., electron density clouds interact through electrostatic interactions divided by the distance between them, similar to simple coulomb interactions, however, the electrons are not point charges. This can be written as a functional of electron density

$$U[\rho] = \frac{1}{2} \int d^3r d^3r' \frac{\rho(\vec{r}) \rho(\vec{r}')}{|\vec{r} - \vec{r}'|} \quad (\text{A.12})$$

Similar to rewriting Equation (A.10), $U[\rho]$ (Equation (A.12)) can be added and subtracted from Equation (A.11).

$$E = T_s + (T_K - T_s) + \int d^3r \rho(\vec{r}) \nu(\vec{r}) + \frac{1}{2} \int d^3r d^3r' \frac{\rho(\vec{r}) \rho(\vec{r}')}{|\vec{r} - \vec{r}'|} + (\langle \psi | V_{ee} | \psi \rangle - U[\rho]) \quad (\text{A.13})$$

With some minor rearrangement Equation (A.13) becomes

$$E = T_s + \int d^3r \rho(\vec{r}) \nu(\vec{r}) + \frac{1}{2} \int d^3r d^3r' \frac{\rho(\vec{r}) \rho(\vec{r}')}{|\vec{r} - \vec{r}'|} + \underbrace{(T_K - T_s) + (\langle \psi | V_{ee} | \psi \rangle - U[\rho])}_{E_{XC}(\rho)} \quad (\text{A.14})$$

Where $E_{XC}(\rho)$ is the Exchange-Correlation term which holds the correction for kinetic energy as well as electron-electron interaction energy. The first term, kinetic energy for the fictitious system can be determined through the Kohn-Sham potential and the second/third term are written explicitly in terms of density. In summary, Equation (A.9) which corresponds to the real system, was simply rewrote with terms from the fictitious system to yield Equation (A.14). Therefore, Equation (A.9) and Equation (A.14) are equivalent. In other words, if Equation (A.9) yields the exact energy, Equation (A.14) yields an approximate energy plus a correctional term. Now two ‘obstacles’ need to be resolved before Equation (A.14) can be used. First, the non-interacting system needs to be possible to determine since we modified Equation (A.9) under the assumption everything of the non-interacting system is easily calculated. Second, the Exchange-Correlation term needs an explicit functional that we can utilize. For the first obstacle, the Kohn-Sham potential, ν_{ks} , is required then the non-interacting system is possible to solve. We begin with the energy of the fictitious system, or, the Kohn-Sham energy which can be determined through the use of the variational principle, where the energy is minimized with respect to the electron density. Then the derivative of the Kohn-Sham energy with respect to density will be zero. Similarly, the real energy, Equation (A.14), can be obtained with the use of the variational principle by minimizing the energy with respect to the electron density; and the derivative of the real energy with respect to density will also be zero. Equating the derivatives of these two equations yields:

$$\nu_{ks}(\vec{r}) = \nu(\vec{r}) + \int d^3r' \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} + \frac{\delta E_{XC}}{\delta \rho} \quad (\text{A.15})$$

Now the first ‘obstacle’ is resolved, since we have an explicit expression for the Kohn-Sham potential which will allow the details of the fictitious system defined earlier to be calculated. However, Equation (A.15) also reinforces the second ‘obstacle’, the Exchange-Correlation term needs to be known. Now for an explicit expression for the Exchange-Correlation term, which accounts for the difference in the actual kinetic energy and the kinetic energy of the fictitious non-interacting system, as well as the difference in the actual electron-electron interactions with the approximation made in Equation (A.12). Typically, the kinetic energy correction is quite minimal, however, the electron-electron interaction correction is significant; both corrections taken together account for less than 10% of the energy of the system. The most fundamental approximation for the Exchange-Correlation term, is the local density approximation (LDA). In LDA the Exchange-Correlation energy of an actual system, i.e., not a non-interacting system, at any point with electron density $\rho(r)$ can be approximated as the Exchange-Correlation energy of a UEG with the same electron density as the position selected in the actual system. Therefore, the Thomas-Fermi energy formulation is extended to include Coulomb interaction energies of electron-electron interactions, which improves the results from a completely non-interacting system, however, is still not accurate enough for many applications. Extensions improving upon this include the generalized gradient approximation (GGA) formulations, which considers the density as well as the density gradient. The GGA functionals can also be extended to consider additional terms, which form the basis of higher level functionals such as, meta-GGA (M-GGA) and hybrid-GGA (H-GGA). However, in any case some sort of analytical expression is obtained for the Exchange-Correlation term,

$E_{XC}(\rho)$, allowing it to be computed in some way from the electron density. Now with the Exchange-Correlation expression obtained, it needs to be recognized that the Kohn-Sham DFT formulation results in a similar procedure as seen with the self-consistent field (SCF) procedure within the HF wavefunction framework. That is, if a trial density is used, the Exchange-Correlation can be determined and the Kohn-Sham potential can be calculated, which can be used to construct the Kohn-Sham orbitals/wavefunction which yields a new electron density. Therefore, this process is repeated until the electron density remains constant from one iteration to the next and then the energy of the system can be calculated (Equation (A.14)). Note, the Kohn-Sham orbitals/wavefunction is calculated, therefore, as mentioned, this procedure does not completely bypass the formulation of the wavefunction, however, it is the wavefunction for a non-interacting system making it much simpler than what is observed in the HF framework. Additionally, it also considers electron-electron interactions which is a huge improvement compared to HF. And as observed with the overall iterative procedure, the computational cost is similar to HF.

A.2.3 Carr-Parrinello Molecular Dynamics & Metadynamics

Hartree-Fock (HF), post-HF (CI, MP, CC, etc.) and density functional theory (DFT) are capable of optimizing a given set of nuclear coordinates using local optimization schemes, as such, the starting nuclear coordinates determine which minimum is found and it is unknown if the configuration corresponds to a global minimum. Therefore, there is a desire to use a similar implementation as molecular dynamics (MD) and supply kinetic energy to the system to overcome energy barriers and explore more of the free energy landscape. Since one of the primary differences between MD and *ab-initio* methods are simply how the potential energy and force are calculated, it is possible to calculate the potential energy and force using electronic structure methods and update the positions of nuclei (which are treated classically under the Born-Oppenheimer (BO) assumption) in a similar fashion as found in MD. Therefore, the Lagrangian can be used to obtain the equations of motion for *Ab-Initio* or Born-Oppenheimer molecular dynamics (AIMD or BOMD) and is written as,

$$\mathcal{L}_{BO} = \frac{1}{2} \sum_i m_i \dot{R}_i^2 - \langle \psi | H | \psi \rangle \quad (\text{A.16})$$

The thermostat is included in Equation A.16 when the canonical ensemble is used. However, since the wave function is dependent on the external potential created by the nuclei, after every “MD step” when the nuclei are adjusted the wavefunction needs to be resolved. This becomes computationally expensive especially when thousands of iterations must typically be completed to have any meaningful MD results. To overcome this Carr and Parrinello (CP) published a paper in 1985 [143] that essentially provided a much more efficient way to obtain the wavefunction after every MD step. They proposed that once the wavefunction of the initial structure is obtained there was a way to “propagate” the wavefunction from the previous MD step and ensure it is similar to the wavefunction that would be obtained if it was recalculated at every step. This is done by considering the two different time scale separations, the classic/slow nuclei and the quantum mechanic/fast electrons. Since the wavefunction is dependent on the external potential created by the nuclei and the nuclei are propagated classically then the change in the wavefunction could also be propagated

classically. To implement this computationally an additional term is included in the Lagrangian to account for the “propagation” of the wavefunction, which is done by treating the wavefunction as a fictitious “classical particle” that possesses a kinetic energy, written as:

$$\mathcal{L}_{CP} = \frac{1}{2} \sum_i m_i \dot{R}_i^2 + \sum_j \frac{1}{2} \mu_j \langle \dot{\chi}_j | \dot{\chi}_j \rangle - \langle \psi | H | \psi \rangle + \text{Constraints} \quad (\text{A.17})$$

Where the first two terms on the right-hand side are the kinetic energy of the nuclei and wavefunction, respectively, and the third term is the potential energy. The fictitious kinetic energy of the wavefunction has no physical meaning; however, it is now possible to obtain the equations of motion and propagate the nuclei as well as the fictitious particle that is the wavefunction.

In some circumstances the system may not be able to overcome energy barriers and the simulation can be “stuck” in a local minimum. An option is to increase the temperature, giving the system enough energy to overcome large energy barriers; however, this can have some negative consequences such as side reactions, not following the minimum energy pathway, or being physically unrealistic. To overcome this, metadynamics [202, 203] can be coupled with CPMD to help the system explore the energy landscape. This is done by dropping potentials throughout the CPMD simulation which “fills” the energy landscape, allowing energy barriers to be crossed as well as reconstruction of the free energy landscape. The energy surface is a function of coordinates and configurations, which can be thought of as discrete variables such as bond distances, bond angles, dihedrals, coordination numbers, etc. These discrete variables that make up the configuration of the system are referred to as collective variables (CVs). Therefore, each system can be composed of dozens or even hundreds of these CVs which determine the configuration of the system and thus the energy. To implement metadynamics specific CVs are selected and the system is biased to provide additional energy to the specified CV. For example, if a reaction is under investigation the CVs need to be defined in such a way that they distinguish the reactants vs the products. Therefore, when additional energy is provided to the CV (reaction coordinate), the desired reaction will be modeled. If ξ is a vector of the CV which corresponds to the section of the energy surface that is under study then the CV will be the new dynamic variable introduced into the system. To do this mathematically the CPMD Lagrangian is extended to include terms to bias the CV as well as include a term that corresponds to the “potential dropping” to “fill up” the energy landscape

$$\mathcal{L}_{MTD} = \mathcal{L}_{CP} + \sum_{cv} \frac{1}{2} m_{cv} \dot{\xi}_{cv}^2 - \frac{1}{2} \sum_{cv} k_{cv} \left[\mathcal{S}(\vec{R}_{cv}) - \xi_{cv} \right]^2 + v_{cv}(t, \xi) \quad (\text{A.18})$$

The first term is simply the CP Lagrangian (Equation A.17). The second term is the kinetic energy of the collective variable and the third term is the harmonic restraining potential of the CV. The fourth term is the potential that is “dropped” to fill up the energy in the CV (ξ) space throughout the CPMD simulation. Therefore, the mass of the CV (m_{cv}) and the coupling constant (k_{cv}) determines how fast the CV is changed throughout the simulation, which can be manipulated. As the simulation proceeds potentials will be added to the simulation to “fill” the energy landscape, which will allow the CV space to be explored systematically and then reconstructed later.

Appendix B

Supporting Information for Chapter 2

B.1 Computational Details for Conformational Screening

The conformational sampling of the model LCC linkages was performed using Car-Parinello molecular dynamics [143] and metadynamics (CPMD-metadynamics). All CPMD and metadynamics calculations in this work were performed with the CPMD software, version 4.3.0 [204] along with the plane-wave-pseudopotential implementation of the Kohn-Sham density functional theory (DFT) [199]. A Martins-Trouiller (MT) pseudopotential [205] using the revised Perdew-Burke-Ernzerhof (revPBE) functional [206, 207] of the generalized gradient approximation (GGA) was used for the exchange and correlation energies for all CPMD calculations, as the revPBE functional has been used for accurate modeling of carbohydrate chemistry [208]. The implemented planewave energy cut-off for energy calculations using the MT pseudopotential, as determined by examining the convergence of the energy, was 70 Ryd. A single k-point (Γ -point) was used for the integration over the Brillouin zone in the reciprocal space. All CPMD calculations were performed at a temperature of 298K and the temperature control was completed using the Nosé-Hoover thermostat [209, 210]. The frequency for the ionic thermostat was selected to be some characteristic frequency of the ionic system to ensure equilibration and the exact value was unnecessary; therefore, it was selected to be 3000 cm^{-1} for all systems, as this is the approximate frequency for the C-H and O-H bond vibrational frequencies. To avoid coupling between the ionic and electronic thermostats the electronic frequency was selected to be 10000 cm^{-1} for all systems. The fictitious electron mass was taken as 300 a.u. for all systems. Short molecular dynamics runs were performed without a thermostat to obtain an approximate value around which the fictitious electronic kinetic energy oscillated. As a result, it was selected to be between 0.004 a.u. and 0.005 a.u. for each system for all subsequent calculations. A timestep of 4 a.u. ($\sim 0.0968\text{ fs}$) was chosen for all calculations. Energies, including the fictitious electronic kinetic energy, were monitored to verify the systems did not deviate from the Born-Oppenheimer surface during the molecular dynamics calculations. Trajectories were visualized using the Visual Molecular Dynamics (VMD) software [211].

To accelerate the sampling of the CPMD calculations as well as access conformations with significant energy barriers, the metadynamics technique [202, 203] was used. A detailed

description of this technique is provided by Laio and Gervasio [212]. In the present work, two torsion angles were selected as the collective variables. Torsion angle 1 (first collective variable) was chosen to rotate the position of the phenyl ring, whereas torsion angle 2 (second collective variable) rotated the hemicellulose moiety. This ensured the system sampled all major conformation changes, i.e., the relative position of the simplified monoglucol to the hemicellulose component was varied. Minor conformational changes such as methyl group orientation and hydrogen bonding direction were accounted for in the dynamics (provided kinetic energy) of the system and adequately sampled without any biasing in the system. The torsion angles selected as collective variables are depicted in Figure B.1.

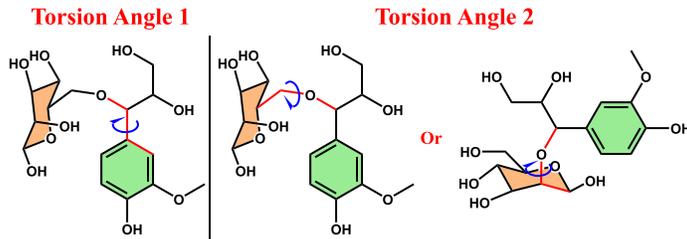


Figure B.1: A schematic of the collective variables (CVs) used in the CPMD-metadynamics calculations. Depending on the connection of the sugar molecule, torsion angle 2 was varied.

The parameters for the metadynamics calculations were chosen following the guidelines as previously outlined [213, 214]. Each fictitious particle, \tilde{s} , is connected to its actual collective variable, $S(x)$, and the extent of coupling between the two is described by a spring with a specified force constant, k_{cv} . This force constant was chosen in such a way that the typical value of the difference between the fictitious particle and collective variable ($\tilde{s} - S(x)$) was smaller than the length on which the free energy varies at the system temperature, T , leading to the condition [212].

$$\langle (\tilde{s} - S(x))^2 \rangle \sim \frac{T}{k_{cv}} \left(\langle \tilde{s}^2 \rangle - \langle \tilde{s} \rangle^2 \right)$$

Where averages were taken at a temperature, T , during a sample metadynamics run without addition of potentials. The fictitious mass, m_{cv} , was selected to ensure the dynamics of the collective variable were slower and thus separable from the nuclear dynamics. It has been shown that the extra terms in the metadynamics extension of the CPMD Lagrangian introduces an additional frequency term proportional to $\sqrt{k_{cv}/m_{cv}}$ [213]. Once the force constant was selected the value of fictitious mass was chosen such that during a sample metadynamics runs of ~ 200 fs without the addition of potentials, the difference between the fictitious particle and collective variables ($\tilde{s} - S(x)$), became minimal [214].

The Gaussian width parameter was chosen as one half the range of fluctuation of the collective variable with the smallest amplitude of oscillation during the sample metadynamics run with no potentials. The height of the potential was kept fixed at ~ 0.63 kcal/mol. The metadynamics timestep at which Gaussian potential added followed similarly to the procedure outlined by Mushrif et al. [214]. Occasionally collective variables struggled to sample the

torsional space and visualizing the trajectory indicated this was typically a result of significant hydrogen bonding occurring between the hemicellulose moiety and simplified monoglucosyl and/or steric interference between the two. In these cases, a scaling factor of 1.2 was implemented for any collective variable that was struggling to sample. Admittedly, the choice of Gaussian potential parameters resulted in a more “hard-driven” simulation; however, since the objective was to screen conformers, rather than identify the global minimum this was deemed an acceptable procedure for the studies objectives.

All metadynamic simulations were performed for a minimum of 1,000,000 CPMD steps and 5000 metadynamic steps. The criteria for rigorously evaluating the convergence of a metadynamic simulation is whether the system begins to rapidly oscillate along the collective variables indicating the free energy surface is filled. However, this was computationally demanding and unnecessary for the present objectives. Therefore, the criteria selected to assess such convergence was the degree each torsional space was sampled. Convergence of the metadynamic simulations were deemed complete if at least one of the torsion angles sampled completely (0-360°) and the other sampled over 90% of the torsional space. The validity of this criteria was reinforced when inspecting the free energy landscape. The free energy surfaces constructed showed all significant minimum wells had dimensions greater than any unsampled region in the simulations, providing confidence no significant conformations were excluded. All structures corresponding to an identifiable minimum were subsequently optimized in all-electron DFT calculations.

The starting set of conformers identified with CPMD+metadynamics were optimized at a M06-2X/6-31G level of theory. All structures within 21 kJ/mol of the lowest energy structure were subsequently optimized at a M06-2X/6-31+G(d) level of theory, this ensured a wide sampling of conformers, and no low energy conformer was excluded from moving forward in the screening process. The top five lowest energy structures were taken for optimization at a M06-2X/6-311+G(d) level of theory, then the lowest energy structure was further optimized at a M06-2X/6-311++G(d,p) level of theory. As the level of theory increased, no dramatic change in the relative trend of conformers was observed.

Due to a reduced computational cost for the reactant compounds, the initial conformational screening was performed directly using DFT calculations at a M06-2X/6-31G level of theory. All structures within 5 kcal/mol of the lowest energy structure identified in literature for glucose [151], mannose [215] and galactose [216] were taken for optimization. On account of the significantly less degrees of freedom associated with arabinose and xylose, conformation sampling of these compounds was done manually where multiple conformers were generated to maximize hydrogen bonding interactions and subsequently optimized. The conformational space of glucuronic acid was sampled by scanning over two torsion angles, depicted in Figure B.2. Torsion angle 1 was rotated at 30° intervals, whereas torsion angle 2 was rotated at 45° intervals, to generate all possible structures within the 360° periodicity of each torsion angle. Note, two scanning calculations were completed where the hydrogen bonding around the 6-member ring was oriented clockwise as well as counterclockwise. In both scanning calculations a crude free energy surface was developed where all observable minimums were selected for further optimization at the M06-2X/6-31G level of theory where no constraints were imposed. Criteria for subsequent optimization of all reactant compounds follows similarly to the LCC moieties.

Conformation sampling of the quinone methide (QM) intermediate was performed manually.

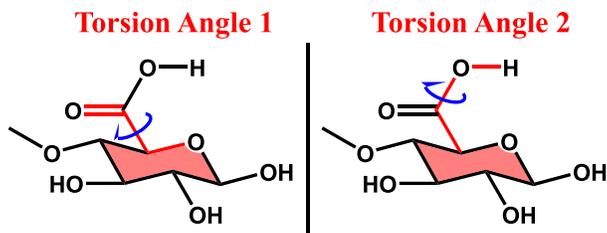


Figure B.2: The schematic of the torsion angles systematically rotated to generate an approximate free energy landscape of glucuronic acid.

B.2 Kinetic Data

Reaction Half-Life Calculations

Under the framework of the transition state theory, one can approximate the reaction rate using the Eyring equation:

$$k_r = \frac{\kappa k_B T}{h} e^{-\frac{\Delta G_a}{RT}}$$

where k_r is the reaction rate constant, κ is the transmission coefficient and assumed to be 1, k_B is the Boltzmann constant, h is Planck's constant, ΔG_a is the activation energy and T is the reaction temperature. Assuming that the reaction is first order, the rate law is given as

$$[A] = [A_o]e^{-k_r t}$$

where $[A]$ is an arbitrary concentration with the subscript, o , denoting the initial state and t is the time. The reaction half-life is then given by

$$t_{1/2} = \frac{\ln(2)}{k_r}$$

Table B.1: The calculated activation energy, rate constant as well as reaction half-life for the nucleophilic addition to the QM intermediate of the most thermodynamically favorable reaction sites on the model compounds considered. Un-bracketed numbers indicate gas phase calculations and bracketed numbers indicate implicit solvent calculations.

Nucleophile (Reaction Site)	Activation Energy (kJ)	Rate Constant (s^{-1})	Half-Life (s)
Mannose (C2)	41.1 [34.5]	3.9E+05 [5.6E+06]	1.8E-06 [1.2E-07]
Mannose (C6)	36.5 [32.5]	2.4E+06 [1.2E+07]	2.8E-07 [5.6E-08]
Glucose (C3)	30.4 [29.5]	3.0E+07 [4.2E+07]	2.3E-08 [1.6E-08]
Glucose (C6)	26.9 [26.1]	1.2E+08 [1.7E+08]	5.8E-09 [4.2E-09]
Galactose (C6)	21.3 [21.7]	1.2E+09 [9.9E+08]	6.0E-10 [7.0E-10]
Arabinose (C5)	16.7 [13.8]	7.2E+09 [2.4E+10]	9.6E-11 [2.9E-11]
Xylose (C3)	27.2 [31.5]	1.1E+08 [1.9E+07]	6.5E-09 [3.7E-08]
Glucuronic Acid (C6)	60.8 [82.3]	1.4E+02 [2.3E-02]	5.1E-03 [3.0E+01]
Water	38.8 [28.8]	1.0E+06 [5.5E+07]	6.9E-07 [1.3E-08]

Appendix C

Supporting Information for Chapter 3

Figure C.1, C.2 and C.3 show the molecular structures of the product compounds resulting from the $C_6 - O_1$, $C_2 - OH$ and $C_3 - OH$ bond breaking pathways, respectively. For further details on the reaction pathways, including the naming used to distinguish each pathway, please see Chapter 3.

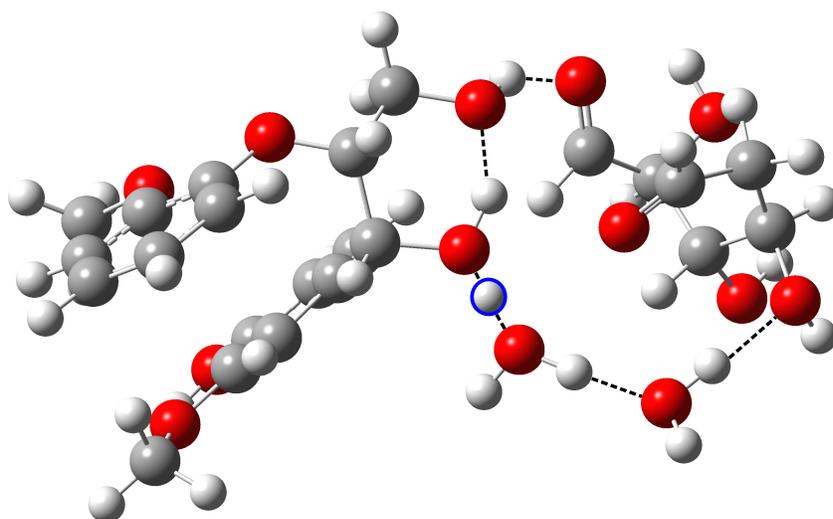


Figure C.1: The product structure resulting from the $C_6 - O_1$ bond breaking pathway. Carbon, oxygen and hydrogen are colored in gray, red and white, respectively. The approximate location of the positive charge is highlighted in a blue circle and the hydrogen bonds are depicted with black dashed lines.

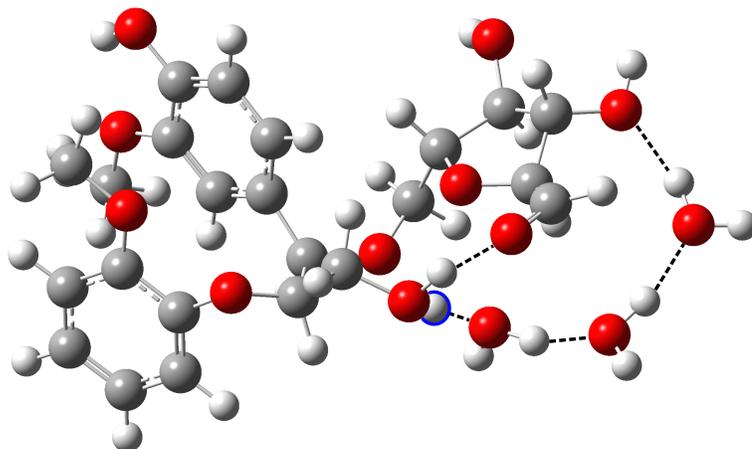


Figure C.2: The product structure resulting from the $C_2 - OH$ bond breaking pathway. Carbon, oxygen and hydrogen are colored in gray, red and white, respectively. The approximate location of the positive charge is highlighted in a blue circle and the hydrogen bonds are depicted with black dashed lines.

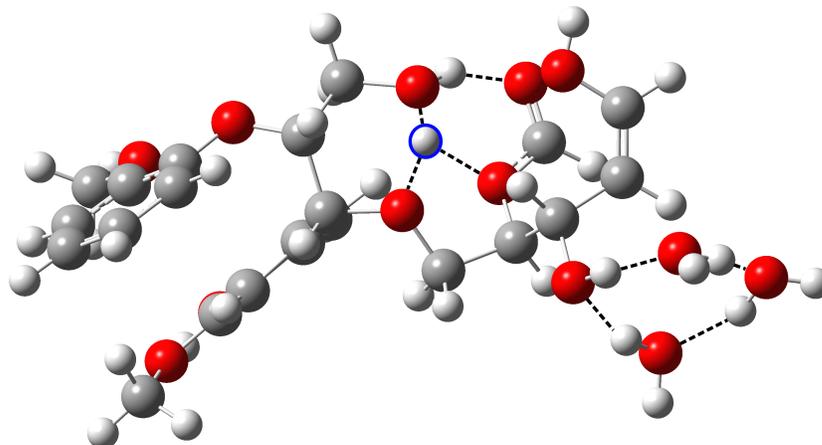


Figure C.3: The product structure resulting from the $C_3 - OH$ bond breaking pathway. Carbon, oxygen and hydrogen are colored in gray, red and white, respectively. The approximate location of the positive charge is highlighted in a blue circle and the hydrogen bonds are depicted with black dashed lines.

Appendix D

Supporting Information for Chapter 4

D.1 Identifying Barrier via Scanning

Figure D.1 shows the relative electronic energy as a function of distance between the carbon 1 location on xylose and the oxygen at the carbon 4 location on lignin, with the energy of the complex separated by 2.97 Angstroms taken as reference. The distance between the carbon 1 location on xylose and the oxygen at the carbon 4 location on lignin will be referred to as the reaction coordinate. As shown in Figure D.1 when the reaction coordinate is at 2.97 Angstroms, the system is at a minimum and this will be referred to as the reactant complex. Incrementally decreasing (scanning) the distance of the reaction coordinate results in a small barrier, ~ 3 kJ/mol at a reaction coordinate distance of ~ 2.6 Angstroms. Further, scanning the distance of the reaction coordinate results in a continuous decrease in energy, towards a product complex. Taking the approximate product structure on the left hand side of the scanned plot, i.e., the structure with a reaction coordinate of ~ 2.1 Angstrom and a relative energy of ~ -9 kJ/mol, and implementing geometry optimization with no constraints reveals a distance of ~ 1.5 Angstroms in the reaction coordinate, suggesting a carbon-oxygen bond forms. In other words, the product complex corresponds to a phenyl glycoside (PG) linkage forming. Although the barrier of ~ 3 kJ/mol is computed in electronic energy, given the minimal magnitude of the barrier, it is assumed the free energy barrier is reasonably approximated by ~ 3 kJ/mol as well.

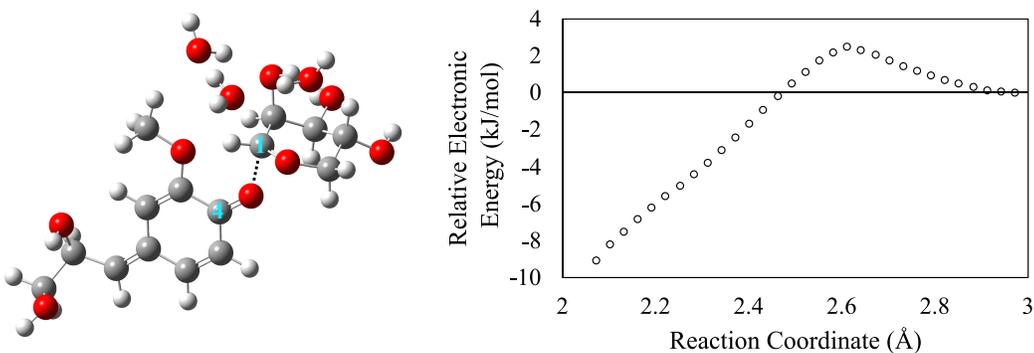


Figure D.1: The relative electronic energy as a function of distance between the carbon 1 location on xylose and the oxygen at the carbon 4 location on lignin (reaction coordinate) with the energy of the complex separated by 2.97 Angstroms taken as reference. Energies reported are at a M06-2x/6-311++G(d,p) level of theory

D.2 Reaction Energetics Temperature Dependence

During biomass deconstruction, elevated temperatures are often employed [18, 19], as a result, evaluating the reaction energetics as a function of temperature is necessary to bring the computational results closer to experimental conditions. Figure D.2 and Figure D.3 shows the activation and reaction free energies, respectively, for cleaving the mannose and xylose PG LCC linkage ($C_1 - O_1$ and $O_1 - C_{4'}$) as well as the competing hemicellulose degradation pathways ($C_2 - OH$ and $C_3 - OH$), as a function of temperature. The circles denote the computationally computed values at 298K. The lines represent the predicted free energies as a function of temperature, and were calculated by assuming enthalpic and entropic contributions were constant at the values computed at 298K. Therefore, the predicted reaction energetics were determined by incorporating the absolute temperature, *i.e.*, $\Delta G = \Delta H - T\Delta S$.

As observed in Figure D.2, the free energy barriers demonstrate an inverse correlation with temperature, *i.e.*, as the temperature is increased the activation energies decrease. Such a result is anticipated for the majority of dissociation reaction. Moreover, the activation barriers of each deconstruction reaction show similar slopes. As observed in Figure D.3, the reaction free energy of every pathway shows an inverse correlation with temperature, indicating as the temperature increases, the reaction energies will become more favorable, characteristic of most dissociation reactions. The $O_1 - C_{4'}$ bond breaking is endergonic and demonstrates a minimal temperature dependence, indicating no feasible temperature will be attainable that will result in the $O_1 - C_{4'}$ bond breaking becoming exergonic. The $C_2 - OH$ bond breaking with mannose is endergonic until $\sim 420K$, where the reaction becomes exergonic. The $C_2 - OH$ bond breaking with xylose as well as $C_3 - OH$ and $C_1 - O_1$ bond breaking for mannose and xylose are exergonic for the entire temperature range explored. As described in the main paper, cleaving the PG linkage, $C_1 - O_1$ bond breaking, is exergonic and possesses minimal activation barriers, < 95 kJ/mol, indicating such a pathway will be the

first deconstruction pathway accessible in acid catalyzed deconstruction of PG linkages. Such a conclusion remains consistent even when considering the reaction energetic's temperature dependence.

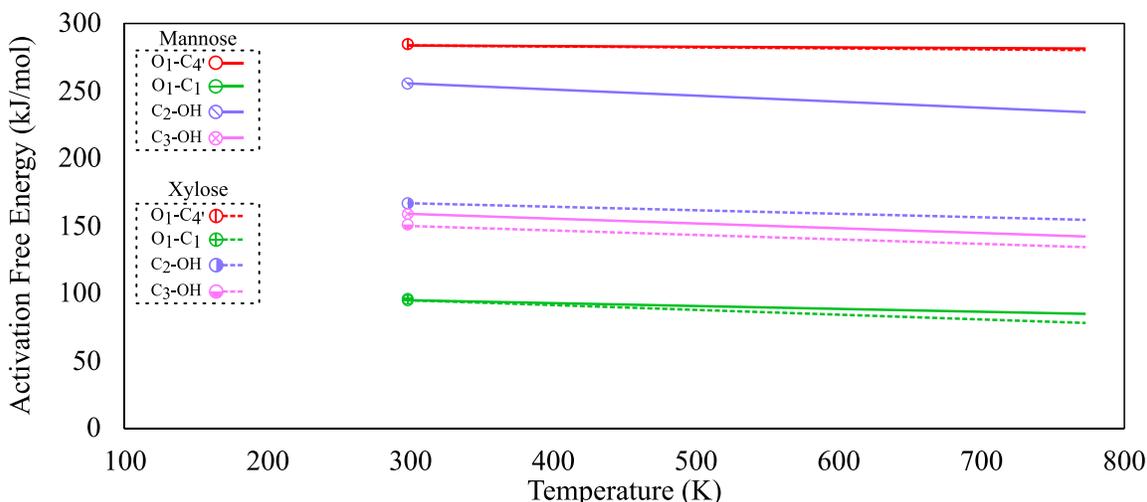


Figure D.2: The free energy barriers for the deconstruction pathways of a phenyl glycoside (PG) LCC linkage as a function of temperature. The circles denote the computationally computed activation barriers at 298K, while the lines are calculated assuming the enthalpic and entropic contributions are constant with the values computed at 298K.

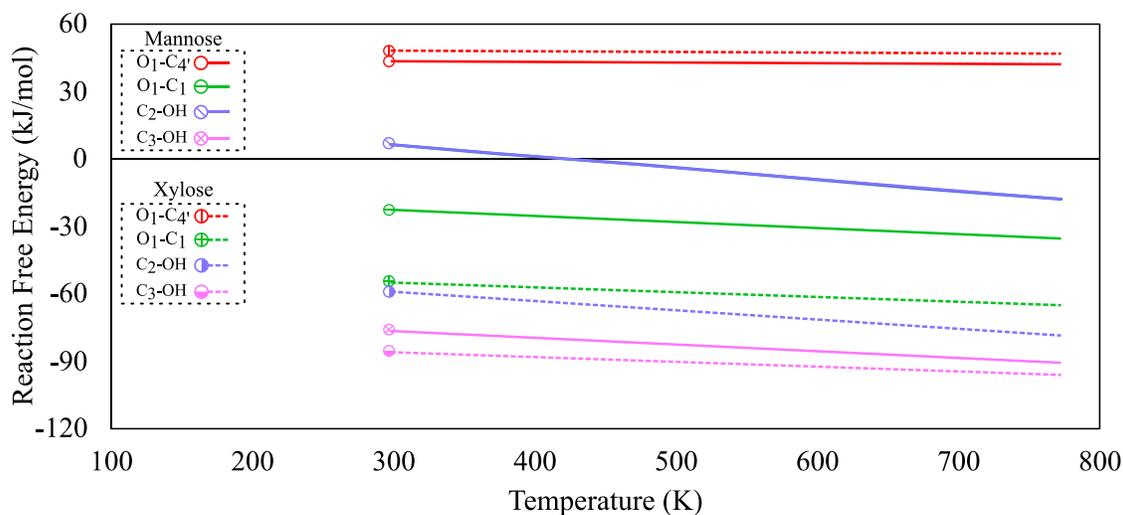


Figure D.3: The reaction free energies for the deconstruction pathways of a PG LCC linkage as a function of temperature. The circles denote the computationally computed reaction energies at 298K, while the lines are calculated assuming the enthalpic and entropic contributions are constant with the values computed at 298K.