Extractives of Douglas-fir and Their Implications for Biofuel Saccharification

Karl R. Oleson

A dissertation

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Daniel T. Schwartz, Chair

Jim Pfaendtner

John C. Berg

Program Authorized to Offer Degree:

Chemical Engineering

University of Washington

Abstract

Extractives of Douglas-fir and Their Impacts for Biofuel Saccharification

Karl R. Oleson

Chair of the Supervisory Committee:

Professor Daniel T. Schwartz

Department of Chemical Engineering

Forestry residues are being investigated as a potential biofuel feedstock since they are an abundant biomass that does not require major land-use change or compete with food resources. This biomass is the waste created from logging and forest thinning operations and consists of branches, treetops, and other unmerchantable timber. Douglas-fir residues have received special attention since Douglas-fir is the most abundant tree in the timberlands of the Western United States. Biofuel from this material has yet to hit full-scale production, but production processes are already being investigated at the pilot plant scale. Creating a biofuel industry from forestry residues is a complex challenge, and there is a need for research into secondary products and process optimization. The best way to go about finding improvements is to holistically connect the challenges observed in processing millions of tons of biomass to the molecular interactions at hand. With this in mind, this research begins by modeling a biofuel process in ASPEN to understand what molecules are present in processing. We then experimentally identify process inhibitors and explore the molecular origins of this inhibition. The specific focus of this research

is on the “extractives” class of molecules in forestry residues. Extractives make up 5% to 25% of the dry weight for different tissues of Douglas-fir, but are overlooked in most biofuel studies. These molecules have been well studied in pulp and paper industry circles and this past knowledge can be used for inspiration for biofuel research. Extractives are both unconsidered secondary products and biofuel process inhibitors. This work first collates the known literature on Douglas-fir extractives into a representative chemical model of Douglas-fir forestry residues. This model is then used to predict which extractives end up in the various process streams of a simulation model of the Northwest Advanced Renewables Alliance (NARA)’s biofuel process. NARA’s biofuel process relies upon a process called saccharification, and the information derived from process modelling is used to begin studying this step. Saccharification is the key step in this process and involves enzymes breaking down biomass into soluble sugars for further conversion into fuels and chemicals. This work shows these enzymes are inhibited by the “tannins” class of extractives in Douglas-fir bark. Initial glucose production from crystalline cellulose is shown to be inhibited ~40%, and specific types of enzyme are individually inhibited ~15-20%. Molecular docking and molecular dynamics simulations were used to investigate how a characterized tannin in Douglas-fir binds onto the most abundantly produced cellulase from *Trichoderma reesei*, Cel7A. We find that the tannin can bind onto the cellulase and obstruct the catalytic tunnel of this enzyme in a variety of different ways with the tannin preferring to bind to sites that contain aromatic or charged amino acid residues. There is still much to learn and explore in this area. Extraction and collection of the tannins as secondary products may be the most appealing option to remove inhibition but may not be economical. Instead, bioengineered tannin resistant enzymes could be appealing, and the information presented here could provide the foundation for that work.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank the following people and organizations for their scientific, monetary, and/or philosophic contributions:

- Dan Schwartz for his guidance, mentorship, and help for which I am forever grateful
- Current and former Schwartz group members: Ikechukwu Nwaneshidu, Trevor Braun, Matthew Murbach, Yanbo Qi, Victor Hu, and Erica Eggleton
- The Pfaendtner research group for their help with molecular simulations. A special thanks to collaborator Kayla Sprenger for her patience and belief in this work as well as thanks to Vance Jaeger for his guidance early on.
- The Pozzo research group for help with UV-Vis work and making the basement of Benson a brighter place to work.
- My committee: Professors Jim Pfaendtner, John C. Berg, and Sharon L. Doty
- Mentorship and teaching from other UW ChemE Faculty not already mentioned, especially Hugh Hillhouse.
- My family for their unwavering support of my PhD ventures and remarkable ability to appear fascinated about my research.
- Agriculture and Food Research Initiative Competitive Grant (No. 2011-68005-30416), USDA National Institute of Food and Agriculture (NIFA) through the Northwest Advanced Renewables Alliance (NARA) for funding part of this work.
- The user facility staff for guiding me in the ways of being considerate to my research neighbors. Particularly Andy Kim and Kameron Harmon

Dedication

To God, my parents, and brother. You guys have helped out every step of the way.

Chapter 1: Forestry Residue Extractives in Biofuel Processing

1.1 Forestry residues are an appealing biomass for biofuels

Forestry residues have been receiving attention as an attractive potential biomass for biofuels. This biomass is the waste created from logging and forest thinning operations and consists of branches, treetops, and other unmerchantable timber. Logging industries are primarily interested in the trunks of trees and, as a rule of thumb, about 40% of a tree (excluding roots) is usually left at the site during timber harvests [1]. These residues are typically collected into "slash" piles at forestry sites and disposed of, commonly by burning. An example of what this looks like is shown in Figure 1-1.

Figure 1-1: Forestry residue (slash) burning after a logging operation along Highway 101 near Quilcene, WA in 2017.

Waste biomass sources such as this are attractive biofuel candidates since they otherwise have limited economic potential or have costs associated with their disposal. This biomass can also avoid issues that other biofuel feedstocks have such as land use change and competition with our food supply[2, 3]. Biofuel production from this material was also not shown to have major impacts on soil quality over short time spans as long as the soil was of good quality previously

(~15 years) and can provide fuel at ~70% lower greenhouse gas emission than petroleum fuels[4]. In 2017, there were approximately 92 million tons a year of this material available for sustainable harvest in the USA (at < $60/ton, Total – Without Federal Land)[5]. Based on current biofuel production yields, there is an opportunity to meet ~25% of U.S. aviation's fuel needs[6, 7] with this material.

1.2 Production of jet fuel from forestry residues

The Northwest Advanced Renewables Alliance (NARA) was an organization created with private industry and educational institutions to investigate jet fuel (biojet) production from forestry residues. Creating jet fuel from non-food biomass such as *Camelina sativa* had been demonstrated by the Navy in 2011[8] and projected prices for similar renewable jet fuels range from $3.80 to $11.00 per gallon[9] ($3.80 from *Jatropha curcas*). NARA sought to make similar progress with forestry residues and focused on designing a process based on modified pulping techniques, adapted biological processes, and oil industry inspired chemical processing. Similar schemes are moving beyond initial lab bench scale studies and are being investigated on pilot-plant scales[10-12] (~50kg biomass in a reactor). NARA's process was recently used to create 1,080 gallons of biojet that was used to power an Alaskan Airlines flight flown from Seattle to Washington D.C[13]. This proof of concept flight is the culmination of numerous advancements in forestry residue chemical processing, logistics, and education that occurred through NARA.

While this flight was a big milestone for forestry residue biofuels, this production process is not yet economical. NARA's economic calculations project the jet fuel to optimistically cost $7.27/gallon[14] (considers a greenfield project, 10% internal rate of return, 30 year project). For comparison, the lowest jet fuel prices within 50 miles of Seattle, WA is 3.55/gal (at Kenmore

Air Harbor, Inc. Seaplane Base, December, 2017)[15]. There is a lot of work that needs to be done if this process is to be made economical in the near future. It is critical to explore ways to increase revenue and decrease process costs. Therefore, exploration of secondary products and ways to more efficiently carry out biofuel production steps should be the focus of future research.

An understanding of NARA's production scheme is necessary in order to find possible secondary products and improvements. NARA chose a fermentation based biofuel route that used pretreatment, saccharification, and fermentation steps to convert the biomass to fuel. Pretreatment of biomass typically involves both physical and thermochemical processes to increase cellulase accessibility to the cellulose during the saccharification step[16]. Different possible pretreatment techniques include dilute acid[17, 18], mild bisulfite[19], organosolv[20, 21], SPORL[18, 22-25], physical pretreatments[26], dissolution of the biomass in ionic liquids[27], steam explosion[21] and many others[28]. Pretreatments typically work better on some feedstocks than others[18], and a version of the mild bisulfite pretreatment became the preferred method for NARA's softwood rich biomass (most of the biomass came from Western United States timberlands)[29]. This pretreatment technique resembles the sulfite process used in the pulp and paper industry[30]. Pretreatment is followed by enzymatic saccharification which was carried out with biomass degrading enzyme blends commonly referred to as "cellulase cocktails" from Novozyme. In this step, cellulose and hemicellulose (the combination of these two are commonly referred to as "holocellulose") are broken down to their constituents monomeric carbohydrates through enzymatic catalysis. Different cellulase cocktails were used as new ones were successively developed and made available to researchers including the Celluclast[19, 22-25], CTec2[18, 19, 31], and CTec3[10, 11, 32] cocktails. The CTec series has been optimized around corn

stover[33] but has also been seen to speed up saccharification for a wide range of biomasses including forestry residues[19]. The early cellulase mixtures used fungal *Trichoderma reesei* enzymes as the workhorse catalysts, but there is less published information on the enzymes used in the newer cocktails. The saccharified sugar stream was then sent to the private company GEVO who carried out fermentation to create isobutanol[34]. GEVO also catalytically upgraded the isobutanol to create the actual biojet, iso-paraffinic kerosene. NARA's process as outlined here allows researchers to take advantage of transferable knowledge from the pulp and paper industry as well as existing fermentation infrastructure. Other biofuel production routes are of course available some of the most promising work coming from the non-fermentation routes such as pyrolysis[35-37].

Similar biomass degradation schemes to NARA's sulfite based pretreatments have been researched for over 70 years, and new innovations may be hard to find in this step. Therefore the best opportunity for new biofuel production improvements may be found in looking at the steps that have been studied less historically, saccharification and fermentation of woody biomass. With regard to new revenue streams, options available for secondary products become limited after NARA's pretreatments. Current secondary products include sulfonated lignins (~$200/dry ton)[14] and activated carbon (~$1,500/dry ton)[14]. The goal of pretreatment is specifically to create sulfonated lignins in order to free up the cellulose. Also, remaining residual solids after processing have little value as anything except as fuel or a carbon source. Since these post-pretreatment secondary products are not enough to make the process profitable, it is worth investigating what possible secondary products could be harvested before pretreatment.

1.3 Extractives in biofuel productions

There is a plethora of phytochemicals beyond the commonly studied cell wall components, holocelluloses and lignin. Phytohormones, chemical defense molecules, and metabolites are all part of the biomass but are not typically analyzed in biofuel studies. These molecules can be extracted from woody tissues with minimal effect on the cell wall and are appropriately named "extractives". These molecules can have value as secondary products and can also harm process rates as inhibitors of the biomass conversion scheme. Extractives have been studied in the literature for over 100 years, and information on them can be found spread out in a wide range of fields including pulping[30], pulp mill wastes[38], leather tanning[39], and other fields. Extractives are species specific, and considering them when optimizing biofuel production could result is new ways to increase revenues and reduce the cost of biojet fuel processing.

Douglas-fir (*Pseudotsuga menziesii*) has been a biomass feedstock of particular interest due to this tree species' prevalence in the timberlands of the Western United States[29]. Douglas-fir contains water, cellulose, hemicellulose, lignin, extractives, ash and protein. Cellulose, hemicellulose, and lignin are typically the focus of biofuel studies since they commonly make up ~90% of the oven dry (o.d.) mass of forestry residues (also known as slash). Extractives are the next largest chemical class and make up 5-25% of the o.d. mass of different tissues in Douglas-fir. These extractives contain phytochemicals classes such as proanthocyanidins and phlobaphenes, waxes, flavonoids, terpenoids, phytosterols, lignans (as opposed to lignin), and many more[40]. Smaller carbohydrates can also be extracted in small quantities along with non-carbohydrate extractives, but they are commonly ascribed to cellulose or hemicellulose during biomass characterization. These molecules have diverse chemical properties and functionalities. Volatile extractive molecules such as the terpenes are largely responsible for the "Christmas-tree" smell in forests and are chemical signalers to bugs and ruminants. Non-volatile, insoluble,

waxes in the bark help protect against desiccation. Tannins protect the bark from being eaten by animals by binding and wrapping up salivary proteins (making them astringent and unpalatable)[41]. Extractives protect a tree against degradation in nature, so it is perhaps unsurprising that they can also impact biological routes to biofuels.

A great deal of information on how these molecules affect NARA's biofuel process can be gained from work done in other fields. Pretreatment techniques that are similar to pulping techniques can draw from the pulp and paper literature to learn about extractives reactions in pretreatment. Examples include soluble dihydroquercetin reacting to insoluble quercetin [30] and condensed tannins, either breaking down to smaller sulfonated components or polymerized to larger polymers depending on conditions[42-44]. In addition, various molecules such as tannins[45], flavonoids[46], and monoterpenes[47, 48] have been studied as antifungal agents and could be a concern for biological steps such as saccharification and fermentation. Saccharification can be expected to suffer from Douglas-fir tannins in particular since tannins in other plant species have been reported to bind and inhibit a wide variety of enzymes such as salivary proteins[49], α-amylases[50], and even cellulases[51]. The biggest concerns for wastewater treatment of biofuel processes will need to pay attention to resin acids and phytosterols since they are known to be toxic or endocrine disruptors[52-55]. Chapter 2 contains a much more in depth review of the 100+ years of literature on Douglas-fir extractives and uses an ASPEN simulation to models how representative extractive molecules propagate through the NARA process.

Modern pretreatment techniques have yet to specifically study what happens to extractives in their processes and if there are negative consequences associated with them. However, it is likely that pretreatment is well optimized with respect to extractives because many pretreatments have been tried and ones that are sensitive to extractives would have been weeded

out. There are also many questions pertaining to extractives and the saccharification and fermentation steps that have yet to be answered. It has remained unclear how much inhibition can occur at biofuel relevant extractive conditions, which extractives can cause inhibition, and how inhibition can occur. This means research into the saccharification and fermentation steps probably have a greater chance of finding improvements. We focus on enzymatic saccharification as a particularly intriguing process step for extractive-related understanding.

1.4 Saccharification and Phytochemical Inhibitors

As mentioned previously, saccharification is the step in which biomass degrading enzymes are used to break cellulose and hemicellulose into sugars. Pretreated Douglas-fir cellulose is broken down to glucose with cellulases and the hemicellulose is broken down to monomeric sugars with hemicellulases. Cellulases used in industrial saccharification typically come from fungi and contain mixtures of exoglucanase, endoglucanase, β-glucosidase, and other enzymes to break down the cellulose[56]. Hemicellulases commonly include enzymes such as β-xylosidase, β-mannanase, acetylxylan esterase, and many more[57-59]. Other enzymes are commonly included in biomass degrading enzyme mixtures such as ligninases which act upon lignin[60] and lytic polysaccharide monooxygenases (LPMOs) which oxidatively induce chain breaks into the cellulose[61].

Identifying how extractives inhibit saccharification is complicated by both the complexity of the enzyme mixtures used and heterogeneous nature of the substrate. Cellulose is the main concern for NARA's saccharification step since pretreatment steps solubilize lignin and partially breaks down the hemicellulose beforehand[10, 19]. Cellulose is made up of long chains of β-D-glucopyranose (glucose) where the repeating unit is two rotated glucose molecules[62]. It is commonly classified into "crystalline" or "amorphous" regions. Crystalline cellulose

corresponds to areas with regular repeating hydrogen bonding between cellulose strands to form sheets while amorphous regions are irregular, non-uniform cellulose areas. Crystalline cellulose in nature comes as a mixture of two crystal structures, I_α and I_β, with algae and bacterial cellulose containing a larger amount of I_α cellulose[63] while terrestrial plants having a larger proportion of I_β[62-64]. Exoglucanases are a family of enzymes considered responsible for breaking down the crystalline cellulose regions (primarily) into cellobiose[65] and cellulases from *Trichoderma reesei* act on the hydrophobic sides of either the Iα[66] or Iβ[67, 68] crystalline regions. The endoglucanase family of enzymes (largely) breaks down amorphous regions into cellobiose or glucose, while the β-glucosidases break down cellobiose into glucose. Cellulase Inhibition studies commonly use model substrates of crystalline cellulose, amorphous cellulose, and cellobiose in order to try and identify which specific cellulases are inhibited.

Saccharification is known to be inhibited by many compounds found in biomass or created during pretreatment, such as polyphenols[51, 69], various saccharides[70, 71], and certain ions[72]. Little is known about how the complex mix of extractive molecules from Douglas-fir bark affects saccharification rates, even though related molecules to the tannins and phenols in Douglas-fir are reported to be strong inhibitors[51, 69, 73-76]. Chapter 3 contains results of saccharification experiments in which sugar yields in the presence or absence of bark extractives.

Neutralization or removal of the tannins could lead to improved saccharification yields. However, common ways of removing such inhibition, such as washing and pressing steps or adding chemical additives to neutralize tannins, all add to the cost of biofuel production[77, 78]. Removal of inhibitors such as tannin and collecting them as a side product would be an ideal solution to this problem. People have looked into what the market for tannins could be in the past and found that they have similar functionality to the quebracho tannins that dominate the

tannin market[79]. Unfortunately this market has been in decline for some time as leather products are displaced by plastics. This has made an industry based solely on Douglas-fir tannins not viable.

If the tannins are to remain in solution for saccharification, it is necessary to understand their effects on enzymatic processes and devise strategies to combat them. Molecular and atomic scale information of how the tannins bind to cellulase could substantially help in determining the best way to do this. To this end, Chapter 4 contains the results from a study using molecular dynamics simulation to analyze how a Douglas-fir tannin binds onto the most abundant cellulase enzyme.

Chapter 2: Extractives in Douglas-fir Forestry Residue and Considerations for Biofuel Production

2.1 Chapter Summary

Forestry residues are a plentiful, low environmental impact feedstock for biofuels and bioproducts. Douglas-fir is the most prevalent tree species in the timberlands of western North America, with approximately 5 million tons of sustainably harvestable forestry residues available each year. These forestry residues are an important potential biomass feedstock containing holocellulose, lignin, protein, ash, and phytochemicals commonly identified as "extractives". The phytochemical extractive category make up 5% to 25% of the dry weight for different tissues of Douglas-fir, but are rarely represented with molecular detail in feedstock models of residues for biofuel or other bioproducts. These extractives contain both primary and secondary metabolites and represent potential revenue sources as side products from processing, but also includes species that are astringent, toxic, endocrine disruptors and/or reactive in similar chemical processes. Within the "extractives" category are phytochemicals such as proanthocyanidins, phlobaphenes, waxes, flavonoids, terpenoids, phytosterols, lignans and many more. This chapter first identifies phytochemical molecules found in different Douglas-fir tissues, then quantities these by category and individual molecular species, to the extent allowed by the literature. We combine the literature into a quantitative, molecularly detailed, mass conserving model for a particular Douglas-fir forestry residue ("slash"). This model is used in a sulfite/bisulfite biofuel process simulation for understanding the molecular partitioning of extractives in different process streams. Model results are used to explore some implications for extractive species in the production of sugars and waste products from Douglas-fir forestry residue feedstock.

2.2 Intro

Pseudotsuga menziesii (Mirb). Franco (common name, Douglas-fir) has been investigated as a feedstock for many natural products industries due to its abundance in western North American timberland. It has received attention recently for its potential to be a large component of a forestry residue-based biofuel industry [29]. Douglas-fir biomass is a large portion of the 75 million tons a year of waste logging residue and forest thinnings that are available for sustainable harvesting in the U.S.A. [5]. Using waste biomass generated in forestry activities is attractive for biofuels because it has fewer potential land-use change and food competition issues than many other biofuel feedstocks [2, 3]. Biofuel production from forest residues has yet to hit full-scale operations, but pilot scale studies are being conducted [80]. However, the biofuel community has yet to incorporate the molecular properties of the diverse "extractive" components of biomass— phytochemicals produced as part of the biological signaling, defense, and metabolism—into their most sophisticated feedstock models and process designs [81]. An improved phytochemical description of Douglas-fir could accelerate biofuel development. The purpose of this review is to provide a quantitative molecular description of Douglas-fir suitable for advancing the development of this field.

Research in the pulp and paper industry has shown that the fraction of the forestry biomass known as "extractives" account for many of the toxic emissions during processing, can participate in undesired side reactions, or inhibit chemical rates during processing [30, 82, 83]. At the same time, some extractives may be valuable side products for a biofuel facility [79]. Extractives have been reported to be in the range of 5-25% by oven dry (o.d.) mass for Douglas-fir, depending on tissue type [84, 85]. In the prominent tissues of Douglas-fir logging slash—heartwood, sapwood, and bark—the most abundant noncarbohydrate extractives can be categorized as

proanthocyanidins, phlobaphenes, flavonoids, waxes, terpenoids, phytosterols, and lignans. The identity, but often not the quantity, of many specific extractive molecules in Douglas-fir have been cataloged (http://tree-chemicals.csl.gov.uk) [86]. Because extractives are particularly prevalent in bark tissue of Douglas-fir, quantitative and comprehensive reviews of extractive phytochemicals were undertaken decades ago for this tissue [87]. While bark is the most extractive-rich component of logging slash, several percent of the mass in sapwood and heartwood can also be comprised of extractive molecules. We estimate that roughly a third of the non-carbohydrate extractive molecules in Douglas-fir forestry residues come from tissues types that have not been subjected to comprehensive, quantitative review.

In order for the biofuel field to identify possible side products, inhibitors, and toxic compounds, an up to date and comprehensive survey of the literature is needed that includes both the known extractives and their expected concentration in biomass. In this work, we review the characterization, possible economic value, and quantification of Douglas-fir extractives, as well as discuss implications of our quantitative feedstock model for biofuel processes based on sulfite/bisulfite pretreatment.

2.3 Extractives species in Douglas-fir and possible uses

The extractives considered here are non-carbohydrates isolated by steam distillation, ether extraction, alcohol extraction, water extraction, or other techniques that do not significantly deconstruct the cell wall. More aggressive extraction techniques can extract soluble carbohydrates such as pectin, but these are likely degraded structural components and are not further considered here. The identified non-carbohydrate extractives of *Pseudotsuga menziesii* (Mirb.) Franco have been assigned here to categories of proanthocyanidins and phlobaphenes, waxes, flavonoids, terpenoids, phytosterols, lignans, and other molecules.

There are two varieties of Douglas-fir, *Pseudotsuga menziesii* (Mirb.) Franco var. *menziesii* and var. *glauca.* Analysis of the foliage, leaves, and branches of the two varieties have shown that the concentrations of individual molecular species differ substantially, but the two generally have very similar molecular species present [88]. Considering this and how studies often do not identify the variety they analyzed, molecules identified in either variety are presented to give a more comprehensive picture of what may be found. The chemical structures of the identified extractives are presented in tables 2- (1-12).

2.3.1 Proanthocyanidins and Phlobaphenes

This category contains various length proanthocyanidin polymers (condensed tannins), and related co-occurring phlobaphenes. The condensed tannins identified are mostly heterogeneous procyanidins. The extender units (main body) in the bark are almost exclusively epicatechin units while the terminal units are mixed epicatechin and catechin [89]. The chain lengths vary and the inner bark has been reported to have condensed tannins with an average of 7 flavonoid units, with tetramers and pentamers of epicatechin and catechin specifically identified [90]. Bonds between individual flavonoids units are most commonly from the C-4 carbon of one to the C-8 of another (shown in the tetramers and pentamers of Table 2-1), but C-4 to C-6 linkages also occur [89]. Another proanthocyanidin type, prodelphinidins, have also been identified in the needles but fewer studies have been conducted on them [91]. The phlobaphenes co-occur with the condensed tannins in tissues such as the outer bark and heartwood. They are water-insoluble phenolic substance comprised of polymeric procyanidins, dihydroquercetin, carbohydrate (glucosyl) and methoxyl moieties [92]. In addition, a very similar phlobaphene substance can be obtained by treating condensed tannins (sometimes called phlobatannins) with acid.

Tannins of Douglas-fir have been extensively studied for use in industries such as leather tanning, drilling muds, ore flotation, ceramics, and cement. However, competition with South American and European tannin production has limited its use in these areas. It is commonly believed that harvesting tannins in a multi-product process could be possible, but an industry based on extracting the tannins alone is not economical [79]. The phlobaphenes have been of less economic interest than the tannins and commonly are investigated indirectly through studies that target tannins, flavonoids, and lignans. Table 2-1 contains the proanthocyanidins specifically identified in literature.

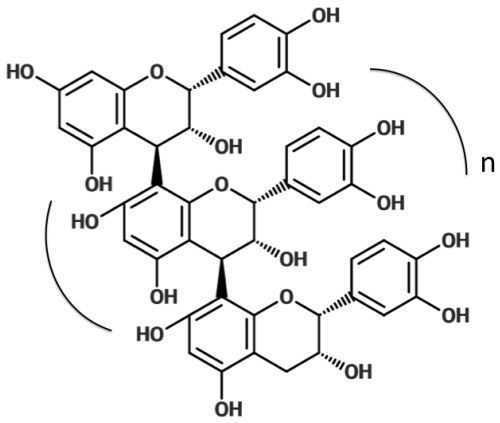
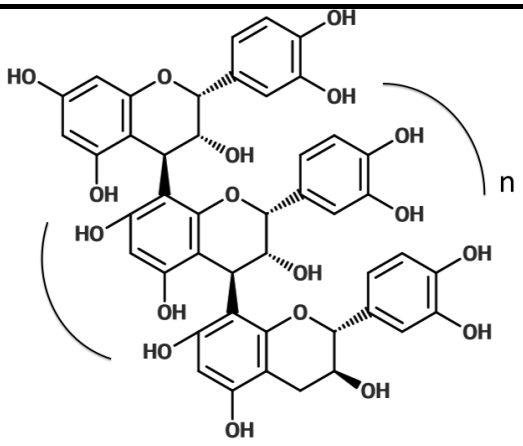
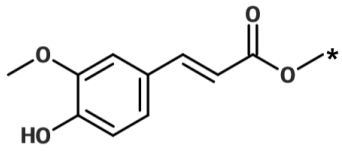
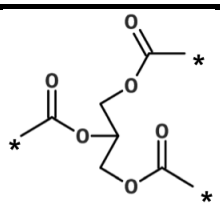
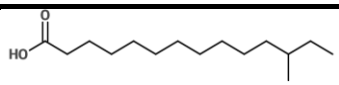
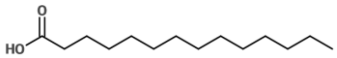
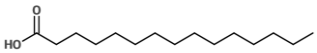
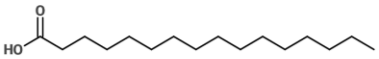
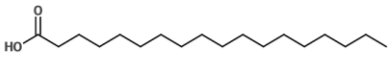
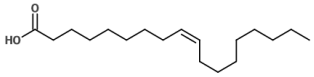
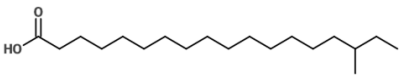
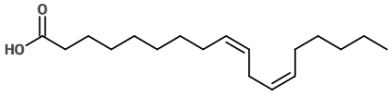
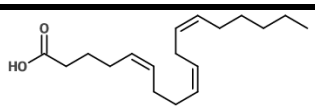
	
n=2; procyanidin tetramer with terminal epicatechin units	n=2; procyanidin tetramer with terminal catechin unit
n=3; procyanidin pentamer with terminal epicatechin units	n=3; procyanidin pentamer with terminal catechin unit
n=7; average inner bark procyanidin	n=7; average inner bark procyanidin

Table 2-1: Identified and Characterized Proanthocyanidins [89, 90]

2.3.2 Waxes

Triglycerides and ferulic acid esters are the most abundant components in the Douglas-fir waxes category. Saponification of the triglycerides releases monocarboxylic acids ranging in size from palmitic acid (C 19:0) to montanic acid (C 28:0), dicarboxylic acids ranging in length from hexadecanedioic acid to tetracosanedioic acid, and alcohols ranging from 1-hexadecanol (C 16:OH) to 1-tetracosanol (C 24:OH). Corresponding alcohols that form upon hydrolysis of the ferulic acid esters are primarily behenyl alcohol and lignoceryl alcohol [93]. Further information on thin-layer chromatography and gas-liquid chromatography on the waxes of the bark was reviewed by Laver et al. [87]. Free fatty acids have also been reported, but typically the concentration is an order of magnitude less than their ester counterparts [94, 95].

Research into the waxes from Douglas-fir in the past has been spurred by interest in using them as a commercial source of natural wax. Possible applications for the wax that have been investigated in the past include polishes, ski wax, ointments, lubricants, soaps, art and sculpture work, preservatives, and many others [79]. Wax is one of the most commonly investigated extractable products from Douglas-fir, along with tannins and dihydroquercetin.

		
Ferulic acid ester; * chain that yields behenyl alcohol or lignoceryl alcohol upon hydrolysis	Triglyceride; * Side chains of various lengths. Hydrolysis yields glycerol, assorted monocarboxylic acids, dicarboxylic acids, and alcohols upon hydrolysis	
		
12-Methyltetradecanoic acid	Tetradecanoic acid	Pentadecanoic acid
		
Hexadecanoic acid	Octadecanoic acid	Oleic acid
		
16-methyloctadecanoic acid	Linoleic acid	Isomeric linoleic acid

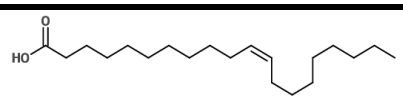
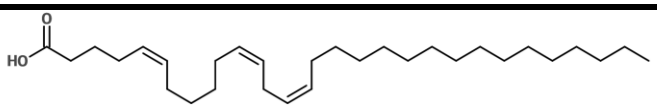
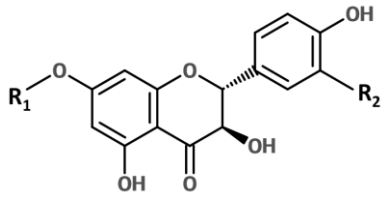
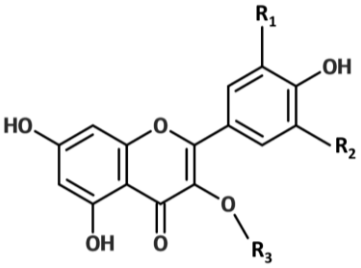
	
11-Eicosenoic Acid	5,11,14-Eicosatrienoic acid

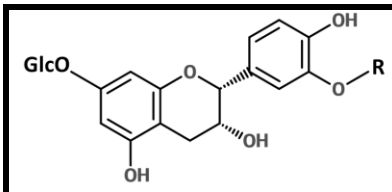
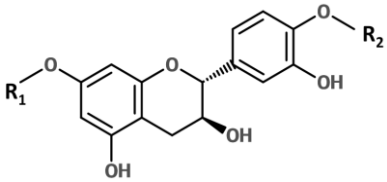
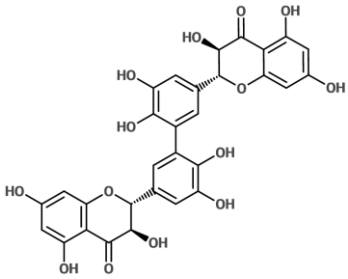
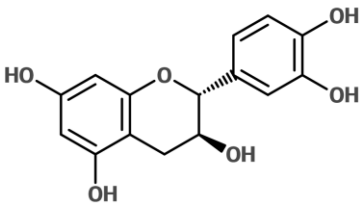
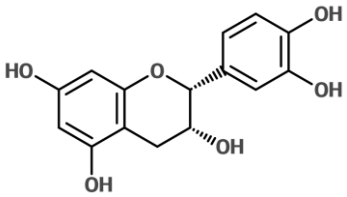
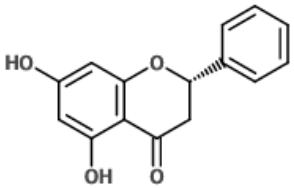
Table 2-2: Identified Waxes [95-97]

2.3.3 Flavonoids

Douglas-fir is known to have a variety of flavonoids and flavono-compounds, including flavanols, flavonol glycosides, O-acylated flavonol glycosides, a biflavonoid, and a flavonolignan. Studies of Douglas-fir also led to the discovery of several of these flavonoid chemistries. [5', 5']-Bisdihydroquercetin was discovered in Douglas-fir and was the first true biflavonoid linked by their B-rings found [98]. The first flavonolignan, pseudotsuganol, was also found in Douglas-fir [99]. In addition, four new novel O-acylated flavonol glycosides (daglesiosides) were recently identified in the needles of Douglas-fir from Poland [100]. Table 2-3 shows identified flavonoid moieties.

The flavonoids in Douglas-fir were first studied to determine pulping inhibitors in Douglas-fir. Dihydroquercetin (Taxifolin) is a relatively abundant flavonoid and was specifically identified as a mild pulping inhibitor, but was a small consideration compared to the difficulties associated with the lignin of Douglas-fir [30]. Dihydroquercetin and quercetin were subsequently studied for economic uses. These chemicals are used in the food industry; however they have a limited market. In 1970, it was speculated that the world market for quercetin probably did not exceed 40,000 pounds [79].

	
R_1=H, R_2=O-β-D-glucopyranosyl; Dihydroquercetin 3'-*O*- glucoside	R_1=H, R_2=OH, R_3=Glc; Quercetin 3'–O-glucoside
R_1= β-D-glucopyranosyl, R_2=OH; Dihydroquercetin 7-*O*-β-D-glucopyranoside	R_1=H, R_2=OH, R_3=Gal; Quercetin 3'-O-galactoside
	R_1=H, R_2=OH, R_3=H; Quercetin
	R_1=H, R_2=H, R_3=H; Kaempferol
R_1= β-D-glucopyranosyl, R_2=H; Dihydrokaempferol 7-*O*-β-D-glucopyranoside	R_1=H, R_2=H, R_3=Glc; Astragalin
	R_1=O-Me, R_2=O-Me, R_3=H; Syringetin
	R_1=O-Me, R_2=H, R_3=H; Isorhamnetin
R_1=H, R_2=OH; Dihydroquercetin	R_1=O-Me, R_2=OH, R_3=H; Larycitrin
R_1=H, R_2=H; Dihydrokaempferol	R_1=OH, R_2=OH, R_3=H; Myricetin

		
R=H; Epicatechin 7-*O*-β-D-Glucopyranoside	R_1=Glc, R_2=H; Catechin 7-*O*-β-D-glucopyranoside	
R=Me; 3'O-Methylepicatechin 7-*O*-β-D-glucopyranoside	R_1=H, R_2=Glc; Catechin 4'-*O*-β-D-Glucopyranoside	[5,5] - Bisdihydroquercetin
		
Catechin	Epicatechin	Pinocembrin

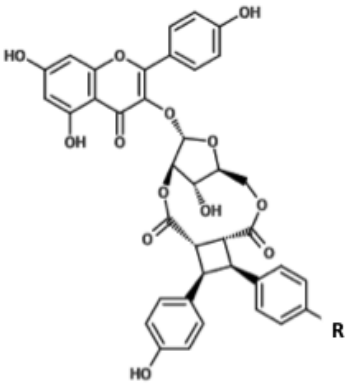
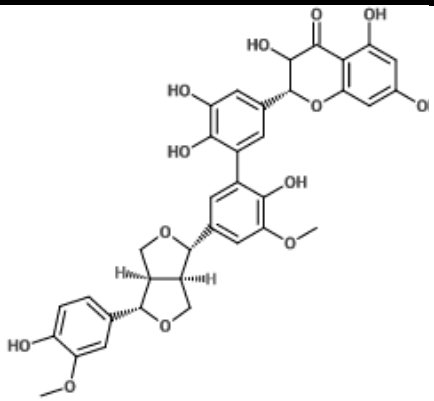
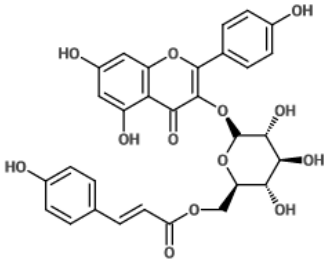
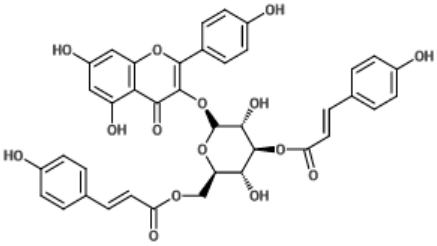
		
R=OH; Daglesioside I		
R=H; Daglesioside II	Daglesioside III	Daglesioside IV
		
Pseudotsuganol	Tiliroside	Ditiliroside

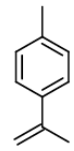
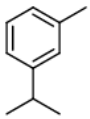
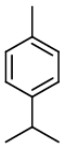
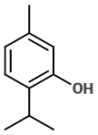
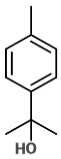
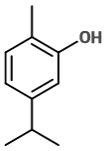
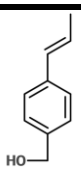
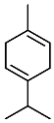
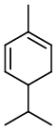
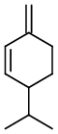
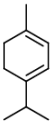
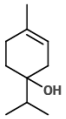
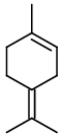
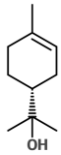
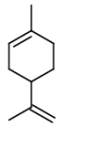
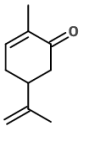
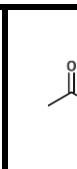
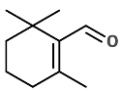
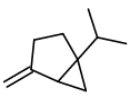
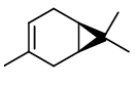
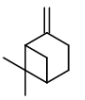
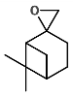
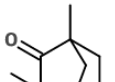
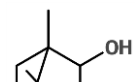
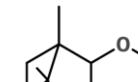
Table 2-3: Flavonoids found in Douglas-fir. *Glc = glucosyl group, Gal= galactosyl group, Me = methyl group. [91, 98-104]

2.3.4 Terpenoids

Various sizes of terpenoids have been identified in Douglas-fir including monoterpenoids, sesquiterpenoids, diterpenoids, and triterpenes. The most abundant and well-characterized terpenoids are the monoterpenoids and diterpenoids. These components are found

throughout the tree in different resins and are a large component of tall oil collected in pulping operations [95]. Tables 2- (4-7) contain the identified terpenoids.

Monoterpenoids and sesquiterpenoids are large components in the “essential oils” from needles, twigs, and branches. These components are known for their volatility and are commonly investigated in the gases above forests. Diterpenoids have been studied in the tall oil for uses such as rosin production and adhesives in asphalt. Diterpenoids have also been identified as one of the most toxic constituents of non-bleached pulp effluent streams [105, 106].

						
P-Cymenene	β-Cymene	P-Cymene	Thymol	p-Cymen-8-ol	Carvacrol	Anethole
						
γ-Terpinene	α-phellandrene	β-Phellandrene	α-Terpinene	Terpinen-4-ol	Terpinolene	
						
α-Terpineol	Limonene	Carvone	Piperitone	Dihydrocarveol	Isopulegol	Isopulegyl acetate
						
β-Cyclocitral	Sabinene	δ-3-Carene	α-Pinene	β-Pinene	β-Pinene Oxide	Fenchol
						
Fenchone	Borneol	Bornyl acetate	Camphor	Camphene	Trans-pinocarveol	

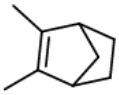
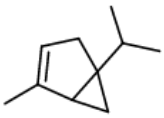
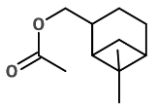
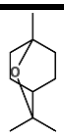
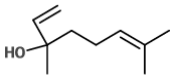
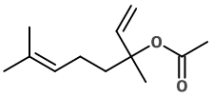
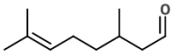
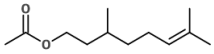
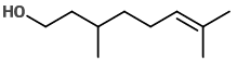
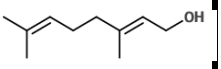
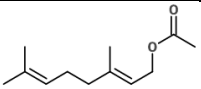
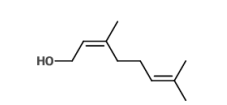
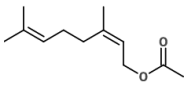
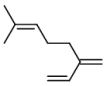
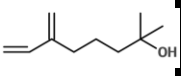
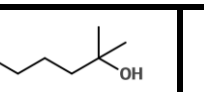
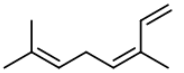
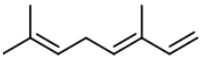
					
Trans-verbenol	Santene	α-Thujene	Myrtanol acetate	Tricyclene	1,8 -Cineole
					
Linalool	Linalyl acetate	Citronellal	Citronellyl acetate		
					
Citronellol	Geraniol	Geranyl acetate	Nerol	Neryl acetate	
					
Myrcene	Myrcenol	Dihydro-myrcenol	Cis-β-Ocimene	Trans-β-Ocimene	

Table 2-4: Monoterpenoids Identified in Douglas-fir [107-112]

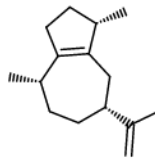
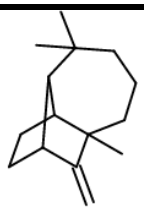
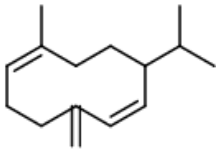
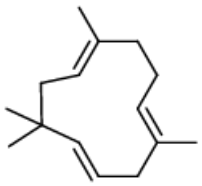
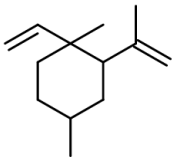
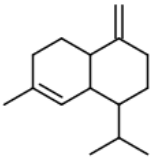
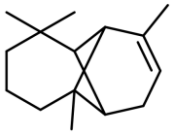
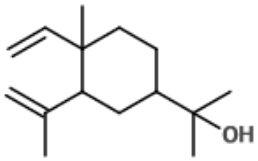
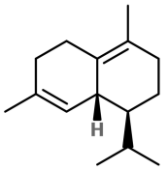
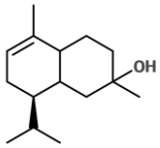
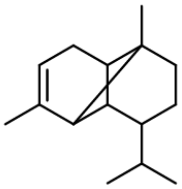
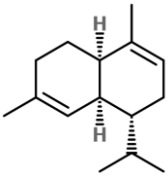
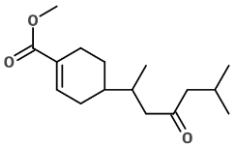
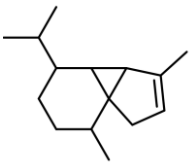
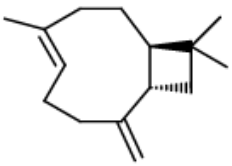
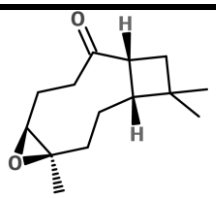
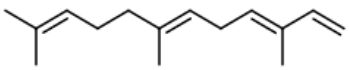
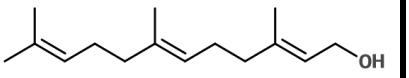
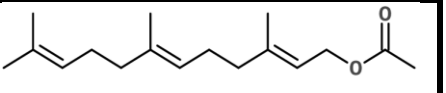
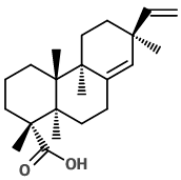
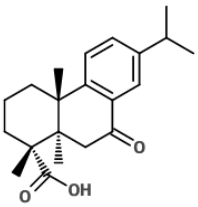
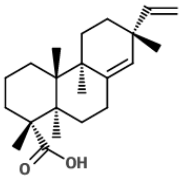
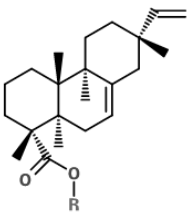
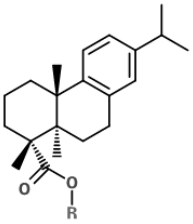
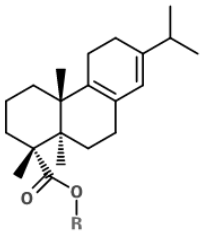
			
α-Guaiene	Longifolene	Germacrene D	Humulene
			
β-Elemene	Cadinene	α-Longipinene	Elemol
			
δ-Cadinene	δ-Cadinol	α-Copaene	α-Muurolene
			
Juvabione	α –Cubebene	β-Caryophyllene	β-Caryophyllene epoxide
			
Farnesene	Farnesol	Farnesyl Acetate	

Table 2-5: Sesquiterpenoids identified in Douglas-fir [107, 109-111, 113]

		
Pimaric Acid	Dehydro-9-ketoabietic acid	Sandaracopimaric Acid
		
R = H; Isopimaric Acid	R=H; Dehydroabietic Acid	R=H; Palustric Acid
R= Me; Methyl Isopimarate	R=Me; Methyl Dehydroabietate	R=Me; Methyl Palustrate

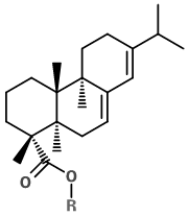
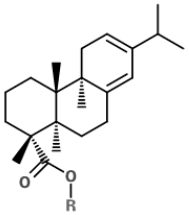
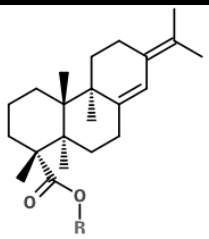
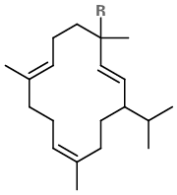
			
R=H; Abietic Acid	R=H; Levopimaric Acid	R=H; Neoabietic Acid	R= H; Thunbergene
R=Me; Methyl Abietate	R=Me; Methyl Levopimarate	R=Me; Methyl Neoabietate	R=OH; Thunbergol

Table 2-6: Diterpenoids currently identified in Douglas-fir [95, 109, 114]

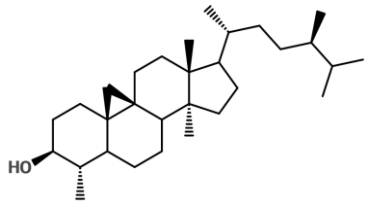
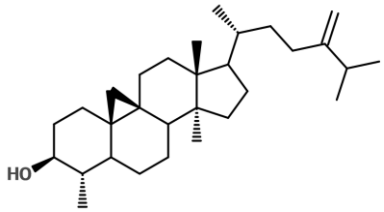
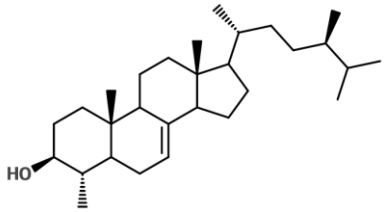
		
24R-Cycloeucalanol	Cycloeucalanol	24R-Methyllophenol

Table 2-7: Triterpenes identified in Douglas-fir [115]

2.3.5 Phytosterols

Both phytosterols and phytosterol esters appear in Douglas-fir. Sitosterol and Campesterol are specifically known to appear in the free state in the n-hexane extract of bark [116], while sitosterol, campesterol, sitostanol, campestanol, cycloartenol, and 24-methylenecycloartanol have been identified upon saponification of Douglas-fir wood [95, 117]. The fatty acid counterparts of the sterol esters have been identified as being in the range of C13:0 through C19:0 [87].

Phytosterols are bioactive compounds and have been examined for their effects in both mammals and fish. In the pharmaceutical field, they have been studied as an effective dietary supplement to reduce cholesterols and suggested mechanisms for this are reviewed by [118]. They have also been studied in pulping effluents and identified as suspected endocrine disruptors. Fish species that have been shown to be affected by high phytosterol concentrations include species such as brown trout [119] and mosquitofish [120].

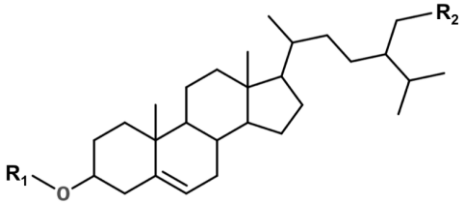
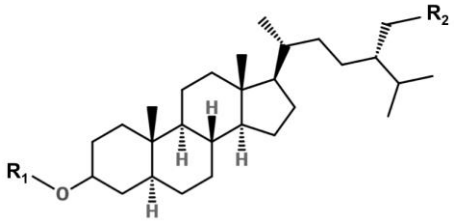
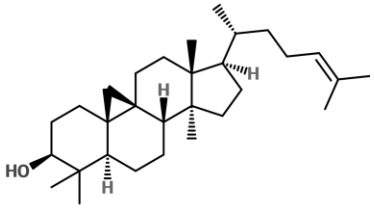
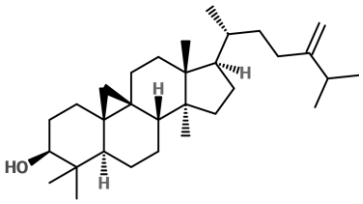
	
R_1=H, R_2=H; Campesterol	R_1=H, R_2=H; Campestanol
R_1=H, R_2=Me; Sitosterol	R_1=H, R_2=Me; Sitostanol
R_1=FA, R_2=H; Campesterol ester	R_1=FA, R_2=H; Campestanol ester
R_1=FA, R_2=Me; Sitosterol ester	R_1=FA, R_2=Me; Sitostanol ester
	
Cycloartenol	24-Methylenecycloartanol

Table 2-8: Phytosterols identified in Douglas-fir [87, 95, 117]

2.3.6 Lignans

Lignans, flavonolignans, and lignan glucosides have been identified in the sapwood, bark, callus resins, and knotwood of Douglas-fir. Like the phytosterols, attention has been paid to them due to their biological activity. Douglas-fir has three lignans that are enterolignan precursors, secoisolariciresinol, lariciresinol, and pinoresinol. These are being studied for their

antioxidant abilities and potential to reduce the risk of certain cancers and cardiovascular diseases [121].

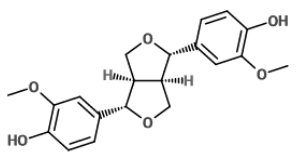
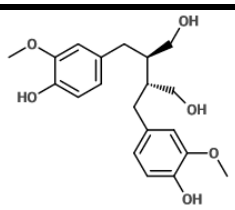
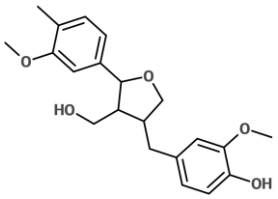
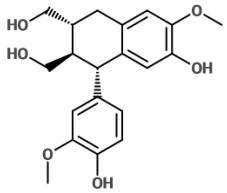
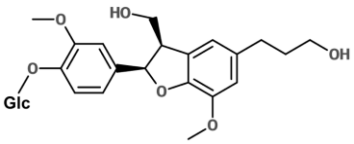
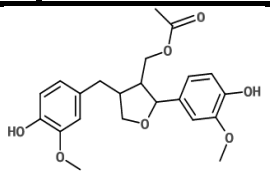
			
Pinoresinol	Secoisolariciresinol	Lariciresinol	Isolariciresinol
			
2,3-Dihydro-2-(4'-O-B-glucopyranosyl-3'-methoxyphenyl)-3-hydroxymethyl-5-(3-hydroxypropyl)-7-methoxybenzofuran		Lariciresinol-9-acetate	

Table 2-9: Lignans identified in Douglas-fir [113, 114]

2.3.7 Others

Various other molecules have been identified in Douglas-fir that do not fall into these previously mentioned classes and appear in needles, buds, cones, shoots, and other developing tissues. These include primary metabolites, molecules that are intermediates in different metabolic pathways (e.g. phenylpropanoid pathway), growth hormones, and other small molecules. This includes molecular classes such as auxins, cytokinins, and many others. Representative molecules of these types reported in the literature are included in Tables 2- (10-12).

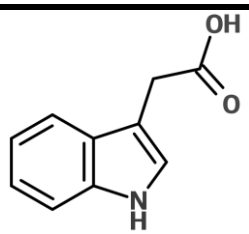
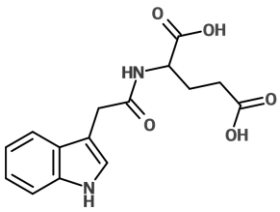
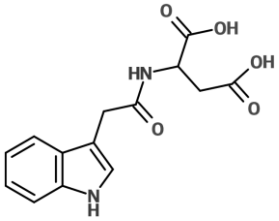
		
Indole-3-acetic acid (IAA)	Indole-3-acetic acid glutamate (IAA-Glu)	Indole-3-acetic acid aspartate (IAA-Asp)

Table 2-10: Identified Auxins [122, 123]

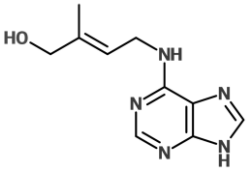
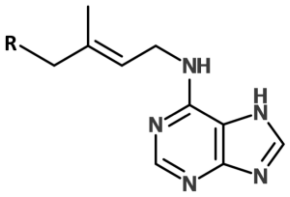
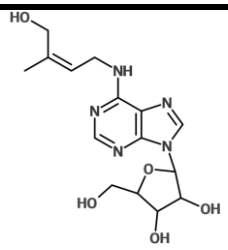
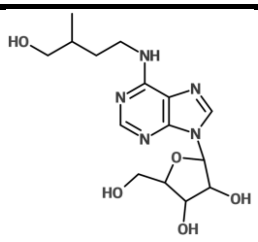
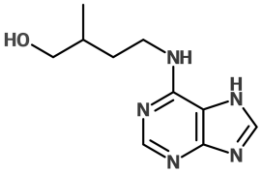
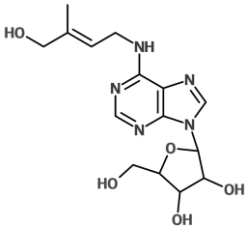
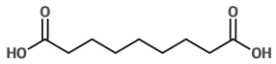
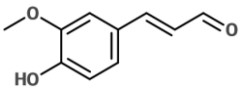
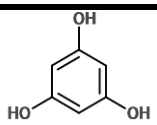
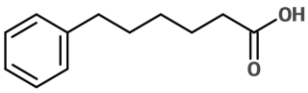
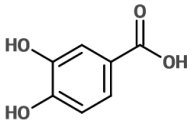
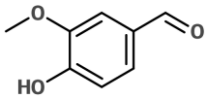
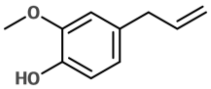
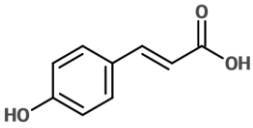
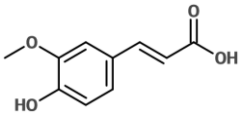
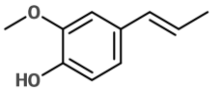
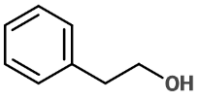
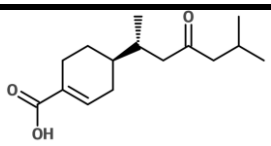
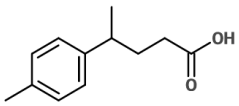
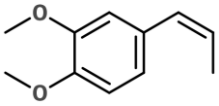
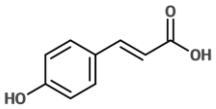
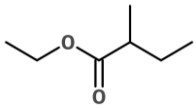
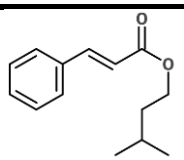
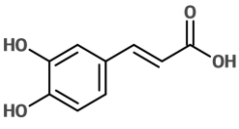
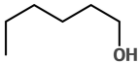
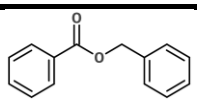
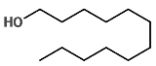
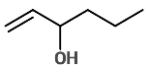
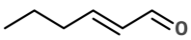
		
Trans-zeatin	R=Glc; trans-zeatin-*O*-glucoside	Cis-zeatin riboside
		
Dihydrozeatin riboside	Dihydrozeatin	Trans-zeatin riboside

Table 2-11: Identified Cytokinins *Glc = glucosyl group [122, 123]

			
Azelaic acid	Coniferyl aldehyde	Phloroglucinol	Butanal
			
Phenylhexanoic acid	Protocatechuic acid	Vanillin	Eugenol
			
4-Hydroxycinnamic acid	Ferulic acid	Isoeugenol	β-Phenylethanol
			
Todomatuic Acid	4-p-Tolylvaleric acid	Methylisoeugenol	p-Coumaric acid
			
Ethyl 2-methylbutyrate	Isoamyl cinnamate	Caffeic acid	Hexanol
			

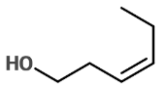
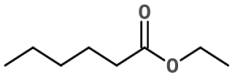
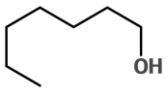
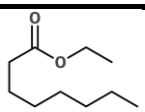
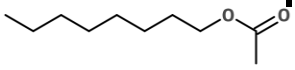
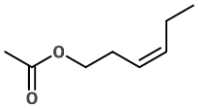
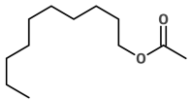
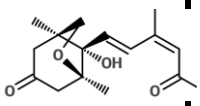
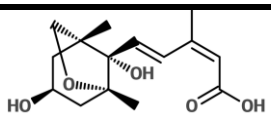
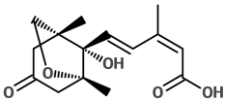
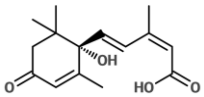
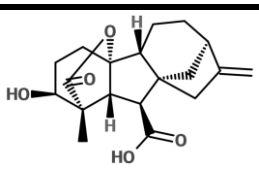
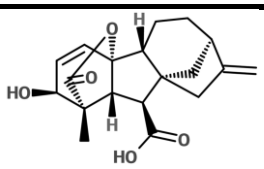
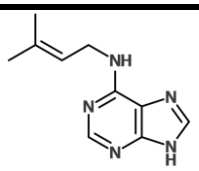
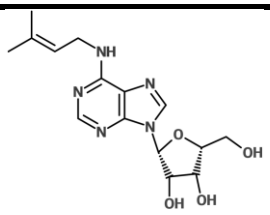
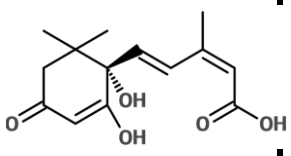
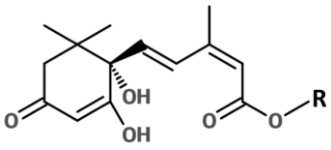
Benzyl benzoate	Dodecanol	Hexen-3-ol	2-Hexenal
			
3Z-Hexen-1-ol	Ethyl caproate	Heptanol	Ethyl caprylate
			
Octyl acetate	3Z-Hexenyl acetate	Decyl acetate	Neophaseic acid
			
Dihydrophaseic acid	Phaseic Acid	Abscisic acid	
			
Gibberelin A_4	Gibberelin A_7	Isopentenyl adenine	
			
Isopentenyl adenosine	7’-hydroxy Abscisic acid	R=Glc; Abscisic acid glucose ester	

Table 2-12: Other molecules reported in Douglas fir. [94, 102, 113, 114, 124]

2.4 Extraction methods and extractive concentrations in slash tissues

Knowledge of the identity and concentrations of the extractives present in Douglas-fir is essential to understanding their implications for biofuel processing and other commercial ventures. In this section we identify how extractives in Douglas-fir have been extracted, and what is known about concentrations of extractives in the tissues of most interest for biofuels (i.e. bark, sapwood, and heartwood). Special emphasis is given to chemical components that could conceivably be extracted as side products during processing, and that are likely to represent ≥0.1% of the biomass feedstock. Comparing differences between the varieties of Douglas-fir is complicated by the fact that many studies do not specifying which variety was studied. However, when the geographical origin of a sample is specified, we mention it as a means to speculate about the variety studied.

2.4.1 Proanthocyanidins and Phlobaphenes

Condensed tannins have been extracted with a variety of solvent including water [39], methanol [102], water-ether [125], water-acetone [101], and water-methanol [94] solutions. Chromatography (paper, thin layer, or HPLC), NMR, MS, and wet-chemical assays often are used to characterize the smaller chain components. Larger chains are often degraded into smaller units before characterization, with common methods of doing this including reaction with phloroglucinol [89] and thiolysis [101]. Colorimetric assays have also been used to characterize total amounts of proanthocyanidins and polyphenols in tissues. The butanol-HCl assay has also been a common characterization tool, in which the proanthocyanidins are colored by iron salts after which their concentration can be measured by their optical absorbance at 550nm [126]. The Folin-Ciocalteu's phenol reagent assay is another colorometric assay that has also been used to identify the total amount of polyphenols in the sapwood and heartwood [101, 127]. Crude

phlobaphenes have been identified in the methanol extract [92], benzene extract [128], and acetone-water solutions [101].

Reports for condensed tannins in Douglas-fir bark have ranged between 7.5 - 18% of the o.d. mass, with higher tannin concentrations observed at the top of the tree than the bottom [85]. A recent report by Pan et al. (2013) of bark from a Washington state pulp mill found a value of 13.2% air-dry mass of tannins by extraction with dichloromethane then a methanol-water solution. In addition to tannins, phlobaphenes have been reported to be 3.2% of outer bark weight by Foo & Karchesy (1989d) by methanol extraction, however they were unable to find phlobaphenes in the inner bark [92].

The sapwood and heartwood proanthocyanidins and phlobaphenes are present at a much lower concentration than in the bark, and have not been as thoroughly characterized. Dellus et al. (1997) extracted polyphenols in heartwood and sapwood in Douglas-fir wood from France with an acetone-water solution. They found that procyanidins (a form of condensed tannins) were 0.1% of the dry weight. Procyanidin identification in the heartwood was much more difficult however, and they instead opted to create a pseudo component called “phenolic polymer”, which they identified at 0.51% of the dry weight. It was believed that the procyanidins oxidized with lignans and other components at the sapwood/heartwood interface to form complex phlobaphenes. While oxidation to form phlobaphenes is likely to occur, both proanthocyanidins and phlobaphenes were probably in the phenolic polymer. Indeed, in 1949 Graham & Kurth did specifically identify tannin materials in heartwood [125]. Heartwood crude tannins were found to be 0.23% of the wood weight in the water-soluble portion of ether extracts. With these two reports, it can be surmised that both species are likely present in the heartwood (an estimate would be 0.2% and 0.4% of the o.d. weight, respectively).

2.4.2 Waxes

Early work on the waxes of Douglas fir was focused on extracting two types of waxes from bark, one that was soluble in hexane or petroleum type hydrocarbons, and another that is isolated using benzene or chlorinated hydrocarbons after the hexane extraction[85, 87, 128]. Specific solvents used include diethyl ether [95] hexane [128], benzene , and dichloromethane [94]. Ferulic acid esters are one of the main components of the waxes and have been shown to be in the hexane extract, but not the benzene soluble wax. Glycerol (indicating triglycerides, post saponification) was specifically found by Kurth (1950) in saponified benzene extracted wax. Foster et al. (1980) studied the sapwood and heartwood using diethyl ether as a solvent and which was used to identify triglycerides and small amounts of free fatty acids.

Early work by Kurth reported that the n-hexane soluble fraction is about 5% of the o.d. weight of bark, while the benzene wax is about 2.5% [85]. These two fractions contain large amounts of ferulic acid ester and triglycerides, but also contain other components such as phytosterol esters. Laver & Fang (1989) studied the ferulic acid esters in the hexane soluble fraction of Douglas-fir bark from western Oregon and reported them to be 1.4% of the bark (dry-weight basis). In Foster et al.'s (1980) diethyl ether extract of sapwood and heartwood from Colorado they found that readily saponifiable triglycerides accounted for 0.319% of the o.d. sapwood and that free fatty acids only accounted for 0.01% of the sapwood or heartwood mass. With these studies it is reasonable to expect that waxes are around 6%, 0.3%, and 0.4% of the o.d. weight of bark, sapwood, and heartwood respectively.

2.4.3 Flavonoids

Flavonoids have been extracted with various polar solvents including alcohols, alcohol-benzene, acetone, hot water [102, 103], alcohol-water [90, 101, 104], and ethyl acetate [99]. Methanol has been a very popular solvent due to its ability to dissolve both hydrophobic flavonoids such as quercetin [129] and more hydrophilic flavonoids such as dihydroquercetin [130]. Flavonoids commonly show up in the same extracts as tannins and also have been found in the benzene soluble wax extract [128]. They also are found in the Folin-Ciocalteu's phenol reagent assay for total polyphenols.

Dellus et al.'s (1997) study of the sapwood and heartwood from France identified total polyphenols (e.g. proanthocyanidins and flavonoids) as 0.7% and 2.4% of the wood dry weight (not o.d. however). The bark of Douglas-fir is known for being rich in dihydroquercetin, with some reports indicating that dihydroquercetin was 4.5-7.5% of the o.d. bark mass [85]. Dihydroquercetin shows up substantially in the heartwood (1.5% dry weight reported), however a glucoside of dihydroquercetin, dihydroquercetin-3'-*O*-glucoside has been reported to be more abundant than the aglycone in the sapwood (0.2% and 0.04% respectively) [101]. With this information and that from the proanthocyanidins and phlobaphenes section, reasonable estimates for flavonoids in the bark, sapwood, and heartwood would be around 6%, 0.5%, and 2% of the o.d. mass, respectively.

2.4.4 Terpenoids

Terpenoids have been isolated through both steam distillation and solvent extraction. Steam distillation has commonly been used to identify monoterpenoids and sesquiterpenoids [88, 107, 108, 110, 112]. However, it is not as reliable a technique for identifying species inside the actual tissue as solvent extraction, since steam distillation may alter certain compounds in the bark [87].

Solvent extraction of Douglas-fir terpenoids has been carried out with dichloromethane [94], diethyl ether [95, 111, 115], chloroform [131], and light petroleum [109].

Foster et al. (1980) studied the resin acids of Douglas-fir from Colorado and found them to be 0.2% and 0.27% of the o.d. weight of the sapwood and heartwood respectively. Pan et al. (2013) studied the resin acids in bark of Douglas-fir and reported them to be 0.4780% of the bark o.d. mass. Resin acids have been the focus of most of the literature due to their toxic nature, and other terpenes were of much less interest. There is a wealth of information on monoterpenoids and sesquiterpenoids in needles and young growth of Douglas-fir around the world, but their presence in the sapwood, heartwood, and bark has been studied much less. However, Erdtman et al. (1968) studied monoterpenoids, sesquiterpenoids, and diterpenoids (acidic and neutral moieties) in pocket resin of wood from Oregon. In the study, α-pinene and limonene were the most abundant monoterpenoids (both 6.7% of the pocket resin), Longifolene was the only sesquiterpenoid identified (0.2% of the pocket resin), levopimaric and palustric acid were the most abundant resin acids (23.8% combined), and thunbergol was the most abundant neutral diterpenoid (11.4%). Until more extensive studies of terpenoids in tissues relevant for logging slash are undertaken, the three mentioned studies can be used to provide an estimate of the amount of terpenoids that can be found in typical bark, sapwood, and heartwood. If resin acids are 0.27% of the heartwood and are 54% of the resin in the wood, than the other terpenoid compounds should correspondingly be about 0.2% of the o.d. weight. Similar deductions with the heartwood and bark can lead to estimates of terpenoids in the bark, sapwood, and heartwood being around 0.9%, 0.2%, and 0.4% of the o.d. mass [94, 95, 109].

2.4.5 Phytosterols

Phytosterols from Douglas-fir have been extracted with dichloromethane [94], diethyl ether [95], and n-hexane [132]. Phytosterols are difficult to characterize precisely because they form esters with a variety of fatty acids. Studies commonly will saponify the phytosterols and then report on the non-ester species they find [95, 117].

Foster et al. (1980) reported that 35% of a fraction of their diethyl ether extract was predominantly sterol esters (0.1925% of the o.d. weight of sapwood). Sitosterol was 25% of the phytosterols, stigmastanol 25%, campestanol 15%, and 5% each of cycloartenol and 24-methylenecycloartanol. In a report by Fischer et al. (1981), they also reported similar yields of 1.2 kg per green tonne of wood with sitosterol at 51% of the phytosterols, campesterol at 30%, sitostanol (stigmastanol) at 14%, and campestanol at 5%. A reasonable expectation for sterols in different tissues is about 0.2-0.3% of the o.d. weight.

2.4.6 Lignans

Lignans have been found in methanol-water extracts [101], ethanol [114], and a method involving successive solvents of increasing polarity [113] They have been reported to be 0.022% of the dry weight of sapwood [101], 4.8% and 7.5% by weight of the callus resin from Finnish Douglas-fir trees [114], and 9% of the of the acetone extract of knotwood from French Douglas-fir trees [113]. Of the extractives classes identified here, it is expected that this class of extractives will show up in the smallest proportions.

2.5 Estimating logging slash composition

Logging slash is comprised of branches, treetops, and other unmerchantable timber; however, the bulk of the available literature has studied extractives in the bark, sapwood, heartwood, and needles. Our estimates for Douglas-fir slash treat the slash as if it were a mixture

of bark, sapwood, and heartwood (needles fall off at the field site, and stumps/roots are not collected for slash). The proportions of bark, sapwood, and heartwood to be used for the compositional slash estimate come from data on a Douglas-fir forestry residue labeled FS-03 in Leu et al. (2013). FS-03 has well characterized glucan and Klason lignin values for the overall slash as well as its bark and wood (sapwood and heartwood not distinguished). Assuming bark and wood make up the FS-03 slash, a system of two equations and two unknowns can be solved with the glucan and lignin to determine bark/wood proportions as has been done in previous reports [133]. This results in a model of approximately 76% "wood" and 23% bark by mass (N.B., there is a small typo or calculational error in [134] that states this as 3% bark). To link FS-03 to the detailed chemical analysis in the literature, "wood" must be further segregated into sapwood and heartwood. We use a mass balance to relate the overall wood carbohydrate data reported in Leu et al. (2013) to the individual Douglas-fir sapwood and heartwood carbohydrate data reported in [84]. This gives an estimate of 57% heartwood and 43% sapwood in the FS-03 wood residues. Combining all these data, the final FS-03 residue compositional model we use is 23% bark, 33% sapwood, and 43% heartwood overall. For later discussion, we will assume that the missing 1% is extra dirt and ash or some other component that falls out of the process. For modeling analysis, we will scale chemical component values by 100/99 to achieve mass balance.

2.5.1 Combined estimates of molecular classes

Table 2-13 takes information from previous sections and accumulates estimated concentrations of extractives in common tissues in Douglas-fir, bark, sapwood, and heartwood. The slash model also presented in Table 2-13 is intended to provide estimates of the compositions in feedstocks of Douglas-fir tissues and logging slash, by category of extractive

compound. Appendix A collates all the specific molecular details described above, by individual extractive compounds, in a self-consistent, mass conserving manner with Table 2-13.

Extractive Type	**Bark**	**Sapwood**	**Heartwood**	**FS-03 Slash**
Proanthocyanidins and Phlobaphenes [85, 94, 101, 125]	13	0.2	0.6	3.3
Flavonoids [85, 101, 102, 125, 135]	*6*	0.5	2	2.4
Waxes [85, 93, 95]	6	0.3	0.4	1.6
Terpenoids [94, 95, 109, 136]	0.9	0.3	0.5	0.5
Phytosterols [95, 117, 136]	0.2	0.2	0.3	0.2
Lignans [101]	*	0.02	ND	*

Table 2-13: Estimates for the concentrations of extractives in Douglas-fir tissues that are likely to be found in forestry slash used for biofuel production (numbers provided are o.d. mass percent). * = not enough information to estimate; ND= not detected

2.6 Implications: Molecular partitioning of extractives in biofuel process streams

The previous sections have shown that extractives are a diverse set of molecules with diverse molecular properties and make up approximately 8% of the o.d. mass of the Douglas-fir slash feedstock. Understanding the concentration and distribution of extractive molecules is an essential step for thoroughly describing a biofuel process. To this end, we have assembled a simple ASPEN simulation that provides a rich molecular description of extractives throughout the process.

A simple process flow diagram for the simulation is presented in Figure 2-1. The process output is a minimally processed (MP) sugar concentrate from Douglas-fir slash and represents the first stage of a two-stage biofuel strategy in which a saccharified and concentrated sugar product is created at one facility then transported to a separate location for large scale fermentation [137].

ASPEN simulations use a mass-conserving model for the slash feedstock composition (Table 2-14). The slash feedstock model includes representative molecules from Table 2-13, or a subclass of extractives (such as diterpenoids and monoterpenoids), where experimentally validated thermodynamics was available from the literature. Values for the carbohydrates, lignin, protein, and ash found in literature for bark, sapwood, and heartwood, were combined in the same proportions as the extractives to give an estimate for slash [84, 134, 138, 139].

Components	Slash (Kg/hr)	SSL (Kg/hr)	Evaporator Waste(Kg/hr)	MP Sugar Stream (Kg/hr)
Water	400.00	296.25	382.16	37.010
Holocellulose	61.919			
Polymeric Sugars				12.013
Monomeric Sugars		7.549		44.909
Lignin	29.596			
Soluble Lignin		7.670	0.003	2.681
Insoluble Lignin				19.242
Furfural		0.666	0.225	0.007
Hydroxymethylfurfural		0.299	0.001	0.105
Acetic Acid		0.448	0.148	0.008
Ash	0.202			0.202
Protein	0.202	0.149		0.052
Tannin	3.131	1.160		0.406
Phlobaphene	0.202			0.202
Epicatechin – (4 β) – sulphonate		1.481		0.518
Dihydroquercetin-3'-O-glucoside	1.212	0.444		0.158
Dihydroquercetin	1.212	0.355		0.645
Quercetin		0.011		0.595
Triglyceride	1.313			0.657
Glycerol		0.039		0.014
Lignoceric Acid				0.659
Ferulic Acid ester	0.303			0.152
Ferulic Acid		0.043		0.015
Behenyl Alcohol				0.098
Palustric Acid	0.202	0.005		0.197
Thunbergol	0.202			0.202
Sitosterol Ester	0.101			0.051
Sitosterol	0.101	0.001		0.129
α-Pinene	0.101	0.037	0.013	0.001
p-Cymene		0.037	0.013	0.001
CTec2 Enzyme				0.468
Non-Carbohydrate Extractives	8.08	3.579	0.026	4.685

Table 2-14: Results of simulation to identify extractive propagation through biofuel processing. The slash feedstock model is representative of the FS-03 forest residue as described above [134]. A basis of 100 kg/hr is used.

Figure 2-1 has four main sections that simulate a sugar production facility: sulfite pretreatment, collection of spent sulfite liquor (SSL) via pressing, enzymatic saccharification, and concentrating the sugar product with an evaporator. A Solid-Liquid Partitioning step is present as a dashed box in the simulation after pretreatment. This step ensures partially soluble solid extractive species are fully equilibrated with the aqueous stream prior to pressing, but it is not a physical operation at a process facility, and hence we denote it's virtual nature with a dashed box. All process simulations used ASPEN Plus V8.6. Thermodynamic property predictions used NRTL with UNIFAC unless direct comparisons to experimental data suggested the need for direct input of key thermodynamic data.

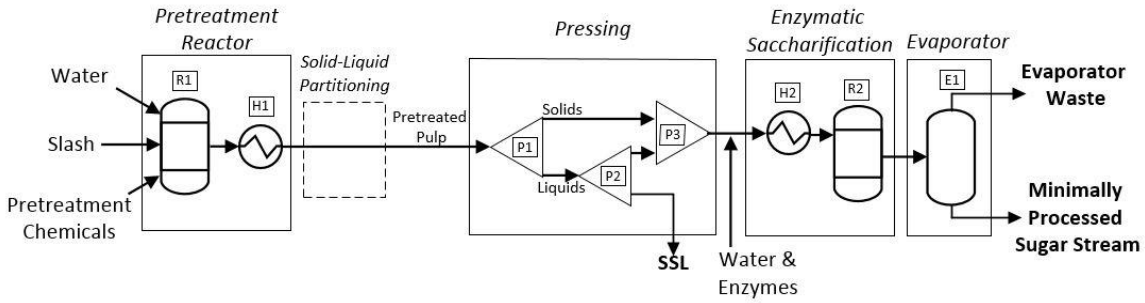


Figure 2-1: Process flow diagram of simulation used to create a minimally processed sugar concentrate from Douglas fir slash. Individual ASPEN blocks have been grouped together in the boxes to identify the process step they are intended to simulate. R1&2 are reactors. H1&2 are heat exchanger. P1&2 are splitters, P3 is a mixer. E1 is an evaporator. The composition of the slash, spent sulfite liquor (SSL), evaporator waste, and minimally processed sugar stream are presented in Table 2-14.

The simulated pretreatment step is intended to be representative of sulfite/bisulfite processes. Yield data from two studies are combined and applied here for estimation purposes. Conversion of holocellulose to polymeric sugars, monomeric sugars, and furans come from a report on mild bisulfite pretreatment from Gao et al. (2013) while lignin reaction to form soluble or insoluble lignin comes from a SPORL pretreatment from Leu et al. (2013).

Chemical reactions involving extractive species in sulfite processing have been identified in various studies. Table 2-15 contains representative reactions that are included in the simulation. The six reactions presented in Table 2-15 were selected to demonstrate the possible effects of pretreatment chemistry on different classes of molecules. Neither kinetic data nor individual yields are available for the conditions we are simulating here, so we simply assume 50% reactant conversion for the purpose of tracking the flow of extractive species reactants and products through the process.

Reactants		Products (Reference)
Dihydroquercetin 3'-O-glucoside + H_2O	→	dihydroquercetin + glucose (monomeric sugar)
2 Dihydroquercetin + $2HSO_3^-$	→	2 quercetin + $S_2O_3^{2-}$ +3 H_2O [30]
Condensed tannin + HSO_3^-	→	condensed tannin + epicatechin – (4 β) – sulphonate[83]
Ferulic acid ester + H_2O	→	ferulic acid + behenyl alcohol
Sitosterol ester + H_2O	→	sitosterol + lignoceric acid
α-Pinene + $2HSO_3^-$	→	p-cymene + $S_2O_3^{2-}$ + $3H_2O$ [30, 140]
Triglycerides + 3 H_2O	→	glycerol + 3 lignoceric Acid

Table 2-15: Representative reactions of extractives and literature citations for more complex reactions than straightforward hydrolysis.

The solid-liquid partitioning and pressing stages are carried out at 25°C and create two streams, a 40% solids stream and a liquid spent sulfite liquor (SSL) stream. Thermodynamic data used for the solid-liquid partitioning step came from [105, 129, 141-143]. Water and enzymes are then added back into the solids stream for saccharification with yield and enzyme information used from the Gao et al. (2013) study. The evaporation stage is modeled with a flash drum at 50°C and produces the MP sugar stream, which is 30% water (mass basis). Since the purpose of the simulation was to evaluate the partitioning of extractives into the product and waste streams, we did not do next-level simulation such as cost or energy integration and optimization, or detailed unit designs, so scale-up factors on the 100 kg/hr mass basis of these calculations are linear.

Extractives partition between the SSL and MP Sugar stream roughly in the ratio of 45:55 in our simulation, and only a small amount of extractives leave in the evaporator waste. Polymeric materials and flavonoids end up in all streams except the evaporator waste. Waxes are likely to remain with the solid phase unless they are sulfonated during pretreatment to form soaps. The molecules of most concern for wastewater treatment, phytosterols and diterpenoids, will predominantly remain with the solid product, but do show up in the SSL and other liquid streams. The most volatile extractives, the monoterpenoids, will be found in the emissions from the evaporator steps. Other emissions in the waste stream from the evaporators include the acetic acid, furfural and a small amount of HMF.

Optimized biofuel processes will need to understand how extractives affect each of the processing steps. What follows is a discussion of extractives of interest for specific biofuel processing steps and considerations for an overall biofuel facility.

2.6.1 Pretreatment

Extractives can undergo a variety of reaction in pretreatment and there is a need to investigate which ones predominantly occur. Dihydroquercetin and condensed tannins have been specifically studied for their reactivity in sulfite processing [30, 83] and it can be expected that other components such as waxes will also be reactive to some extent, depending on details of the pretreatment conditions [144]. Treatment by sulfuric acid further polymerizes condensed tannins while treatment by sulfite breaks them up into smaller sulfonated units [43, 83, 125]. With sulfite/bisulfite biofuel treatments using chemicals such as sulfuric acid, sulfite, and bisulfite [80, 137], it is unclear which preferentially happens. In addition, pretreatment products can have very different aqueous solubilities then the starting material (e.g. dihydroquercetin glucosides vs. quercetin), affecting the feasibility of extracting them for side products after pretreatment. Dihydroquercetin and its glucosides can be expected to be 0.5-5% of the slash feedstock mass, with this same range expected for the polymeric polyphenol and waxes categories. Soluble tannins have more economic potential then the solid phlobaphenes, but which chemistries are observed and their kinetics still need to be studied.

2.6.2 Saccharification

Non-cellulosic components are often lumped into a lignin, or Klason lignin, category and this category is known to bind to cellulases and obstruct the saccharification step in biofuel production [145]. Extractives may be some of the key culprits. Depending on the amount of bark in the slash and processing, saccharification can expect from 10-100 g of proanthocyanidins and phlobaphenes/kg o.d. dry sugars, 10-100g/kg (o.d. dry sugars) of flavonoids, 10-100g/kg of waxes, 1-10 g/kg of diterpenoids and monoterpenoids, and 1-10 g/kg of phytosterols. The proanthocyanidins and flavonoids are of particular interest to saccharification since some are

known for their astringent properties and have been seen to hinder cellulases [146] . Many of these compounds have significant water solubility, and could be removed through washing or pressing. However, some biofuel pretreatment techniques produce lignosulfonates that act as possible surfactants to aid saccharification and would also be lost [147]. Research comparing the effects of these two competing processes is needed to determine whether washing/pressing the biomass after pretreatment is advantageous or not.

2.6.3 Extractives sent to fermentation

Our simulation and other literature indicate that sulfite based degradation leaves many of the extractives in the solid phase throughout processing [148]. This corresponds to large proportions of extractives following the holocellulose through the biofuel process and into the fermentation feed in our simulation. Extractives that show up in significant quantities in the feed for fermentation include waxes, proanthocyanidins and phlobaphenes, hydrophobic flavonoids, diterpenoids, and phytosterols. Diterpenoids are of particular interest because of their known toxicity [105] as well as dihydroquercetin for its known antifungal properties [46]. Our simulation estimate that the concentrated sugar product could be 1-5% by mass proanthocyanidins, phlobaphenes and flavonoids (combined), 1-5% waxes, 0.1-0.5% terpenoids, and 0.1-0.5% phytosterols, depending on specific details of the slash feedstock. If these components are found to significantly affect fermentation processes, then removing the solids through decanting becomes an attractive processing step.

2.6.4 Waste Treatment

Extractives have been found to be some of the most toxic components of mechanically and sulfite pretreated pulp waste streams and will also likely have a similar role for biofuel

waste. This toxicity is due to compounds such as the diterpenoids whose parts per million aqueous solubility are toxic for fish [105] and phytosterols which are possible endocrine disruptors [120]. In addition, our simulation indicates that aqueous waste streams are likely to carry significant amounts of extractive molecules such as procyanidins and soluble flavonoids. Of the gas emissions, the mono-terpenoids are the only ones that appear in these streams appreciable amounts (~10:1 abundance in evaporator waste compared to MP sugar stream). These extractive compounds are well known in the pulp and paper industry however and methods for handling them can be adapted from established techniques [143, 148, 149].

2.7 Conclusion

Studies on the composition of Douglas-fir have been conducted for over a century and provide a wealth of information on the phytochemical within this feedstock. There is considerable literature available for Douglas-fir extractives in general, but there are still large gaps in our knowledge such as the differences between extractives in var. *menziesii* or var. *glauca* in tissues relevant for biofuels. Also, thermodynamics parameters for many extractive molecules have yet to be determined. Optimized biofuel processes will need to understand how extractives affect each processing step. Extractives are both possible revenue sources and inhibitors at different point in chemical processing. The accumulated knowledge presented here provides a starting point for incorporating next-level details of extractive chemistry, biochemistry, and process-science in our understanding of biofuel production.

2.8 Appendix 1

2.8.1 Sapwood, Heartwood, and Bark Detailed Models

Molecularly specific compositional models for softwood, hardwood, bark, and slash were assembled and used to come up with the specific compositional estimates in Table 2-13. This was done by using a combination of literature data, heuristics, and mass balances to estimate the concentrations of individual molecules for extractive classes that available literature can justify to be at least 0.1% of the forestry biomass (o.d. basis). This excludes classes such as triterpenes, lignans, and sesquiterpenes [101, 109, 115]. These models are presented in Table 2-16.

The concentrations of the most commonly studied sugars are included for the different tissue types (i.e., glucan, mannan, xylan, galactan, and arabinan). Analyses of other units of hemicellulose such as, 4-O-methyl-glucuronic acid, glucuronic acid, galacturonic acid, and rhamnose have been carried out for the heartwood and sapwood [84, 150] and values for these components are included; these compounds are not regularly reported in studies of the bark. Lignin values reported here come from studies that measured acid soluble and insoluble lignin [84, 139]. Protein values have been reported for the sapwood, heartwood, and cambium [84, 138]. The value for protein in the cambium has been used here to estimate the protein composition of the entire region we label “bark”. We used reported ash values for sapwood, heartwood and bark [84, 134].

In many instances data for the composition of specific extractive compounds were not directly quantified in the literature for sapwood, heartwood, and/or bark. When this was the case, estimates were made based on a synthesis of the best available data. In the cases of flavonoids and phytosterols in the full model, we found data for the total amount of that extractive category in a tissue, and then estimated the individual molecular composition using mass-conserving heuristics and other molecular characterizations we found in diverse literature reports. With the diterpenes and monoterpenes, we use reported values for components that have been quantified

in the sapwood and heartwood [95] and estimated other components from their known relative abundance compared to the quantified compounds.

Dellus et al. (1997) reported the total polyphenol extractives (polymeric polyphenols and flavonoids) in sapwood and heartwood as .7% and 2.4% of the dry mass respectively. In this report, they were also able to individually identify the species responsible for 52% and 86% of the polyphenols in each tissue type, respectively. We took the individual species identified, and had them represent the whole polyphenol category (excluding phlobaphenes though) by scaling their values up proportionally to meet the total measured polyphenol values. To account for the complex polymeric polyphenols in a manageable way in this model, we separated the polymeric polyphenols into water soluble "condensed tannins" and water insoluble "phlobaphenes". Specific types of condensed tannins or phlobaphenes are lumped into one these categories rather than using the diverse nomenclature associated with these compounds. For example, the procyanidin measurement of the sapwood from Dellus et. al (1997) are assigned to the condensed tannin category in the full compositional model and Table 2-16, and the phlobatannins from Graham & Kurth (1949) were used for condensed tannins in the heartwood (~0.2%). The polymers of dihydroquercetin, procyanidins, and other polyphenols Dellus et al. reported in the heartwood are assumed here to be comprised of condensed tannin and phlobaphene. The phlobaphene value in the heartwood was found by taking Dellus et al. (1997)'s adjusted values for "complex polymers" and subtracting Graham & Kurth (1949)'s value for phlobatannin (condensed tannin). In the bark, condensed tannins have been reported to be in the range of 7.5- 18% [85] and an intermediate value of 13% is assumed here to be similar to a recent report on bark characterization [94]. Studies have been undertaken to identify more polyphenol

species in the bark [102], but because of limited quantitative data, the only bark flavonoid value in our model is for Dihydroquercetin/Dihydroquercetin-glucoside (a value of 6% assumed here [85]).

Fats and Waxes values come from reports interested in the wax or tall oil and mass balances when total quantities of a category were known. Ferulic acid esters have been reported to be 1.4% of the dry weight of the bark [93], but reported data was not found for sapwood and heartwood. Ferulic acid esters and triglycerides show up in the hexane soluble wax of the bark (5-7.5% by dry mass [85], an intermediate value of 6% assumed here) so the portion of hexane soluble wax not attributed to ferulic acid esters was used for the Triglycerides/Free Fatty Acid section of our bark model [85]. Foster et al. (1980) reported triglycerides in the sapwood were 58% of the neutral section of the diethyl ether extract (corresponding to 0.319% dry weight), and this same proportion of triglycerides in the neutrals was assumed for the heartwood (0.4%) [95]. The free fatty acids also reported by Foster et al. (1980) for sapwood and heartwood were added to the triglycerides for the Triglycerides/Free Fatty acid portion of the full compositional model. For further information on the constituents of fats and waxes, see [87, 94, 95].

Foster et al. (1980) identified values for acidic diterpenes (resin acids) in the sapwood and heartwood and Pan et al. (2013) in the bark with their values used in the model. The resin acid data was paired with a study by Erdtman et al. (1968) which reported the chemical composition of a wood resin sample to estimate the non-acidic diterpenes and monoterpenes. Values for the total phytosterols (including phytosterol esters) in the heartwood and sapwood come from Foster et al. (1980), i.e. 35% of their “neutrals” class in the sapwood and heartwood are used as phytosterols. We assume that 50% of these phytosterols are sterol esters before saponification. This is combined with a report from Fischer et al. (1981) which identified the individual species of sterols present after saponification. The sapwood phytosterols values are

also used as estimates for those in the bark since resin components are known to be similar in the sapwood and bark [136].

To complete the mass balance, the unaccounted remainders of the compositional models are treated as soluble carbohydrates such as pectin and called “Extractable holocellulose” in Table 2-16.

Chemical Species	**Bark**	**Sapwood**	**Heartwood**	**Slash**
(values correspond to % o.d. mass)				
Cellulose & Hemicellulose Components [84, 139]				
Glucan	*27.7*	*42.95*	*40.27*	37.9
Mannan	*3.0*	*11.90*	*11.11*	9.4
Xylan	*2.3*	*3.93*	*4.40*	3.7
Galactan	*1.6*	*2.56*	*2.41*	2.2
Arabinan	*3.4*	*0.58*	*0.56*	1.2
4-O-Methylglucuronic acid	*	*1.00*	*0.94*	0.7
Acetyl group	*	*1.2*	*0.57*	0.6
Galacturonic Acid	*	*0.09*	*0.11*	0.08
Glucuronic Acid	*	*0.00*	*0.08*	0.03
Extractable holocellulose	0.6	5.36	7.84	5.5
Total	38.6	69.6	68.3	61.3
Non-Holocellulose/Extractive Components [84, 134, 138, 139]				
Lignin	*34.5*	*28.62*	*27.66*	29.3

Protein	*0.22*	*0.18*	*0.16*	0.2
Ash	*0.7*	*0.14*	*0.08*	0.2
Total	35.4	28.9	27.9	29.7

Polymeric Polyphenols *85, 94, 101, 125*

Condensed Tannin	*13*	0.2	0.2	3.1
Phlobaphenes	*	*	0.4	0.2
Total	13	0.2	0.6	3.3

Flavonoids *85, 101, 102, 125, 135*

Dihydroquercetin/ Dihydroquercetin-glucosides	*6*	0.5	1.8	2.3
[5',5']-Bisdihydroquercetin	*	*	0.2	0.07
Dihydrokaempferol	*	*	0.03	0.01
Pinocembrin	*	*	0.003	0.001
Quercetin-3'-O glucoside	*	0.002	0	0.0007
Total	6	0.5	2.0	2.4

Fats and Waxes *85, 93, 95*

Triglycerides/Free Fatty Acids	4.6	*0.329*	0.4	1.3
Ferulic Acid Esters	*1.4*	*	*	0.3
Total	6	0.3	0.4	1.6

Diterpenes *94, 95, 109*

Dehydroabietic Acid	*0.3347*	*0.016*	*0.027*	0.09
Thunbergol	0.1	0.04	0.06	0.06
Palustric Acid	*0.0141*	*0.047*	*0.062*	0.05

Isopimaric Acid	*0.0679*	*0.012*	*0.073*	0.05
Abietic Acid	*0.0256*	*0.032*	*0.051*	0.04
Neoabietic Acid	*	*0.028*	*0.033*	0.02
Isopimarinal	0.04	0.02	0.02	0.02
Dehydro-9-ketoabietic acid	*0.0252*	*	*	0.01
Levopimaric Acid	*	*0.012*	*0.006*	0.01
Methyl Levopimarate and Methyl Palustrate	0.02	0.009	0.01	0.01
Sandaracopimaric Acid	*	*0.006*	*0.01*	0.01
Thunbergene	0.02	0.007	0.009	0.01
Methyl Isopimarate	0.005	0.002	0.003	0.003
Pimaric Acid	*0.0105*	*	*	0.002
Methyl Dehydroabietate	0.004	0.001	0.002	0.002
Methyl Abietate	0.002	0.001	0.001	0.001
Total	0.7	0.2	0.4	0.4

Phytosterols *95, 117, 136*

Sitosterol Fatty Acid Esters	0.05	0.05	0.06	0.05
Sitosterol	0.05	0.05	0.06	0.05
Campesterol Fatty Acid Esters	0.03	0.03	0.04	0.03
Campesterol	0.03	0.03	0.04	0.03
Sitostanol Fatty Acid Esters	0.01	0.01	0.02	0.01
Sitostanol	0.01	0.01	0.02	0.01
Campestanol Fatty Acid Esters	0.005	0.005	0.007	0.005

Campestanol	0.005	0.005	0.007	0.005
Total	0.2	0.2	0.3	0.2

Monoterpenes [95, 109, 136]

α-Pinene	0.02	0.02	0.03	0.03
Limonene	0.02	0.02	0.03	0.03
Terpineol	0.02	0.02	0.03	0.02
β-Pinene	0.007	0.007	0.01	0.01
δ-3-carene	0.002	0.002	0.003	0.002
Total	0.1	0.1	0.1	0.1

Extractive Totals

Noncarbohydrate Extractives	26	1.5	3.8	8.0
Extractives	26.6	6.9	11.6	13.5

Table 2-16: List of individual molecules in Douglas-fir for classes expected to be >0.01% of the o.d. mass. Slash values fit to the 99% characterization of Leu et al. [134]. Italicized values are taken directly from literature, while values that are not italicized are estimated from literature.

Chapter 3: Inhibition of Trichoderma reesei Cellulases by Aqueous Douglas-fir Bark Extractives

3.1 Chapter Summary

Extractives are a class of molecules that incorporate many diverse phytochemicals, including soluble components such as flavonoids, proanthocyanidins, and terpenes. Though concentrated in the bark, extractives can make up 8% of the non-carbohydrate dry mass of actual Douglas-fir forestry residues. Here we explore the role of Douglas-fir bark extractives and some individual extractive components such as taxifolin (a flavonoid), tannin (a proanthocyanidin), α-pinene and abietic acid (terpenes) on the initial rates of saccharification using cellulases from *Trichoderma reesei*. Initial rates of glucose production from crystalline cellulose are inhibited roughly 40% when bark extract and tannin are present; none of the other extractive components produced a measureable inhibition. The use of crystalline cellulose, amorphous cellulose, and cellobiose as substrates suggests that each major enzyme in the cellulase mixture—exoglucanase, endoglucanase and β-glucosidase—is inhibited by 15-25% in the presence of bark extract and tannin. β-glucosidase activity follows Michaelis-Menten Kinetics and Lineweaver-Burk plots were used to show that bark extract inhibits β-glucosidase via an uncompetitive binding mechanism. Inhibition of this type would indicate the tannin is affecting the substrate-enzyme complex. This work shows that extractives from Douglas-fir can inhibit saccharification at biofuel relevant concentrations and that the inhibition from whole extracts closely resembles the inhibition seen from condensed tannins. Tannin removal or neutralization techniques should be the focus for mitigating extractive inhibition from this biofuel feedstock.

3.2 Background

Forestry residues are an abundant, low environmental impact biomass feedstock that does not drive major land-use changes or compete with food resources [2, 3]. Douglas-fir residues are especially attractive in the Western United States as a biofuel feedstock owing to the species' pre-eminence in timberlands[29]. Pilot scale biofuel processes based on pretreatment and saccharification of this feedstock are in development[10]. Saccharification of pretreated Douglas-fir feedstocks relies on cellulase enzymes to hydrolyze the cellulose fraction into glucose. This process is known to be inhibited by many compounds found in biomass or created during pretreatment, such as polyphenols[51, 69], various saccharides[70, 71], and certain ions[72]. Cellulases used in industrial saccharification typically come as a "cocktail" that contain exoglucanases, endoglucanases, β-glucosidase, and other enzymes to break down the biomass[56]. Exoglucanases are a family of enzymes considered responsible for breaking down the crystalline cellulose regions (primarily) into cellobiose[151]. The endoglucanase family of enzymes (largely) breaks down amorphous regions into cellobiose or glucose, while the β-glucosidases break down cellobiose into glucose.

Little is known about how the complex mix of extractive molecules from Douglas-fir bark affects saccharification rates, even though similar molecules are reported to be strong inhibitors [51]. Douglas-fir extractives include tannins and many different phenols which are reported or expected inhibitor[51, 69, 74-76, 152]. Removal or neutralization of the specific inhibitors could lead to improved saccharification yields. However, the drawback to this is that common ways of removing such inhibition, such as washing and pressing steps or adding chemical additives to neutralize inhibitors, all add to the cost of biofuel production[77, 78]. A clearer understanding of the extent of inhibition would be useful for making decisions on how to handle these molecules.

This study seeks to quantify and characterize how water soluble Douglas-fir bark extractives inhibit exoglucanase, endoglucanase, and β-glucosidase activity at relevant biofuel concentrations. In forestry residues, non-carbohydrate extractives can make up roughly 8% of the total dry mass[40]. To get at the question of inhibition by extractive molecules, initial rate saccharification experiments were performed in the presence and absence of whole bark extracts and compared to experiments on several individual extractives expected in the bark extract[40]. Individual extractives that were investigated include phenolic species (tannin and taxifolin) and defensive molecules in the bark (α-pinene and abietic acid[47, 52]). We present the inhibitory impact of these Douglas-fir extractive molecules on cellulase enzymes from *Trichoderma reesei*, a non-proprietary but industrially relevant enzyme cocktail[19].

3.3 Results

Experiments were conducted to assess how extractives affect exoglucanases, endoglucanases, and β-glucosidases from *Trichoderma reesei*. To do this, crystalline cellulose, amorphous cellulose, and cellobiose were used as model substrates for initial rate saccharification experiments. These better defined substrates make it easier to pinpoint the origin of inhibition rather than using the more complex forestry residue biomass. Experiments with whole bark extracts, a condensed tannin supplement, taxifolin, α-pinene, or abietic acid are compared to experiments without any extractives.

3.3.1 Screening inhibition of saccharification by several extractive classes

The effects of different extractive molecules on the initial rate of glucose production in our standard enzyme cocktail are shown in Figure 3-1A for crystalline cellulose substrate. Bark extract and tannin both cause approximately 40% inhibition of glucose production while

taxifolin, α-pinene, and abietic acid showed no statistically significant inhibition. The statistical significance of inhibition was determined by comparing the mean yields for the control experiments (no extractive) and the test experiments (with extractive) through a two-sample t statistic using a conservative number of degrees of freedom (i.e., number of replicates in a single experiment-1). Differences are considered statistically relevant in this study if they are above a 90% confidence level. Figure 3-1A shows that most of the inhibitory effect of bark extractives on glucose production from crystalline cellulose can be tentatively assigned to tannins.

Exoglucanase is the primary cellulase in our enzyme cocktail responsible for initiating saccharification of the crystalline cellulose substrate. Glucose production from crystalline cellulose is primarily produced by the simplified two-step mechanism[151]:

$$\text{Crystalline Cellulose} + (x\text{-}1)\ H_2O \xrightarrow{\text{Exoglucanase}} x\ \text{Cellobiose} + x\ H_2O \xrightarrow{\beta\text{-Glucosidase}} 2x\ \text{Glucose}\ , \qquad \text{Eq. 1}$$

where x denotes the length of the cellulose chain (number of polymerized cellobiose units). The first step in Eq. 1 can be inhibited by the cellobiose or glucose products, so the use of initial reaction rate studies here helps ensure this effect is minimized.

In our crystalline cellulose, amorphous cellulose, and cellobiose initial rate saccharification experiments we used the Celluclast cellulase cocktail with no excess β-glucosidase added. This is counter to common practice in other studies[19, 24, 153] but allows us to more easily monitor the intermediate product cellobiose in Eq. 1. The cellobiose values can indicate if β-glucosidase and/or exoglucanases are inhibited. To identify if exoglucanases are inhibited specifically, we examine the total amount of cellobiose produced throughout the initial rate experiment.

"Equivalent cellobiose" can be computed by adding the measured cellobiose produced to one half the amount of measured glucose, per the stoichiometry of Eq. 1. Values for inhibition of equivalent cellobiose production are shown in Figure 3-1B. Both bark extract and tannin cause approximately 15% reduction in the total cellobiose product. This value serves as an upper bound for tannin binding inhibition of the exoglucanases and is in line with previous inhibition studies[51, 154]. The actual inhibition can be smaller due to slightly elevated cellobiose concentrations experienced because of β-glucosidase inhibition or because a modest percentage of glucose (<15%) is directly generated without going through the cellobiose intermediate[151]. Since the % inhibition for equivalent cellobiose and glucose are different by a significant amount, it means that the β-glucosidases are also inhibited by the bark extract and tannin.

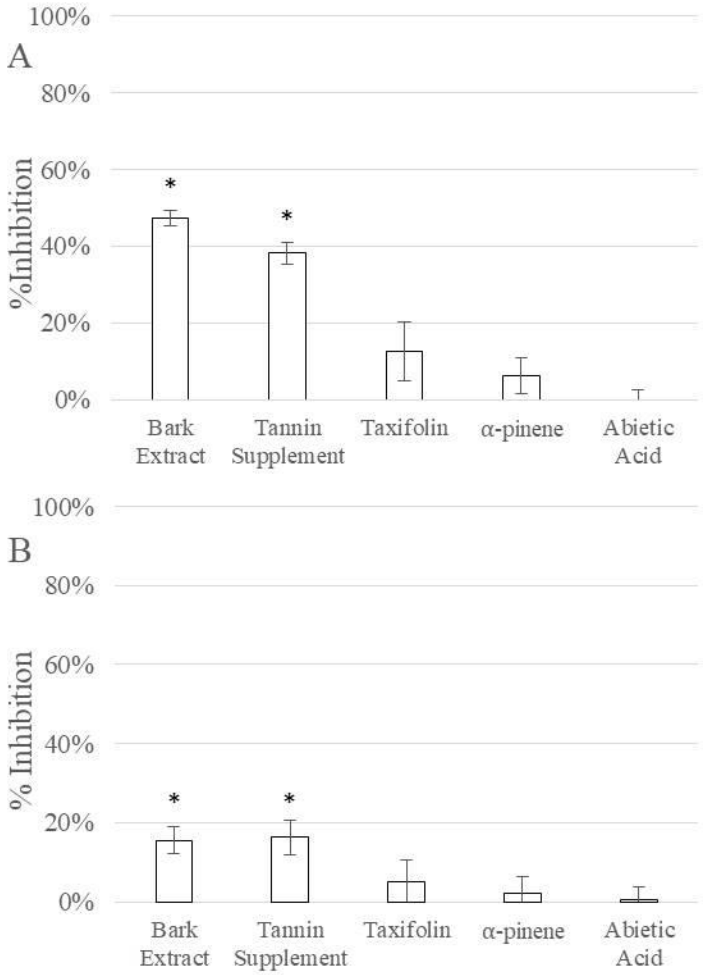


Figure 3-1**:** Inhibition of crystalline cellulose hydrolysis after 1 hour. A. Inhibition of the observed glucose due to the addition of extractives. B. Inhibition of the equivalent cellobiose due to addition of extractives. Error bars are standard deviation derived using error propagation for quotients. * denotes >95% confidence that the values is different than experiments without extractives. Both the bark extract and tannin supplement lead to significant inhibition and very similar inhibition amounts.

The mechanism for exoglucanase inhibition could come from two possible sources: (i) inhibitory binding of the extractives to the cellulases or (ii) inhibitory binding directly to the cellulose surface. To test if these extractives strongly adsorb onto the surface of cellulose and

affect the initial rates, experiments were conducted in which crystalline cellulose and bark extracts were first mixed together for 1 hour, allowing extractives to adsorb onto the cellulose. The extract supernatant solution was then removed and replaced with extractive-free buffer, leaving only the strongly, or irreversibly, bound extractives on the cellulose surface. Under these circumstances, inhibition would constitute evidence that extractives strongly or irreversibly bound to the cellulose surface are part of the inhibitory mechanism. Pre-adsorption experiments of this nature showed no statistically relevant inhibition however. Thus, the binding of extractives to cellulose is weak, highly-reversible, or may be wholly unimportant to inhibition. Data for testing the role of direct adsorption on the cellulose is presented in Supplementary Information. These experiments suggest that the most plausible route for inhibition is extractive binding directly to the cellulases in solution.

Parallel inhibition screening experiments can be performed using amorphous cellulose as the substrate, resulting in a more direct probe of extractive interactions with the endoglucanase component in our Celluclast cocktail. The primary mechanism for glucose production from amorphous cellulose substrate is a direct route[155]:

$$\text{Amorphous Cellulose} + (2x-1)\,H_2O \xrightarrow[\text{Endoglucanase}]{} 2x\ \text{Glucose} \quad \text{Eq. 2}$$

where x has the same meaning as Eq. 1. Though Eq. 2 is the dominant pathway, it has been reported that up to 25% of glucose production via endoglucanase and amorphous cellulose can involve a cellobiose intermediate, akin to Eq. 1[155].

Figure 3-2 shows that the bark extract and tannin cause approximately 20% reduced glucose production from amorphous cellulose after 6 hours. The longer saccharification time was

necessary to achieve comparable total sugar production, along with sufficient reduction in viscosity to make HPLC analysis convenient. Taxifolin, α-pinene, and abietic acid had no statistically relevant inhibition.

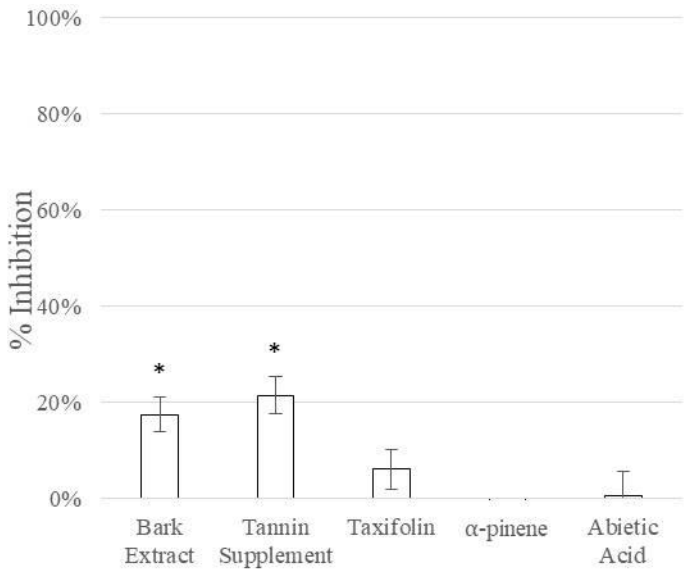


Figure 3-2: Inhibition of amorphous cellulose hydrolysis after 1 hour. A. Inhibition of the observed glucose due to the addition of extractives. B. Inhibition of the equivalent cellobiose due to addition of extractives. Error bars are standard deviation. * denotes >95% confidence that the values is different than experiments without extractives. Bark extract and tannin supplement both cause inhibition of the endoglucanases.

The inhibition of amorphous cellulose hydrolysis by bark extract closely resembles the inhibition seen from the tannins/phenols in the tannin supplement. Considering that none of the other extractives had inhibition comparable to the bark extract, this implies that the bark's own tannins and phenols lead to inhibition. Condensed tannins are known exoglucanase, endoglucanase, and β-glucosidase inhibitors[51, 69] and are the most abundant aqueous phenolic extractives in Douglas-fir bark[40]. Taxifolin and taxifolin derivatives are the second most

abundant aqueous phenolic, but taxifolin was not seen to cause inhibition. This means that tannins are the primary saccharification inhibitors in the bark extractives.

The data in Figs 3- 1A and 3-1B suggest that inhibition of β-glucosidase accounts for roughly half of the inhibition associated with converting crystalline cellulose into glucose. Even in a complex mix of enzymes, we can largely isolate the effects of bark extract and tannin on β-glucosidase by using cellobiose as the substrate. Figure 3-3 shows glucose production inhibition when cellobiose substrate, Celluclast enzyme cocktail and potential inhibitory molecules bark extract, tannin, and taxifolin are examined after 1 hour.

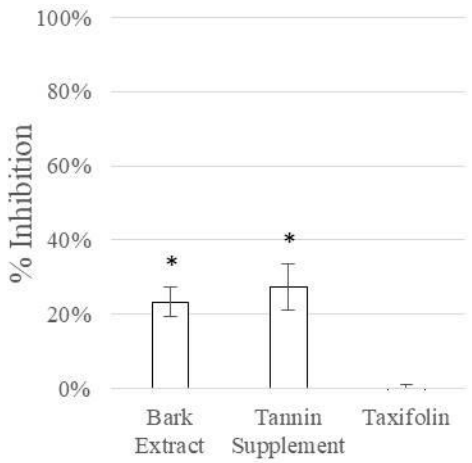


Figure 3-3: Inhibition of cellobiose hydrolysis after 1 hour. * denotes >95% confidence that the value is different than experiments without extractives. Glucose production was inhibited by both the bark extract and the tannin supplement by approximately 25%.

Experiments in Figure 3-3 were conducted at 7g/L cellobiose, a value that is similar to what is observed in solution at the 1 hour exoglucanase hydrolysis experiments. The glucose yield after one hour with cellobiose is a fairly direct measure of β-glucosidase activity, because the activities of the endo- and exo-glucanases towards cellobiose are small[156, 157]. The data show that

bark extract and tannin both caused approximately 25% inhibition of glucose production. Understanding the mechanistic origins of this inhibition of β-glucosidase is more straightforward than with the exo- and endo-cellulases, since the production of glucose from cellobiose (rather than cellulose) involves soluble small molecule reactants and products. Thus, mechanistic insights can be gleaned from Lineweaver-Burk plots.

3.3.2 β-Glucosidase Lineweaver-Burk Plots

Lineweaver-Burk plots can be used to more quantitatively examine β-glucosidase inhibition to gain mechanistic insights (Figure 3-4). Data for β-glucosidase activity as a function of inhibitor concentration and cellobiose concentration are presented with concentrations of starting cellobiose ranging from 4 g/L to 14 g/L. This covers the range of cellobiose values that could be observed during crystalline cellulose hydrolysis. A trend line fit with linear regression leads to a line of slope 1.92 ± 0.16 (standard deviation) and intercept of 165 ± 9 for the case without extractives. The fitted line for the 2.6g/L extractive solution had a slope of 1.95 ± 0.31 and intercept of 197 ± 18.

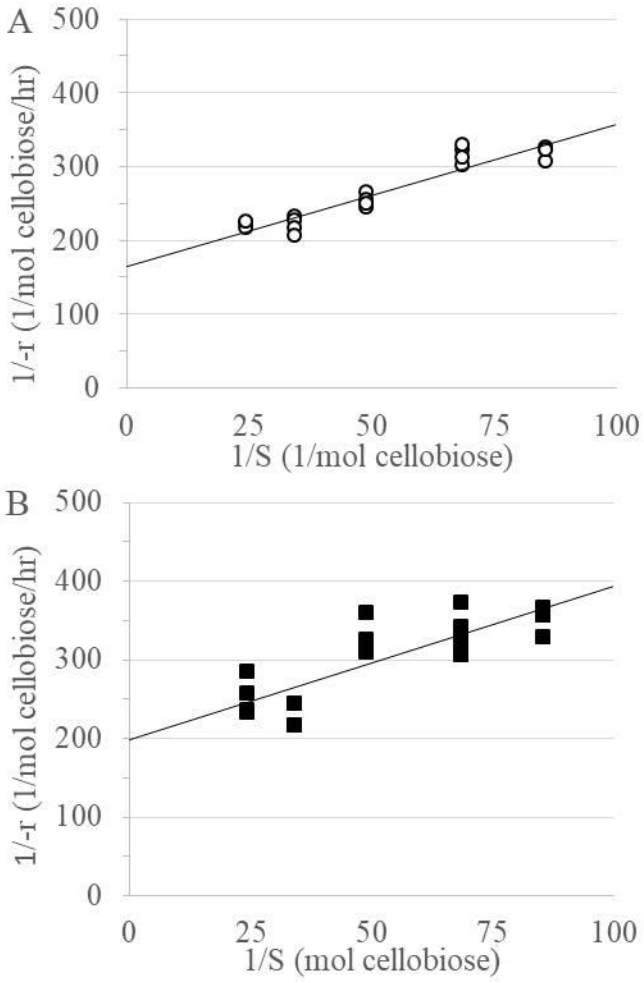


Figure 3-4: Lineweaver-Burk Graphs with individual data points displayed. A) Experiments without extractives. B) Experiments with 2.6 g/L extractives

The y-intercept for the full extract case is larger than the no-extractive case and this is statistically relevant based on a one-sided t-statistic. However, the slopes are not statistically different. When the slopes of Lineweaver-Burk plots are identical, but the intercepts differ, it implies that the enzyme (β-glucosidases) experience uncompetitive inhibition[158]. Uncompetitive inhibition was also found for tannin interactions with α-amylases[50].

3.4 Discussion

Tannins reported in literature for Douglas-fir share many structural features to that of the tannins in the tannin supplement. The tannins in the supplement are derived from grape seed and both contain procyanidins polymers of primarily (-)-epicatechin and (+)-catechin subunits[90, 159]. They differ in average tannin size however, with grape seed tannins reported to be made up of an average of 3.2 subunits[159] while the tannin from Douglas-fir inner bark has been reported to have an average of 7 subunits[89]. Regardless of the difference, they appear to inhibit cellulases a similar amount when their UV-Vis absorption peaks at 280nm are matched up. This shows that absorption at 280nm is clearly associated with inhibitors and could be developed into a simple and effective predictive tool for estimating solution inhibition. This would be much faster determination than carrying out a saccharification experiment for in-line testing at a biofuel facility.

With uncompetitive β-glucosidase inhibition, the tannins do not compete with the substrate for the catalytic site, but instead interacts in a manner that stabilizes the enzyme-substrate complex. This can be rationalized in part by considering that the inner bark tannins in Douglas-fir are much larger than the cellobiose or glucose molecules[89]. Tannins may simply be too large to travel into the catalytic sites and must affect the catalytic site indirectly. It would also mean that the tannin has minimal impact on the ability of the cellobiose to enter the catalytic site of the enzyme. Instead tannins would block the cellulase's ability to form and release glucose.

Getting at the question of whether inhibition is competitive, uncompetitive, or mixed is non-trivial for exo- and endo-glucanases in a complex cocktail of enzymes with a large macromolecule substrate such as cellulose. This limits us to mostly speculating on the inhibitory mechanism for exo- and endo-glucanases at this point, but it opens interesting questions that could likely be answered by molecular simulation or careful 2D NMR studies.While it is clear

we can relate the inhibition of glucose production to the tannin content of the bark extract, the mechanistic role tannins play in inhibiting the desired saccharification chemistry is less clear. It is noteworthy that the inhibitory effects of bark extract and tannin on the initial rate of glucose production from amorphous cellulose (Fig. 3-2) is comparable to the inhibition of "equivalent cellobiose" production from crystalline cellulose (Fig. 3-1B). Based on the two-step reaction scheme in Eq. 1, inhibition of "equivalent cellobiose" is driven by reduced activity of exoglucanase, whereas inhibition of the direct reaction scheme in Eq. 2 is associated with reduced activity in the endoglucanase. In short, it appears that extractive and tannin inhibition of endo- and exo-glucanse rates is comparable; this may provide some mechanistic insights into the origins of inhibition. In particular, exoglucanases utilize two functional domains, the carbohydrate binding module and the catalytic module. In contrast, the carbohydrate binding module for endoglucanases is not functional with amorphous cellulose substrates, so inhibition of endoglucanase must be associated with decreased activity of the catalytic module. Based on this and our observation that inhibition is comparable for both the endoglucanase and endoglucanase, it suggests that inhibition of the catalytic module may be the main driver for both enzymes. That said, there are distinct differences in the structure of the catalytic module for exoglucanase and endoglucanase, namely, exoglucanase has a catalytic tunnel whereas endoglucanase has a catalytic cleft. Given the different steric constraints for a tannin molecule in the tunnel vs. cleft and the different sizes of grape seed tannins vs Douglas-fir tannins, these results seem to imply that inhibitory binding does not occur directly at the catalytic site.

3.5 Conclusions

Douglas-fir bark extracts were shown to cause between 15-25% inhibition of *Trichoderma reesei* exoglucanases, endoglucanases, and β-glucosidases at biofuel relevant conditions. Tannins

are the molecules primarily responsible for the observed inhibition. Exo- and endo-glucanase inhibition seems to be mainly associated with the activity of the catalytic module rather than the affinity of the carbohydrate binding module to cellulose. β-glucosidase inhibition appears to be uncompetitive in nature based on Lineweaver-Burk plots. In addition, UV-VIS adsorption at 280nm, which is a commonly used to estimate phenol concentration in the wine industry, coincided well with the inhibitory influence of tannins in solution and could likely be the basis for monitoring saccharification inhibitors.

Tannins are the most potent extractive saccharification inhibitors. Cellulase hydrolysis can be improved by either removing or neutralizing the tannins, but extra process steps and chemical additives add to the cost of biofuel production. Bioengineering cellulases to be more resistant to tannins or using more tannin-resistant enzymes could be a way to reduce inhibition without extra process steps. Further research into understanding how tannins bind to the cellulases could lead to new insight on how to remove the inhibition or predict if different enzymes are more likely to be tannin resistant.

3.6 Methods

3.6.1 Materials

Bark was collected from an approximately 20-year-old *Pseudotsuga Menziesii* (var. *menziesii*) tree grown in Hoodsport WA. The bark was a mixture containing the outer bark, inner bark, and cambium collected from heights all along the trunk. The bark pieces with diameters and lengths in the range of 10^{-3} m to 10^{-4} m, a range of particle sizes observed in biofuel processing[160], and were stored under refrigeration. Citrate buffer, cycloheximide, tetracycline, Avicel PH-101 (~50 µm particle), low viscosity Carboxymethylcellulose sodium salt, D-(+)-

cellobiose (>98%), Abietic acid (~75%), α-pinene (98%), and cellulase from *Trichoderma reesei* (Celluclast, Cat. No. C2730) were purchased from Sigma-Aldrich. Taxifolin (>96%) was purchased from Abcam. Tannin used in this study was derived from grape seed extract that was highly phenolic (95%), with roughly two-thirds specifically associated with oligomeric proanthocyanidins. The source of tannin was Nature's way Tru-OPCs tables, crushed with a mortar and pestle as reported previous[69].

3.6.2 Extractive solutions and buffers

Bark was dried between 60-70 ˚C until the mass did not decrease by more than 0.1% in 10 minutes. DI water was added to 75g/L of the dried bark and was mixed in a shaker table at 200 rpm and 50^{o}C for 12 hours. The solids were removed by filtering through 0.45-micron PES filters, producing extract solutions that contained 2.7 ± 0.3 g/L dry extractives. An example UV-Vis spectrogram of the bark extract is shown in Figure 3-5 (0.1 mm quartz cuvette), where the peak at 280 nm had an absorbance of 0.25±0.05 (95% confidence interval), a wavelength commonly used to estimate phenolic concentrations[161]. Tannin solutions were formulated to have a comparable phenolic concentration with the bark extract through matching the UV-Vis absorption at 280nm. Tannin solutions were made by adding 3.1g/L of the grape seed tannin to DI water, mixing in a shaker table at 200 rpm and 50^{o}C for 12 hours, and then filtering out the solids. A spectrogram for the tannin solution is present in Supplementary Information. Solutions with taxifolin, abietic acid or α-pinene had the extractive added close to their solubility limit (1000 mg/L, 2.75mg/L, and 5mg/L respectively)[52, 162, 163], then mixed in the shaker table at 50^{o}C for 2, 12, or 2 hours respectively. Buffer chemicals were added to create 0.05M citrate buffers with 0.015% w/v cycloheximide and 0.02% w/v tetracycline.

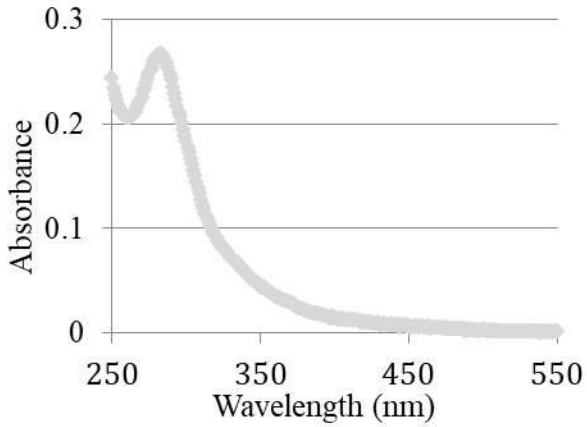


Figure 3-5: UV-Vis spectrogram of bark aqueous extract. Absorption in the range of 250-350nm is characteristic of many phenolic molecules, and 280nm is commonly used to estimate phenol concentration.

3.6.3 Initial inhibition saccharification experiments

Buffer solutions with or without extractives were preheated in the shaker table at the experimental conditions (50^{o}C and 200rpm) and added to substrates preheated in capped Erlenmeyer flasks. Cellulose substrate concentrations were either 250g/L Avicel PH-101 (crystalline) Cellulose or 250g/L Carboxymethyl sodium salt (amorphous cellulose). Cellobiose was used as a substrate at 7g/L to mimic values reached near the end of the crystalline cellulose experiments. Solutions combined with substrates were mixed in the shaker table for 1 hour before cellulase addition. Owing to the viscosity of the amorphous cellulose solutions, additional mixing was required, including a 30 second vortex mix upon buffer addition to the cellulose, stirring with a spatula upon cellulase addition, then vortexed for another 30 seconds after stirring. All concentrations used here were chosen to be representative of a bark rich feedstock undergoing biofuel processing (liquids to solids ratio of 4:1, bark to other solids ratio of 23:76)[40]. After the preheating and mixing steps, 60 ml/L *Trichoderma reesei* cellulase cocktail was added

to match previous studies[19]. Samples were taken at 1 hour for the crystalline or cellobiose experiments and after 6 hours for the (initially) viscous amorphous cellulose solution. Cellulases were denatured by placing the sample in boiling water bath for 20 minutes. The samples were then centrifuged and the supernatants analyzed via HPLC with an RSO-Oligosaccharide Ag+ (4%) column. More information on HPLC analysis can be found in the Supplementary Information. Triplicate or more experiments were used to evaluate errors in all reported data.

3.6.4 Extract adsorption on cellulose

Buffer solution containing extract and free of extract were prepared as described above. Crystalline cellulose was added at 250g/L and allowed to mix in a shaker table at 50^{o}C and 200rpm for one hour. The solutions were centrifuged to separate solid cellulose and the supernatant. The supernatant was discarded and fresh extractive-free buffer was added back into the residual cellulose solids. The solutions were then preheated, saccharified, and processed as described above to see if adsorption of extractives on cellulose was inhibitory.

3.6.5 Extract-β-glucosidase Lineweaver-Burk plots

Lineweaver-Burk plots for the β-glucosidase-extractive interactions were created by carrying out saccharification experiments as described above at 14, 10, 7, 5, and 4 g/L cellobiose. Bark extract solutions were used at a concentration of 2.6 g/L extractives.

3.7 Supplementary Information

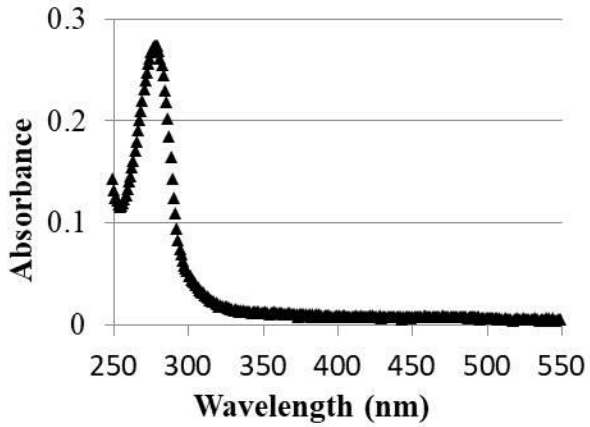


Figure 3-S1: Example Spectrogram for the tannin supplement solution.

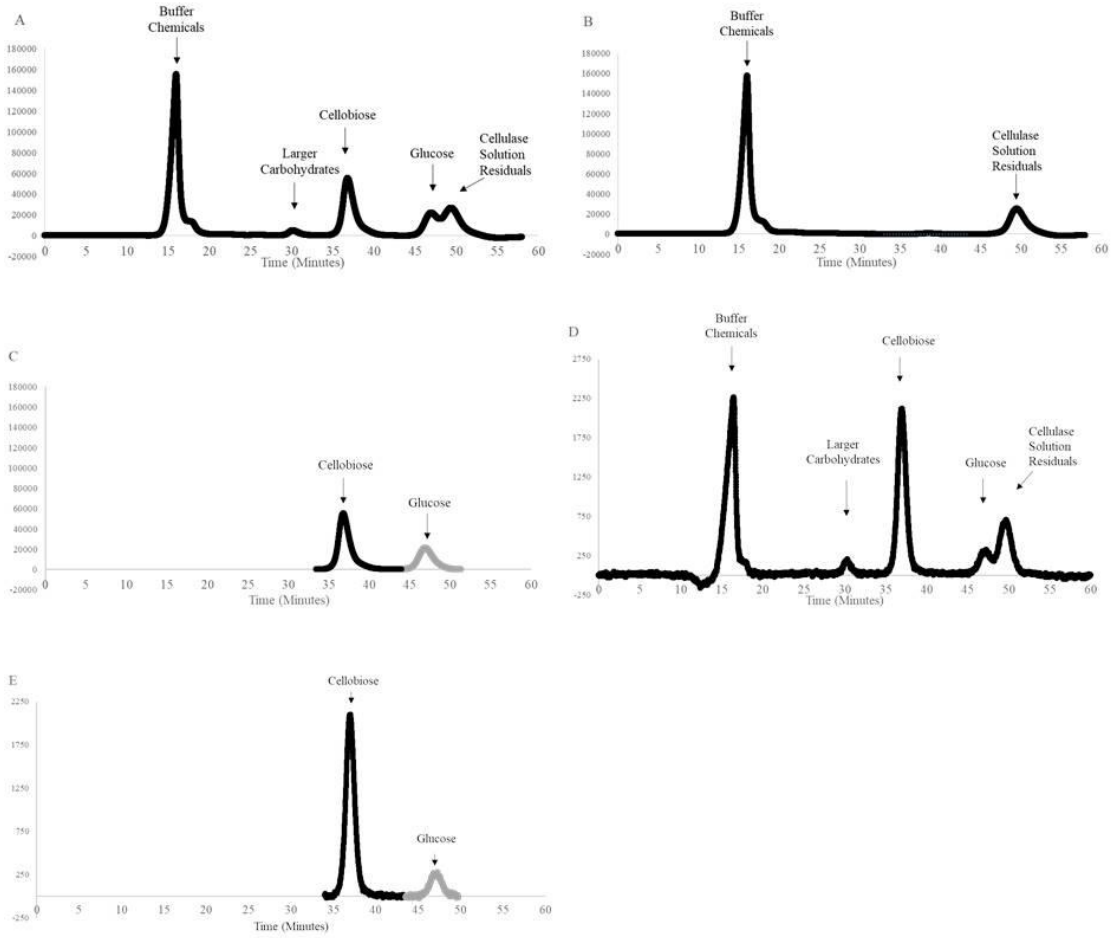


Figure 3-S2: Chromatograms from crystalline cellulose saccharification. A. Chromatogram of an experiment in which saccharification of crystalline cellulose occurred is present in Figure 3-S2A. B. Chromatogram of the buffer solution with remaining denatured enzyme solution but no substrate. C. Peaks integrated for carbohydrate determination, derived from subtracting the buffer and then subtracting the remaining baseline. D. Chromatogram from saccharification of 14g/L cellobiose. E. Resultant peaks integrated for sugar determination from 14 g/L cellobiose.

A Shimadzu Prominence UFLC System with a RID-10A refractive index detector was used for our HPLC analysis to determine sugar concentrations. An RSO-Oligosaccharide Ag+ (4%) column was used operating at 80°C with water as the mobile phase. Chromatograms were

collected for the saccharification experiment samples (Figure 3-S2A) as well as solutions of the buffer and denatured enzymes (Figure 3-S2B). The experimental chromatograms had the buffer chromatograms subtracted from them then the remaining baseline subtracted. An example of the resultant peaks is shown in Figure 3-S2C. The processed peaks were integrated and glucose and cellobiose standards were used to determine the glucose and cellobiose yields. Solutions with Avicel or Carboxymethyl cellulose were not found to contain cellobiose or glucose until cellulases were added. Cellobiose solutions were also not found to have glucose before cellulase addition. Bark extracts were seen to contain less than 1g/L of glucose or cellobiose.

In experiments in which a better fit of the cellulase solution residuals was desired (such as to get better accuracy in the Lineweaver-Burk experiments), the cellulase solution residuals peak for the individual experiments were modeled with a Chesler-Cram curve[164] and then subtracted from the experimental chromatogram. An example chromatogram for the cellobiose saccharification experiments is shown in figure 3-S2D. All parameters for the Chesler-Cram curve came from fitting the equation to the buffer-cellulase solution first, then changing the peak height parameter (A) to the cellulase solution residual peak of the experiment. An example chromatogram for a cellobiose saccharification is shown in Figure 3-S2D. Then the peak height was changed to the peak cellulase height of a given solution, and the resultant curve subtracted from the saccharification chromatogram. Figure 3-S2E shows the results of such processing.

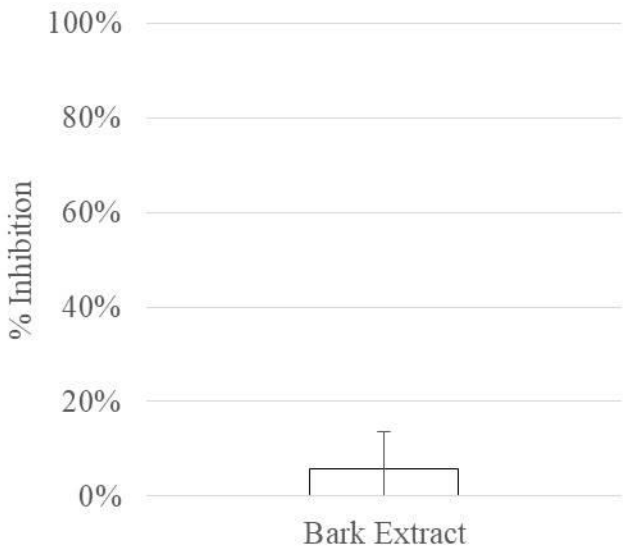


Figure 3-S3: Equivalent cellobiose derived from crystalline cellulose in which the extractive containing buffer was replaced by extractive-free buffer before saccharification.

Chapter 4: Inhibition of the Exoglucanase CEL7A by a Douglas-fir Tannin

4.1 Chapter Summary

Douglas-fir forestry residues have been a popular feedstock for biofuel studies and condensed tannins are expected to make up ~3% of the o.d. mass of this feedstock. Condensed tannins are well known for their ability to interact with proteins and can bind and inhibit cellulase enzymes used in biofuel processing. In this study, we use molecular dynamics simulations to investigate how characterized condensed tannin from Douglas-fir bark binds to the exoglucanase CEL7A from *Trichoderma reesei*. Our simulations indicate that tannin binding primarily occurs on the catalytic module as opposed to the carbohydrate binding module. The tannin can bind in a variety of different places/poses in/or near the openings of the catalytic tunnel and is large enough to interact with multiple parts of the cellulase surface at once. The analyzed tannin further prefers to bind to loops around the catalytic region and has affinity for aromatic and charged amino acid residues. These mechanistic insights could help inform rational design of tannin-resistant cellulases.

4.2 Introduction

Forestry residues are being investigated as a potential biofuel feedstock since they are an abundant biomass that does not require major land-use change or compete with food resources[2, 3, 5]. Douglas-fir residues have received special attention since Douglas-fir is the most abundant tree in the timberlands of the Western United States[29]. Biofuel production from this material has yet to hit full-scale production, but production processes are already reaching the pilot plant scale[10]. Better understanding of how the individual molecules in this complex feedstock interact with individual processing steps can lead to ways to improve these processes.

Many leading biofuel technologies rely upon a process called saccharification in which enzymes break down biomass into soluble sugars for further conversion into fuels and chemicals. Cellulases from *Trichoderma reesei* are the backbone enzymes in many saccharification cocktails and these enzymes have been shown to be inhibited by the tannins in Douglas-fir in the previous chapter. Tannins are water soluble polyphenols which are toxic towards fungi, bacteria, and yeast[45]. These polyphenols can inhibit extracellular microbial enzymes and are expected to make up about 3% of the dry weight of forestry residues(7.5-18% of the o.d. bark mass)[40]. Exoglucanase, endoglucanase, and glucosidase enzymes have been specifically reported to be inhibited by these Douglas-fir tannins. This inhibition could be removed through previously described methods such as the use of sacrificial proteins or surfactants[69], pulp washing[40, 165], tannin degrading microbes and enzymes[166], or steps to remove bark from the biomass. However, these techniques involve recurring process costs that drive up biofuel price and are as such undesirable. One route to remove the inhibition without recurring costs would be to have microorganisms express cellulases that are more resistant to tannins. This could be done by either incorporating cellulases from other organisms that are more tannin-resistant, such as incorporating β-glucosidase from *Aspergillus niger*[73] instead of from *Trichoderma reesei,* or bioengineering *Trichoderma reesei* cellulases to more effectively resist tannins. Bioengineering cellulase has found success in the past, including the creation of more thermostable cellulases, improved enzymatic binding to cellulose, and increased hydrolytic rate[167]. A better understanding of the mechanism(s) by which Douglas-fir tannins cause enzymatic inhibition is needed in order to find new and/or improved methods for designing tannin-resistant enzymes.

Enzymatic processes in the past have been built around the limitation of the enzymes. Molecular simulations are an increasingly important route to instead design the enzyme around

the process[168]. Molecular modelling methods have been used to investigate enzyme improvements such as tunnel mutations[169], surface charge modification[170], and enantioselectivity change[171]. Simulations have also been used to identify and study cellulase improvements[167] such as improved cellulose transport into and out of catalytic areas[172] and carbohydrate binding modules with stronger binding affinity[173]. In this study, we use molecular dynamics (MD) to investigate the means by which an experimentally characterized Douglas-fir tannin binds to the most abundantly produced cellulase in *Trichoderma reesei* cellulase cocktails, CEL7A[174]. This exoglucanase degrades the crystalline portions of cellulose starting from the reducing end of cellulose strands. It consists of both a carbohydrate binding domain and a catalytic module that includes a catalytic tunnel. The tannin investigated here is a polyphenol consisting of four epicatechin units and has previously been investigated in literature for commercial exploitation of tannins[90]. Tannin binding sites were suggested on CEL7A with the molecular docking algorithm used in Patchdock[175, 176] and these suggested sites were confirmed or rejected based on if the tannin could remain at the binding site for at least 50ns in a MD simulation. The binding sites were then analyzed to identify common types of amino acids and/or secondary structure motifs within cellulase that could be the focus of future mutagenesis studies aimed at improving cellulase resistance to tannins.

4.3 Methods

4.3.1 Tannin, Enzyme Structures and Binding Sites

The simulated tannin was identified and characterized in a study by Foo and Karchesy[90]. Gaussview[177] was used to create the complete simulated structure of the tannin and perform simple energy minimization of the starting structure. Force field parameters (Gromos96 G43a1)

for the tannin topology were derived through the Dundee PRODRG2 server web site version PRODRG2.5β[178]. Crystal structures of the carbohydrate binding domain (1CBH[179]) and catalytic module (1CEL[180]) of CEL7A from *Trichoderma reesei* were obtained from the protein data bank. Ligands, crystallographic waters, and modified residues were removed from the crystal structures before submitting them for binding site analysis to Patchdock and subsequent MD simulations.

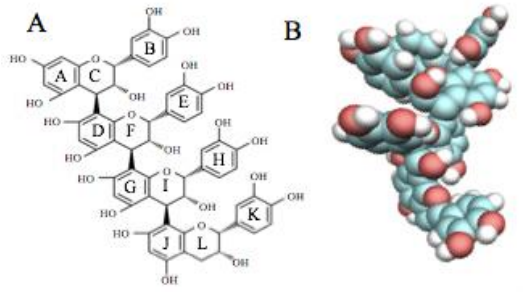


Figure 4-1: A) Structural formula of the simulated tannin and letter names of the individual rings. B) Space-filling model of the simulated tannin shown in Visual Molecular Dynamics (VMD[181]). Carbon, oxygen, and hydrogen atoms are shown in cyan, red, and white coloring, respectively.

Possible binding sites were first generated by Patchdock[175] and then validated with molecular dynamics simulations. Patchdock suggests possible binding sites by considering both geometric fit and atomic desolvation energy and finds binding sites on the rigid crystal structure at a reduced computational cost compared to starting with the tannin unbound in solution. The crystal structure was used in Patchdock so other studies can conveniently compare the binding rating of other molecules or tannin structure to the tannin structure in question. This was followed with molecular dynamics simulations to assess if the complex enzyme shape deviations over time

make the site untenable[182]. The first 20 structures from all complex type options were considered. Of these, only unique results in which the tannin did not overlap protein residues were used for simulations. This resulted in 10 unique tannin-enzyme starting structures for MD simulations with the catalytic module, and 14 for MD simulations with the carbohydrate biding module. Suggested tanning-binding sites 1-10 from Patchdock are sites on 1CEL with site 1 corresponding to the highest Patchdock-rated site and site 10 to the lowest rated site. Suggested sites 11-24 are on 1CBH and are similarly numbered according to how they were rated by Patchdock. The initial placements of the tannins around 1CEL and 1CBH are shown in Figure 4-2.

4.3.2 Molecular Dynamics Simulations Protocol

GROMACS4.6.5[183] was used for all MD simulations along with the SPC water model[184] and GROMOS96 force field (G43a1)[185] to model the enzyme. The GROMOS96 force field has been shown to produce binding results in good agreement with experimental data of other tannin-peptide and globular enzyme systems[41, 186-188]. Simulations of 1CEL had an average of ~23,300 water molecules while 1CBH had an average of 5,700. The catalytic module was neutralized by the addition of 22 Na^+ atoms, also modeled with the GROMOS96 force field. A steepest descent energy minimization of initial coordinates for each system was run for 5,000 steps. NVT equilibration was run for 2 ns at 298 K with the Bussi-Donadio-Parrinello thermostat[189]. This was followed by 2 ns of NPT equilibration at 298 K and 1 bar with the same thermostat and Berendsen barostat[190]. Production simulations in the NPT ensemble were then carried out at the same temperature and pressure with the same thermostat and the Parrinello-Rahman barostat[191]. Long-range electrostatic interactions were calculated with particle mesh Ewald summations (PME) using a cutoff of 1.3 nm. Lennard-Jones interactions were calculated over 1.2 nm and

shifted. The LINCS algorithm was used to constrain hydrogen bonds. Neighbor lists were updated every 10 steps with a cutoff of 1.3 nm, and full periodic boundary conditions (PBC) were used in all simulations. Production simulations were initially carried out for 10 ns with a time step of 2 fs. If the tannin did not diffuse away from the site during that time, simulations were extended for another 40 ns. Binding sites were confirmed as being relevant if the tannin molecule remained in the site for the entire 50 ns, a similar time to what has been used to study tannin-enzyme binding interactions previously[186]. Analysis of the simulation trajectories was performed over the last 40 ns. Root mean square deviation (RMSD) values of the backbone atoms of the enzyme from the crystal structure were calculated as a function of time with a GROMACS tool.

4.3.3 Binding Analysis Techniques

Enzyme-tannin interactions were analyzed by tracking the individual interactions between amino acid residues in enzyme binding sites and the individual eight outer aromatic rings (A, B, D, E G, H, J, and K) of the tannin molecule known to be important in binding[192-194]. Distances were calculated between the center of mass of specific amino acids and either a phenol ring with ortho-hydroxyl groups (B, E, H, K rings) or a phenol ring with meta-hydroxyl groups (A,D,G, J rings).

Calculations of both the frequency (i.e., per-residue occupancy) and stability (i.e., residence time) of tannin-enzyme interactions were carried out with in-house Python scripts. Trajectory data in the form of atomic coordinates was stored at 0.02 ns intervals during the simulations and was used to track amino acid residues in contact (i.e., within 4 Å) of the eight tannin rings. Changing the interaction length criteria between 3.9 - 4.1Å did not significantly change the results, and information on this is presented in Figure S4-4. Interactions were deemed relevant if a particular amino acid was in contact with a tannin ring for at

least 90% of the simulation (i.e., with an occupancy value of at least 90%). This value was chosen because it would roughly correspond to an energy difference greater than 2kT in order to change states based on a Boltzmann distribution of energies per equation 1. This would also roughly correspond to the values of a weak, 5 KJ/mol, hydrogen bond.

$$\frac{p_i}{p_f} = e^{(\varepsilon_j - \varepsilon_i)/kT}, \quad \text{Eq. 1}$$

where p is the probability of occupying a state, ε is the energy of a state, k is the Boltzmann constant, and T is temperature [195].

Residues were considered to be on the surface of the enzyme if they were within 5 Å of water molecules. Secondary structure analysis of the enzyme crystal structure was performed with the STRIDE[196] functionality in VMD.

4.4 Results and Discussion

4.4.1 Confirmed Binding Sites on 1CEL and 1CBH

MD simulations were conducted to identify tannin binding sites on the catalytic domain (1CEL) and carbohydrate binding module (1CBH) of exoglucanase CEL7A. Binding sites were first suggested through the molecular docking program Patchdock. These sites were then confirmed or rejected based on if an amino acid residues could remain within 4Å of a tannin ring for >90% of the molecular dynamics simulations, similar to previous enzyme – tannin research[197]. The starting structures from Patchdock suggested four possible regions for tannin binding: 1) in/near the exit of the catalytic tunnel of 1CEL [sites 1-5], 2) in/near the entrance to the catalytic tunnel of 1CEL [sites 6-10], 3) the hydrophilic side of 1CBH away from the connection to the linker region [sites 11-14], 4) the hydrophobic side of 1CBH [sites 15-24]. The structures are presented in Figure 4-2. Sites 1-5 were the highest rated by Patchdock and had scores between 6000-8000 while simulations 6-24 were in the range of 2000-6000.

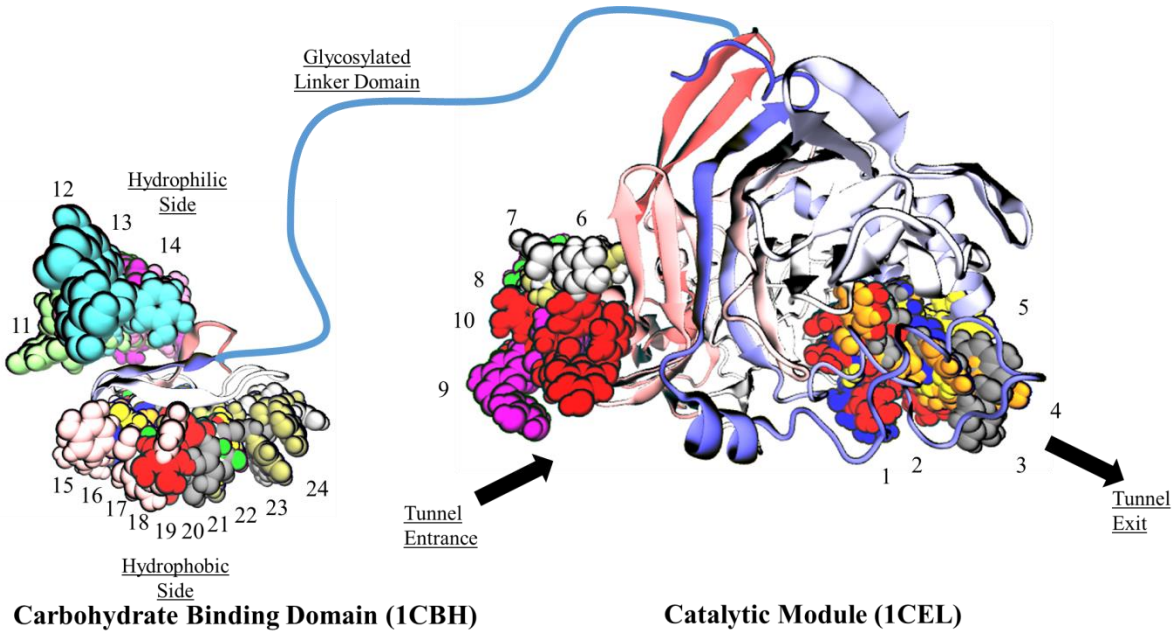


Figure 4-2: Suggested tanning-cellulase binding sites from Patchdock. The carbohydrate binding domain (1CBH) and catalytic module (1CEL) simulations were conducted separately. Each simulation contained a single tannin molecule and either 1CEL or 1CBH. Suggested binding sites are numbered 1-24 and the number appears next to the approximate location of the tannin on the cellulase. The linker domain was not simulated. Proteins are visualized in VMD and colored with the "index" option. Each tannin molecule is shown in a different color.

Binding sites were confirmed if one of the analyzed tannin rings remained within 4 Å of at least one amino acid residue for at least 90% of the simulation. The confirmed binding sites are shown in Figure 4-3. Eight of the 10 suggested sites for 1CEL were confirmed after 50 ns while only one of the 14 sites was confirmed for 1CBH. Sites were confirmed near/in the entrance and exit of the catalytic tunnel as well as on the hydrophobic side of the CBM. The nine unique tannin-cellulase combinations indicate that the tannin binding is not limited to a single site, but can bind to the cellulase in a variety of orientations at several different sites.

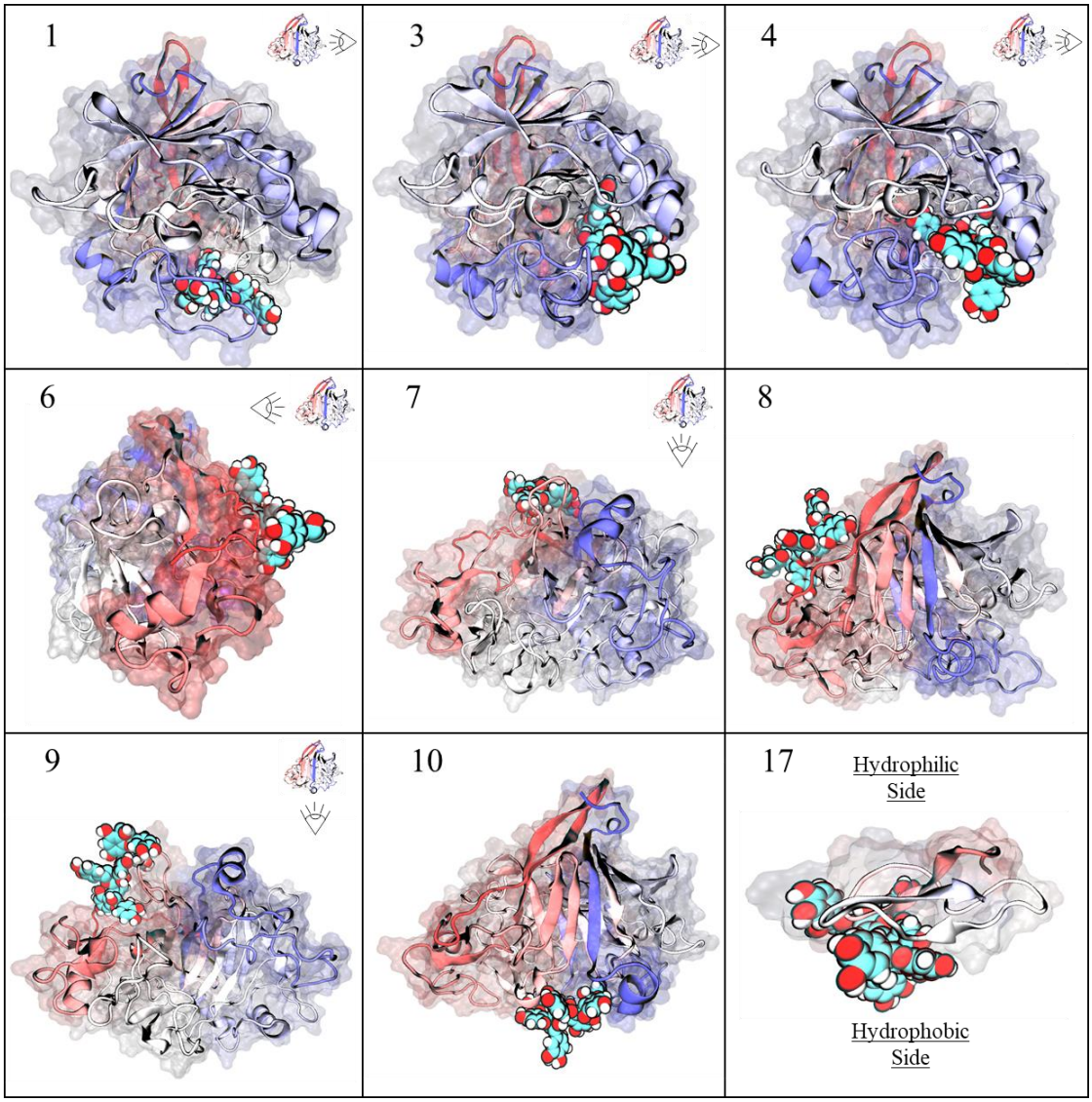


Figure 4-3: Confirmed binding sites for tannin on CEL7A (1CEL and 1CBH). Sites have at least one amino acid residue within 4 Å of a tannin ring for >90% of the observed time. 1CEL or 1CBH are displayed in a position to best show off the binding position. Multiple tannin sites were confirmed near the entrance and exit to the catalytic tunnel on 1CEL but only one site was confirmed on 1CBH on the hydrophobic side. Sites are labeled as in Figure -2, and coloring is as described in Figures 4-1 and 4-2.

4.4.2 Binding Site Characterization

Possible methods for designing a tannin-resistant enzyme can be found by identifying the specific amino acid residues and secondary structure motifs of the enzyme that are most amenable to tannin binding. The metric of percent occupancy – defined it as the fraction of frames over the analyzed MD trajectory for which a tracked tannin ring is within 4 Å of an enzyme residue[198, 199] – was used to determine the specific amino acids in each binding site. The number of residue-tannin ring interactions in the simulations are tabulated in Table 4-1. Specific tannin-residue occupancy values are presented in Table 4-S1 and 4-S2 of the Supplementary Information (SI). The results show binding sites had as few as 6 residue-tannin ring interactions and as many as 19. We note that tannins are large molecules that can interact with multiple residues simultaneously or even "sandwich" a residue between the different rings (see Figure 4-S1 in SI), hence some residues were counted more than once in Table 4-1. In addition, most of the interactions were not sustained over the entire length of the simulations, but involved the tannin molecules switching binding from one residue to the next in a particular binding site. Only five residue-tannin part interactions were observed to last throughout the entire length of the 40 ns simulations while 106 interactions were observed to last for >90% of the time.

Based on this information, it appears tannin binding is dependent on a network of interactions to form an overall stable complex. Cooperative binding of multiple phenolic tannin rings has been demonstrated for salivary proteins by Charlton et al.[200] Hagerman and Butler even showed that tannins can exhibit a thousand-fold lower affinity for alanine, glycine, and proline molecules individually compared to actual proteins[201, 202]. It is also statistically likely that tannin binding is initiated by a series of lesser strength interactions with stronger connections developing over time.

Binding Site	# with 100% Occu.	# with >95% Occu.	# with >90% Occu.
1	0	5	8
3	0	13	17
4	0	13	19
6	2	10	13
7	0	11	15
8	3	11	12
9	0	3	6
10	0	5	8
17	0	5	8

Table 4-1: Number of amino acid residue-tannin ring interactions involved in the binding sites identified by Patchdock and confirmed in MD simulations. The number of individual residue-tannin ring pairings that have 100%, >95%, and >90% occupancy are reported. In order to illustrate the trait of binding, each residue – tannin ring interaction is included even if a residue interacts with more than one tannin ring.

The number of times specific amino acid residue types in 1CBH and 1CEL were involved in binding tannin rings are shown in Table 4-2. If a residue had interactions with multiple tannin rings during a simulation, only the highest resulting occupancy was counted and reported. This was done in order to clearly assess what types of residues make up the binding sites and gives us

a value of 97 residue-tannin interactions as opposed to 106 residue – tannin "ring" interactions. Binding sites were seen to be comprised of combinations of aromatic, hydrophobic, polar, negative, and positive residues. The binding site of 1CBH was made primarily of hydrophobic residues, with six out of eight involved residues falling into that category. The binding sites of 1CEL were most commonly made up of polar (29 out of 89 residues) or hydrophobic residues (26 out of 89 residues). These findings imply tannin's multifaceted structure and diverse chemical functionalities allow it to interact with the enzyme's surface in multiple ways and adapt to find multiple binding sites.

In previous studies, Hagerman and Butler have reported that tannins have specific affinity for proline[201] while Adamczyk et al. reported they have affinity for arginine[202] residues. While we do see those two types of residues frequently involved in the binding sites, their limited availability on the surface indicates they are not solely responsible for binding. There are 19 proline residues out of the 434 total residues in 1CEL and only six proline residues are located in or near the catalytic tunnel of 1CEL. Four of these prolines were involved in binding sites however, with some of these specific residues part of more than one of the confirmed binding sites. In addition, there are also only seven arginine residues in all of 1CEL and three of them were found to be involved in binding sites. Overall, glycine and threonine residues were the most commonly observed residues in Cel7A to bind tannin rings. However, they were much more prevalent on the enzyme surface (7 out of 50 GLY on the surface involved and 9 out of 37 THR on the surface). Visual analysis of the simulation trajectories with VMD shows threonine can hydrogen bond with the tannin either through the carbonyl group on its backbone or the hydroxyl group on its side chain, and examples of each are present in Figure 4-S2 of SI. Glycines can hydrogen bond with the tannin through the hydroxyl group on its backbone.

		1CEL			1CBH		
Amino Acid Type	Residue Species	100% Occu.	>95% Occu.	>90% Occu.	100% Occu.	>95% Occu.	>90% Occu.
Aromatic	Phe	0	4	4	0	0	0
	Trp	0	3	4	0	0	0
	Tyr	2	5	6	0	1	1
	His	0	1	2	0	0	0
	Total	**2**	**13**	**16**	**0**	**1**	**1**
Hydrophobic	Ala	0	5	6	0	0	0
	Leu	0	2	2	0	0	0
	Val	0	4	4	0	0	0
	Ile	0	0	0	0	0	1
	Gly	0	4	9	0	3	4
	Pro	0	4	5	0	1	1
	Total	**0**	**19**	**26**	**0**	**4**	**6**
Negative	Asp	0	5	6	0	0	0
	Glu	0	5	7	0	0	0
	Total	**0**	**10**	**13**	**0**	**0**	**0**
Polar	Asn	0	1	1	0	0	0
	Cys	0	3	4	0	0	1
	Gln	0	4	4	0	0	0
	Met	0	1	1	0	0	0
	Ser	0	3	7	0	0	0
	Thr	1	8	12	0	0	0
	Total	**1**	**20**	**29**	**0**	**0**	**1**
Positive	Arg	1	2	3	0	0	0
	Lys	1	1	2	0	0	0
	Total	**2**	**3**	**5**	**0**	**0**	**0**

Table 4-2: Number of occurrences of an amino acid being involved at a binding site on 1CEL or 1CBH.Unlike Table 4-1, only one interaction between a specific residue and tannin per simulation are counted towards these values

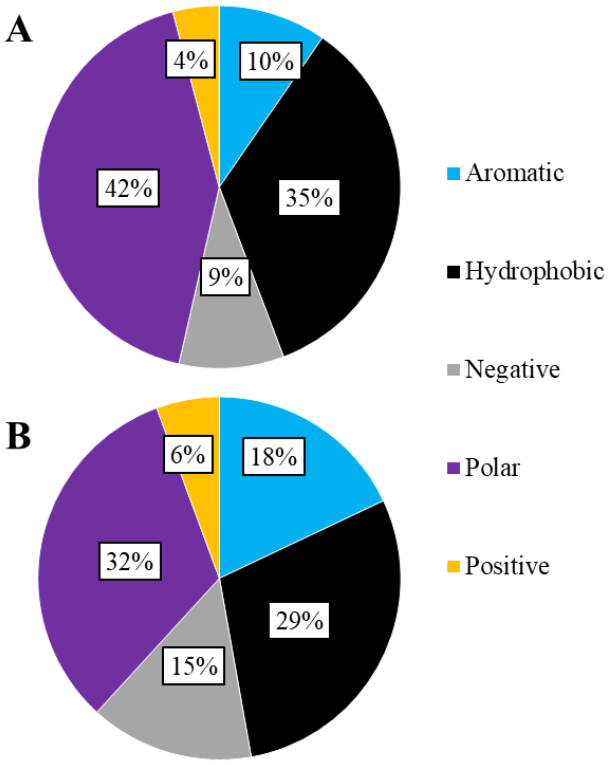


Figure 4-4: Amino acid proportions from the 1CEL simulations. A) Proportions of amino acid types on the surface of the catalytic module. B) Proportions of amino acid types involved in binding sites (>90% Occu.).

The condensed tannin analyzed here was capable of interacting with either the side chain of an amino acid or hydrogen bonding with the carbonyl groups on the enzyme's backbone. These interactions give the tannin an ability to interact with every type and species of amino acid residues to some extent, or at least not be repelled by specific ones. This does not mean that the tannin binds with equal strength to all residue types, however. For example, if the tannins bound equally strongly to all amino acid residues, we would expect the proportions of residue types involved in the binding sites to closely match the proportions of amino acid residues found on

the surface. Figure 4-4 contains pie charts of the proportion of surface residues types on 1CEL and residue types seen in the binding sites suggested by Patchdock and confirmed with MD. Charts for 1CBH are present in Figure 4-S5 of SI. These results show the tannin has stronger affinity for the negative, positive, and aromatic residues on 1CEL than expected based on the enzyme surface average. The affinity for both the negative and positive residues, as opposed to just one, occurs because the tannin ring's hydroxyl groups have a hydrogen atom and a lone pair that can be oriented differently to optimally align for either amino acid type. Affinity for aromatic residues is understandable since tannins are rich in aromatic rings that are capable of ring-ring stacking (sometimes referred to as pi-pi stacking). Stacking between the phenols and the non-aromatic proline ring has also been reported[203, 204], and examples of these are shown in Figure 4-6.

Tannin binds to the hydrophobic side of 1CBH, and correspondingly 6 of 8 residues involved were in the hydrophobic category (4 were GLY). This is a higher proportion of hydrophobic residues than the 14 hydrophobic residues out of 36 total exposed to water. However, it does not as clearly illustrate tannin trends since the information is based on only one site, and 1CBH contains no negatively or positively charged amino acid residues.

In a similar manner, tannin preference for secondary structure elements can be determined. Figure 4-S6 in the supplementary information shows the proportions of surface motifs on 1CEL along with the proportions of secondary structure motifs in the binding sites. Charts for 1CBH are present in Figure 4-S7 of SI. These results show binding sites consist of more flexible loops than would be expected based on surface residue averages. Tannins are well known for their ability to complex with peptide chains and other polymers[41, 186] and the tannin in this study correspondingly prefers the areas that are more peptide-like. A similar preference for

loops on the globular protein trypsin was found by Gonçalves et. al[187]. This could be due to the loops' flexibility allowing them to find optimal amino acid-tannin interactions. 1CBH was seen to have 4 out of its 8 binding residues in the loop regions whereas the enzyme has 20 out of 36 total residues in loops.

The number of interactions between individual tannin rings and enzyme residues in each secondary structure motif was also tracked. However, a single static structure was used for docking sites, while the tannin structure was observed to be quite flexible during the simulation that may lead to biases towards specific rings. This information is presented in Figure 4-S8 in the SI nonetheless.

The loops that the tannin interacts with on 1CEL are primarily the ones that make up the sides of the catalytic tunnel. These loops are immensely important for interacting with cellulose at the tunnel entrance and threading the cellulose through the tunnel[205-207]. Tannin binding to or near these loops could cause inhibition by obstructing cellulose access to the catalytic tunnel, affecting the ability of the cellulase to move cellulose along tunnel subsites, or affecting the actual cellobiose production process. These simulations focus on the native folded structure and RMSD calculations of the 1CEL and 1CBH backbone were used to rule out denaturation of the enzyme at these time scales. Denaturation occurring on much longer time scales cannot be ruled out however. This information is presented in Figure 4-S3 in the SI and indicates either competitive or mixed inhibition as a possible inhibition mechanism.

4.4.3 Stability of Confirmed Binding Sites and Most Stable Interactions

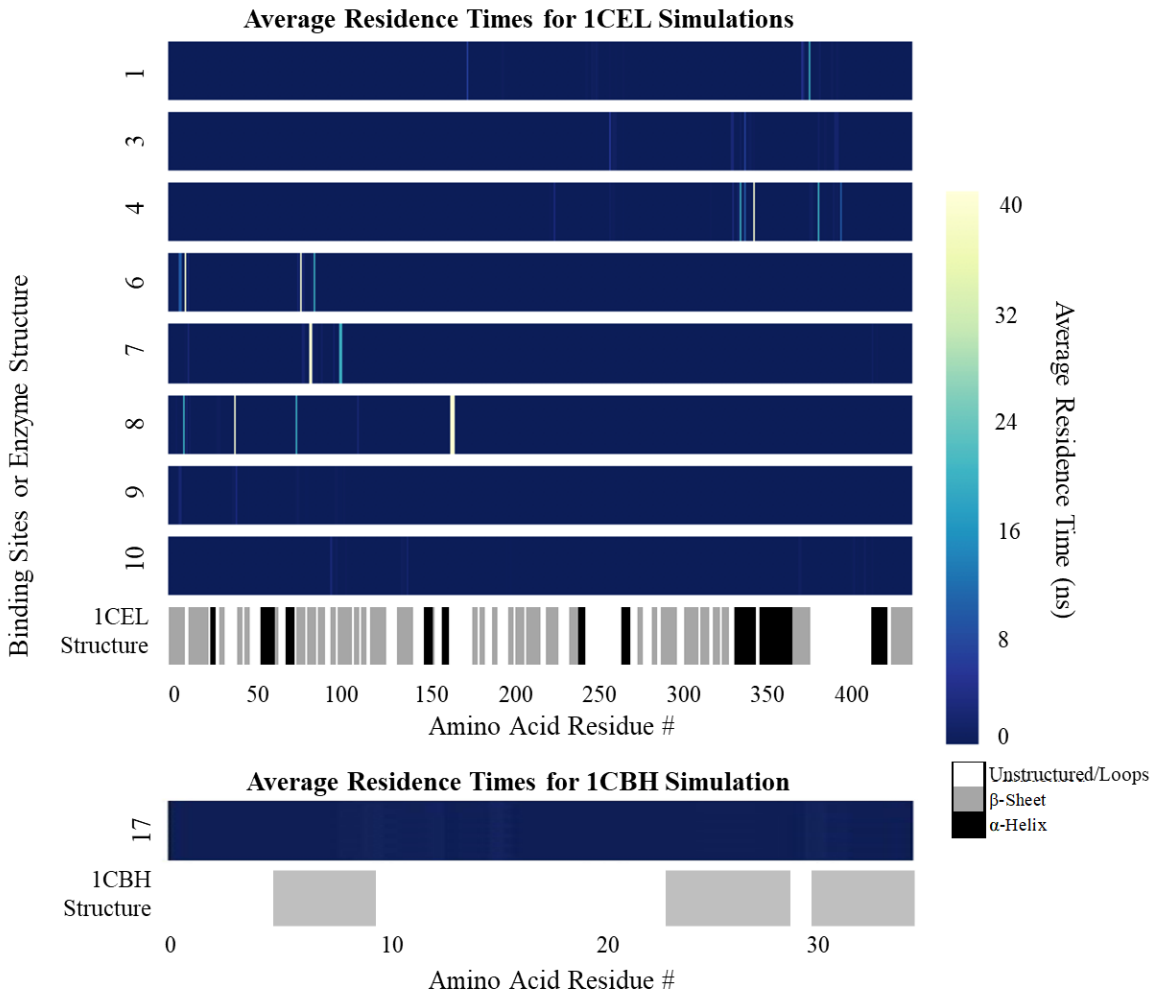


Figure 4-5: Residency information for each confirmed catalytic module simulation. Lighter colors imply stronger binding. Secondary structure information is included at the bottom of the simulations for 1CEL or 1CBH. This information came from the last 40ns of the simulations, so values approaching 40ns imply long term binding interactions. If a residue interacted with multiple rings of the tannin, only the longest average length interaction is reported. Simulations 4, 6, 7, and 8 on 1CEL involve the longest lasting interactions. Amino acid residues on 1CBH were observed to have interactions with tannin rings lasting < 2 ns.

An effective strategy for bioengineering a tannin-resistant cellulase might be to remove or weaken the strongest tannin-enzyme binding sites. The strongest binding sites result in interactions that last the longest in our simulation. Percent occupancy is not necessarily the best measure for this, since steric constraints can cause easily broken enzyme-tannin interaction to be revisited with high frequency[199]. We use residence time, defined as the average time a residue interacts with a tannin ring, as a measure of interaction stability in our simulations. Residence times information for residues in the binding sites are presented in Figure 4-5 as a heat map and actual residence time values are presented in Table 4-S3 and 4-S4 in SI. Longer interactions (lighter colors in Figure 4-5) imply there is a larger energy barrier for the tannin to reach a new state and thus the binding state is stronger. Illustrations of the longest interactions are presented in Figure 4-6 for the four binding sites with the longest residence time interactions. Similar illustrations for the other 1CEL binding sites are present in Figure 4-S9 in SI. From this figure, it can be seen that simulations 4, 6, 7, and 8 have the longest lasting binding interactions (residence times $\geq$ 39 ns) and lead us to believe these are the most important binding sites. These binding sites are in the exit end of the catalytic tunnel or interacting with the loops on the outside of the catalytic tunnel. It can also be seen from Figure 4-5 that the 1CBH binding site (binding site 17) is comprised of several weaker interactions of the order of 1 ns. This indicates that the binding site on 1CBH is expected to be much weaker than binding to 1CEL.

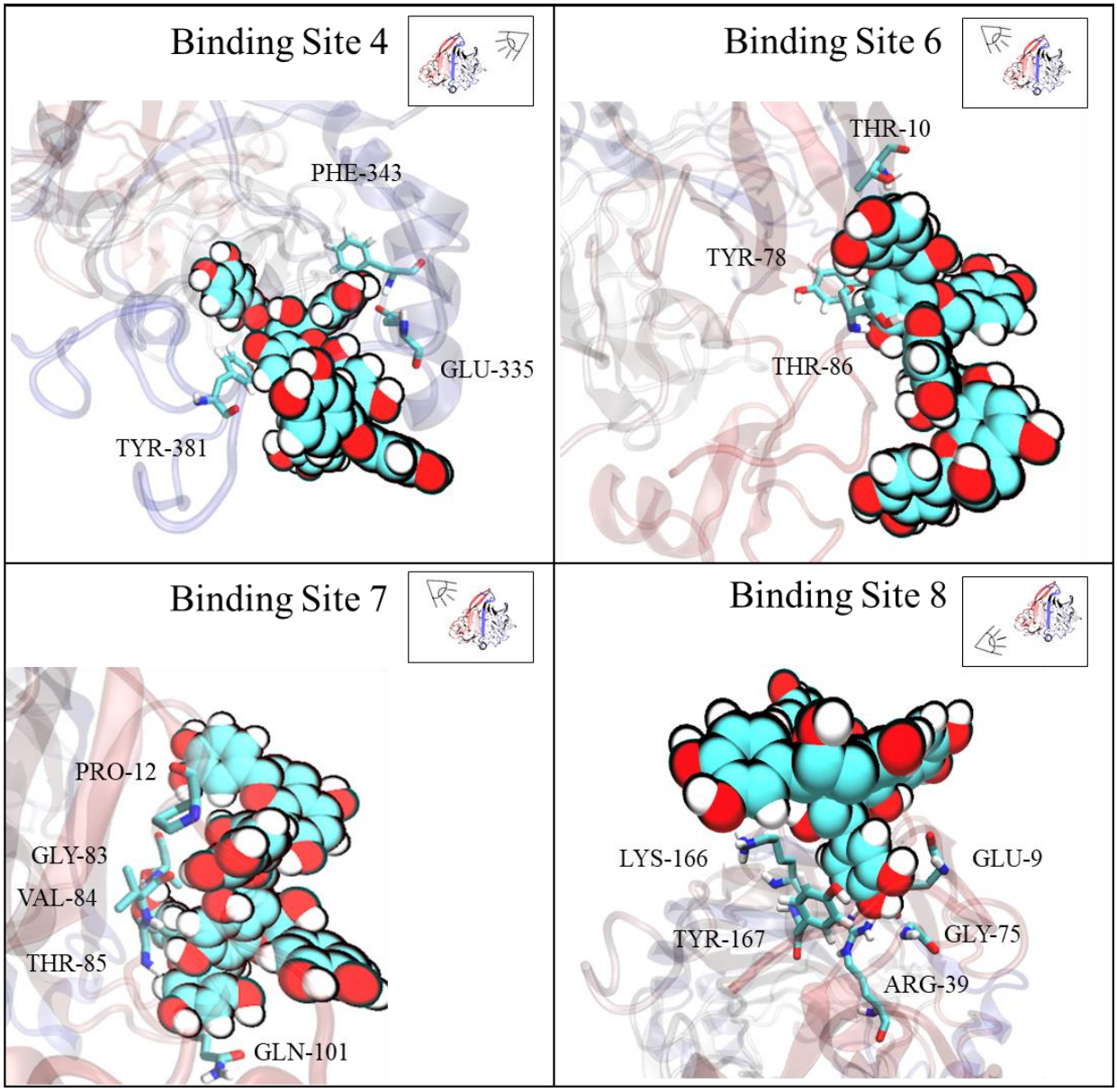


Figure 4-6: Illustrations of most stable binding interactions seen in Figure 4-5. Ring-ring interactions are illustrated in all simulations (TYR-381, PHE-343, TYR-78, PRO-12, and TYR-167). Hydrogen bonding is illustrated in Simulation 6 and 7 (THR-86 and VAL-84). Charged residue interactions with tannin aromatic rings are illustrated in Simulation 4 and 8 (GLU-335, GLU-9 and ARG-31).

The interactions that correspond to the brightest bands in Figure 4-5 are shown in Figure 4-6. It appears the most stable interaction in binding site 4 comes from the aromatic side chain of PHE-343 perpendicularly aromatic stacking with ring H of the tannin. Ring H also overlays GLU-335 in such a way that the non-charged parts of the side chain interact with the center of the tannin ring and the negatively charged portions are close to the hydrogens on the ring. On the other side of the tannin there is also binding by alternating offset-perpendicular aromatic interaction between TYR-381 and Rings J, K, and G. Binding site 6's longest lasting interactions involve hydrogen bonds alternating between THR-10 and rings B, D, and G. There is also an interaction that most closely resembles parallel-displaced aromatic stacking between TYR-78 and Ring D. Binding site 7's most stable interactions involve the H ring on/off hydrogen bonding with the backbone carbonyl of VAL-84 and offset-perpendicular ring interactions with PRO-12. It is expected that the cooperation of these residues and other hydrogen bonding around VAL-84 leads to large residence times as opposed to the VAL-84 interaction being unexpectedly strong. There is also on/off hydrogen bonding between the A ring and multiple carbonyl groups that involve the backbone of GLN-101 and surrounding residues. Binding site 8 has the tannin anchored with the H ring sandwiched between GLU-9 and parallel aromatic stacking with TYR-167. The charged head group of GLU-9 also orients itself away from the pi orbitals of the tannin ring and remain closest to the hydrogens on the ring. In addition, the side group of ARG-31 is between TYR-167 and GLU-9 with the guanidine group pointed towards the H ring's hydroxyl groups.

Based on this information, it appears that hydrogen bonding networks, ring-ring interactions, and ring-charged group interactions can all contribute significantly to a binding site. Hydrogen bonding displays an on/off nature with even the most stable hydrogen bonding having

the correct hydrogen alignment for no longer than a nanosecond. This implies that hydrogen bonding is significant, but is likely a weaker contributor to binding. Interestingly, amino acid residues involved in the hydrogen bonding interaction can still remain close to the tannin for long periods of time. These long lasting residues that only interact through hydrogen bonding probably benefit substantially from cooperative binding effects. Ring-ring and ring-charged group interactions typically lasted much longer than the hydrogen bonding's on/off switching, but the interactions do not always align in textbook fashion due to geometrical constraints. Considering all this, the loops of the exoglucanase may be the major target of tannins for two reasons: 1) their flexibility can allow for better alignment with amino acid residue rings and charged groups and 2) backbone carbonyl groups are less likely to already be occupied in hydrogen bonding with other protein backbone elements.

4.5 Conclusion

Based on these results tannin binding onto CEL7A is expected to occur at a variety of sites primarily on the catalytic domain of the exoglucanase. The strongest binding sites were shown to be in the exit of the catalytic tunnel and on the outside of loops of the catalytic tunnel. Sites were seen to consist of a minimum of 6 amino acid-tannin interactions and as many as 19. These multi-residue interactions cooperate to stabilize the binding structure. Residues in the binding sites were most commonly hydrophobic or polar types. The tannin was also shown to interact with residue side chains that are aromatic, positively charged, or negatively charged and preferred these residues more than would be expected based on the proportion of these residues on the enzyme surface. However, hydrogen bonding to the enzyme's backbone carbonyl groups gives the tannin an ability to interact with every type of amino acid residue, even if the amino acid side chain is not conductive to binding. A variety of ring-ring interactions were observed

with amino acid side groups and these interactions were more stable than the hydrogen bonds. There was also stronger preference for the unstructured and loop regions than α-helix or β-sheet than would be expected based on the makeup of the proportion of secondary structure facets on the surface. Binding sites were primarily comprised of loops around the catalytic tunnel where the enzyme has some flexibility to find optimal alignments between the amino acids and tannin aromatic rings. Inhibition thus appears to result primarily from either clogging the catalytic tunnel or interfering with the loops responsible for guiding the cellulose onto catalytic residues. This would imply a competitive or mixed inhibition mechanism. The glycosylated linker domain, with its prevalence of rings, hydroxyls, and high flexibility, could also be a binding partner for tannins and tannin binding to this region would be interesting future work.

Mutagenesis work could investigate making CEL7A tannin-resistance through modifying the types of residues found on the enzyme surface or possibly altering the secondary structure in a way that does not affect enzyme stability or function, or. Removal or modification of aromatic and charged residues could be another route; however this may be difficult for the charged residues due to their impact on enzyme solubility. Removal of any aromatic residue on the surface of the cellulase that is not directly involved in interactions with the cellulose could be a route to weaken binding sites on the tannin. Creating loops that are more highly structured along the catalytic tunnel could reduce the ability of these areas to optimally align with the tannin rings.

4.6 Supplementary Information

4.6.1 Full List of Occupancy Values

		Simulation #							
Residue #	Residue Type	1	3	4	6	7	8	9	10
5	THR						97.2%		
7	GLN				99.9%			98.6%	
8	SER						91.1%		
9	GLU						99.95%		
10	THR				100.0%	95.5%			
11	HIS				94.7%		98.9%		
12	PRO				99.0%	99.2%			
29	THR						95.3%		
38	TRP							95.5%	
39	ARG						100.0%		
40	TRP							99.5%	
74	ASP						96.0%		
75	GLY						99.95%	93.1%	
76	ALA				98.0%				
77	ALA				95.7%				
78	TYR				100.0%				
79	ALA				98.9%	99.0%			
80	SER					94.8%			
81	THR							90.1%	
83	GLY					99.95%			
84	VAL				97.4%	99.95%			
85	THR				95.0%	98.8%			
86	THR				99.95%	91.0%			
87	SER					94.1%			
90	SER					97.8%			
95	PHE								96.1%
96	VAL								96.2%
97	THR					98.5%			
99	SER							90.7%	
101	GLN					99.9%			
104	VAL								94.3%
111	MET						99.3%		
137	PRO								95.9%
140	LEU								99.0%
166	LYS						100.0%		
167	TYR						100.0%		
175	GLN	99.7%							
224	ALA			94.1%					
226	THR			99.2%					
245	GLY	94.5%							

246	THR	94.9%							
248	SER	97.3%							
249	ASP	95.2%							
251	ARG	92.7%							
259	ASP		99.5%	97.6%					
260	GLY		92.6%	90.6%					
261	CYS		98.0%	95.0%					
262	ASP		97.7%						
330	TYR		99.2%						
331	CYS		99.3%	99.2%					
334	GLU			97.8%					
335	GLU		98.0%	99.9%					
337	GLU			92.8%					
338	PHE		99.7%	99.8%					
340	GLY		95.3%						
341	SER		96.6%						
343	PHE			99.95%					
371	TYR	99.6%							
376	TRP	99.95%		93.1%					
377	LEU			97.3%					
381	TYR		94.0%	99.9%					
382	PRO			91.1%					
385	GLU		93.9%						
390	PRO			98.1%					
391	GLY		99.0%						
392	ALA		98.8%						
394	ARG			99.8%					
402	GLY								93.7%
404	ASP								90.4%
408	GLU								97.2%
413	ASN					97.8%			
415	LYS					93.3%			

Table 4-S1: Occupancy values for residues on the confirmed binding sites of 1CEL. Only values >90% occupancy are reported. If a residue interacts with two different tannin rings, only the largest occupancy value is reported.

Residue #	Residue Type	Simulation #17
8	CYS	93.1%
9	GLY	96.9%
10	GLY	96.4%
11	ILE	93.4%
12	GLY	96.8%
15	GLY	94.3%
30	PRO	97.9%
31	TYR	94.2%

Table 4-S2: Occupancy values for residues on the confirmed binding site of 1CBH. Only values >90% occupancy are reported. If a residue interacted with multiple tannin rings, only the largest occupancy value is reported.

4.6.2. Example Interactions

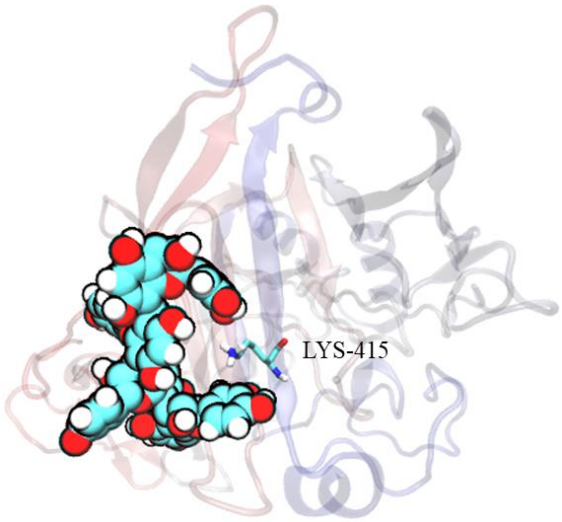


Figure 4-S1: Scene from the simulation of Binding Site 7. LYS-415 sandwiched between rings of the tannin molecule.

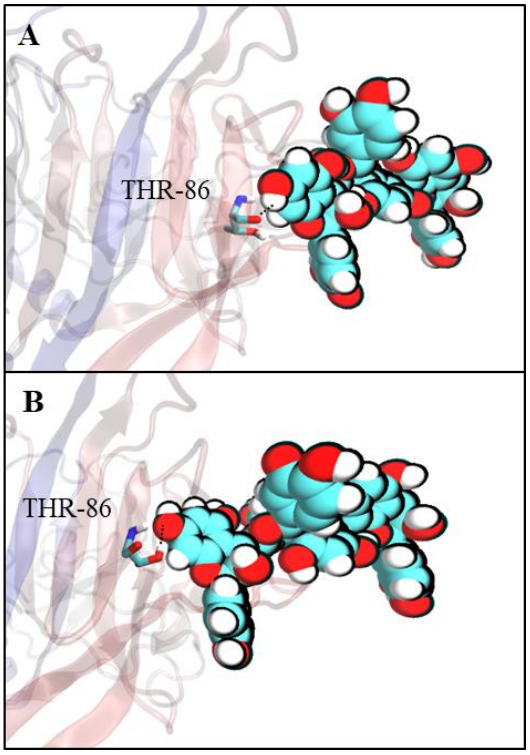


Figure 4-S2: Tannin hydrogen bonding THR-86 at different times in binding site #6. A) Hydrogen bonding to the carbonyl group of the backbone. B) Hydrogen bonding with the hydroxyl group of the side chain.

4.6.3. RMSD Information

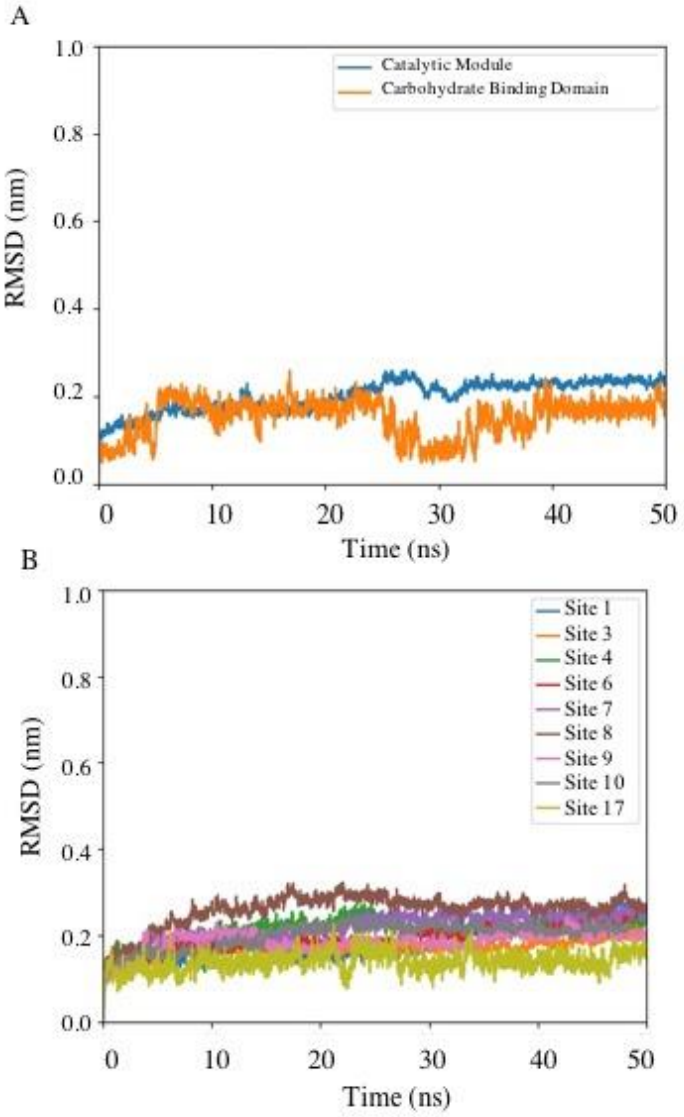


Figure 4-S3: RMSD values for the enzyme backbones referenced to the start of the production simulations. A) Values for the cellulase parts without any tannin. B) Values for the cellulase in the presence of the tannins.

To investigate whether tannin binding leads to denaturation of the protein, RMSD calculations of the Cel7A protein backbones are presented in Figure 4-S3. No overarching structural change to the enzyme counterparts are observed. This indicates that any slight rearrangement that may be observed in the final binding structures of Figure 4-3 occur primarily in regions that naturally have a larger range of motion, such as side chain elements and loops on the enzyme.

4.6.4. Occupancy Proportions

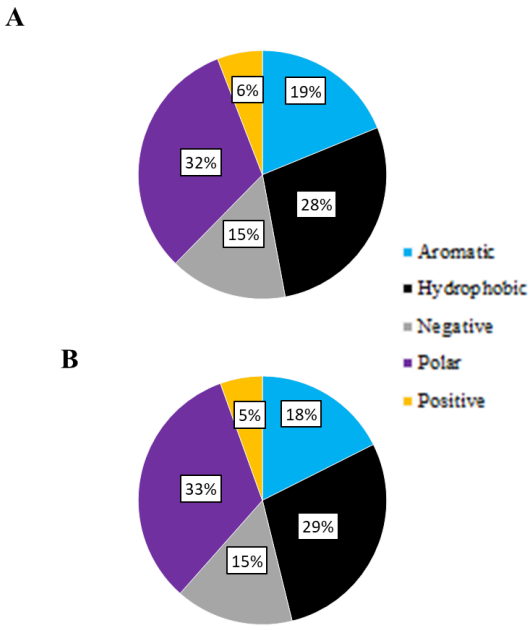


Figure 4-S4: Amino Acid Proportions in 1CEL binding sites for different interaction lengths. A) Proportion of interactions at 3.9Å distance. B) Proportion of interactions at 4.1Å distance.

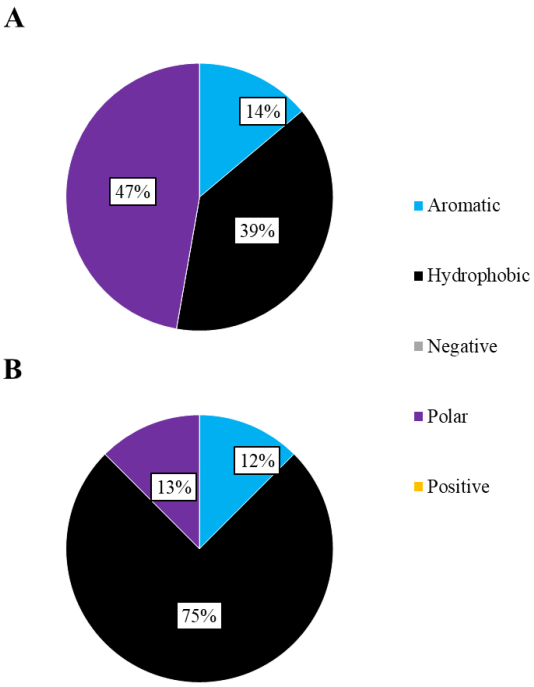


Figure 4-S5: Amino Acid Proportions from the 1CBH simulations. A) Amino acid proportions exposed to water. B) Amino acid proportions constituting the binding site.

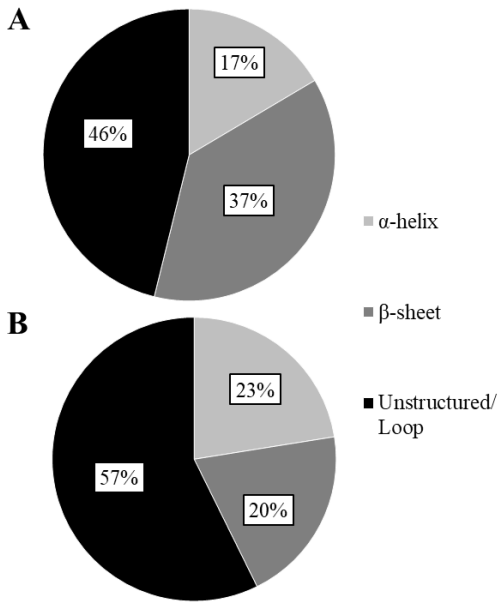


Figure 4-S6: Secondary structure proportions from the 1CEL simulations. A) Proportion of secondary structures on the surface of the catalytic module. B) Proportion of secondary structures involved in binding sites (>90% occupancy).

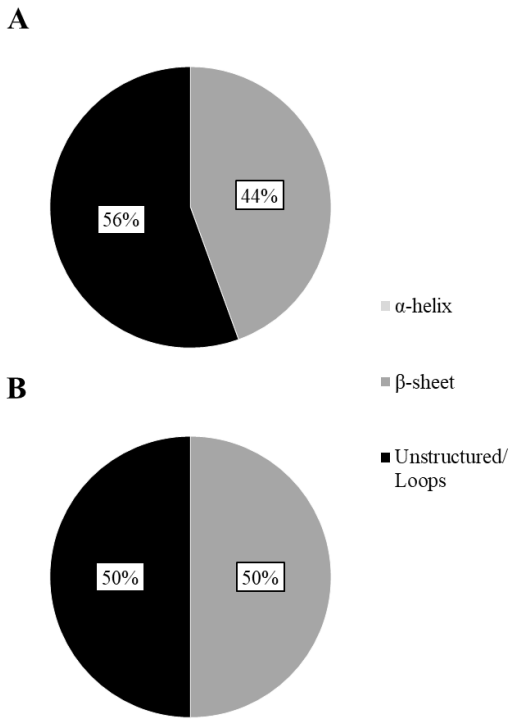


Figure 4-S7: Secondary structure proportions from the 1CBH simulations. A) Proportion of secondary structures on the surface of the carbohydrate binding domain. B) Proportion of secondary structures involved in binding sites (>90% occupancy).

4.6.5. Occupancy Information Specific to Each Analyzed Ring

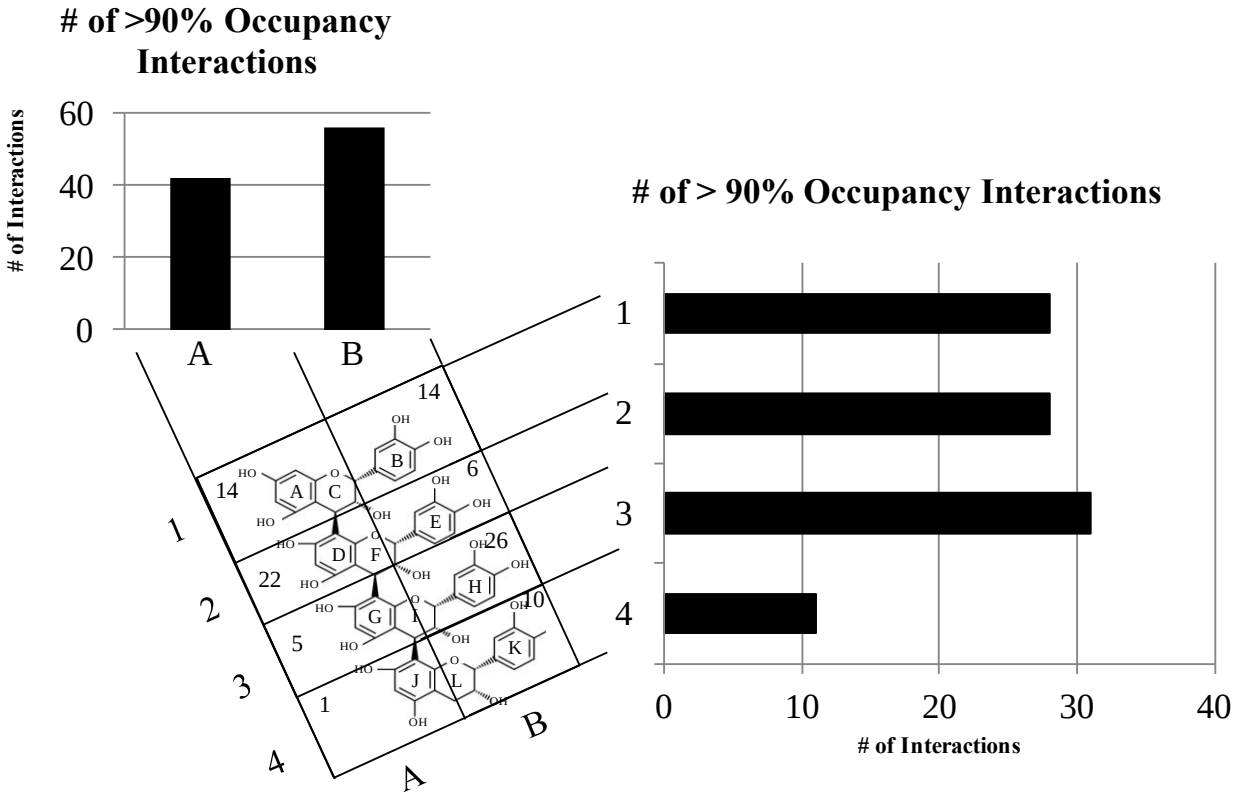


Figure 4-S8: Parts of the tannin involved in binding. Numbers inside the boxes overlaying the tannin indicate the number of times that segment was involved with an enzyme residue (>90% occupancy).

Previous studies have indicated that the phenol ring with the ortho-hydroxyl groups (B, E, H, K rings) are essential for binding to proteins such as bovine serum albumin[192-194]. In our simulations, we see a similar trend in which the B, E, H, K rings had 56 observed binding sites compared to the phenol rings with meta-hydroxyl groups, A. D, G, J rings, with 42 binding interactions. This could be due to either 1) the flexibility in the center ring allows the B ring to explore more sites during the energy minimization, 2) the two hydroxyl groups on the ortho-

phenol working together to maintain hydrogen bonding better than the meta-hydroxyls or 3) the A ring of the terminal flavonoid is unfairly biased against during the molecular docking procedure. The tannin has a fair amount of flexibility during the simulation, but only one static structure is used in Patchdock. It is possible that this biases the tannin against the J ring. More simulations with different tannin conformations that reflect the tannin flexibility would be needed to confirm the trends seen in these simulations.

4.6.6. Longest Residence Time interactions for other 1CEL Binding Sites

Table 4S-3: Residence Times of Binding Sites

		Binding Site #							
Residue #	Residue Type	1	3	4	6	7	8	9	10
2	SER	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	ALA	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
4	CYS	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
5	THR	0.0	0.0	0.0	0.1	0.0	0.7	0.0	0.0
6	LEU	0.0	0.0	0.0	0.1	0.0	0.0	0.4	0.0
7	GLN	0.0	0.0	0.0	10.0	0.0	0.1	2.1	0.0
8	SER	0.0	0.0	0.0	0.2	0.0	0.6	0.0	0.0
9	GLU	0.0	0.0	0.0	0.1	0.0	20.0	0.0	0.0
10	THR	0.0	0.0	0.0	40.0	0.5	0.3	0.0	0.0
11	HIS	0.0	0.0	0.0	0.4	0.1	1.9	0.0	0.0
12	PRO	0.0	0.0	0.0	2.0	2.7	0.0	0.0	0.0
13	PRO	0.0	0.0	0.0	0.1	0.0	0.9	0.0	0.0
14	LEU	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
15	THR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
16	TRP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
17	GLN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
18	LYS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
19	CYS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25	CYS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
26	THR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

27	GLN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
28	GLN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
29	THR	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0
30	GLY	0.0	0.0	0.0	0.0	0.0	0.6	0.0	0.0
31	SER	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
32	VAL	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
33	VAL	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
34	ILE	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
37	ASN	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0
38	TRP	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.0
39	ARG	0.0	0.0	0.0	0.0	0.0	40.0	0.2	0.0
40	TRP	0.0	0.0	0.0	0.1	0.0	0.0	3.3	0.0
41	THR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
42	HIS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
47	SER	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0
48	THR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
49	ASN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
51	TYR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
53	GLY	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
73	LEU	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
74	ASP	0.0	0.0	0.0	0.0	0.0	0.5	0.1	0.0
75	GLY	0.0	0.0	0.0	0.1	0.0	20.0	0.5	0.0
76	ALA	0.0	0.0	0.0	2.2	0.0	0.0	0.7	0.0
77	ALA	0.0	0.0	0.0	0.5	0.0	0.0	0.2	0.0
78	TYR	0.0	0.0	0.0	40.0	0.1	0.0	0.0	0.0
79	ALA	0.0	0.0	0.0	1.7	1.9	0.0	0.0	0.0
80	SER	0.0	0.0	0.0	0.0	0.4	0.0	0.3	0.0
81	THR	0.0	0.0	0.0	0.0	0.1	0.0	0.3	0.0
82	TYR	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0
83	GLY	0.0	0.0	0.0	0.0	40.0	0.0	0.0	0.0
84	VAL	0.0	0.0	0.0	1.2	40.0	0.0	0.0	0.0
85	THR	0.0	0.0	0.0	0.4	1.6	0.0	0.0	0.0
86	THR	0.0	0.0	0.0	20.0	0.3	0.0	0.0	0.0
87	SER	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0
88	GLY	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
89	ASN	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0
90	SER	0.0	0.0	0.0	0.0	1.4	0.0	0.0	0.0
91	LEU	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
92	SER	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
93	ILE	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

94	GLY	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
95	PHE	0.0	0.0	0.0	0.0	0.0	0.0	0.1	3.0
96	VAL	0.0	0.0	0.0	0.0	0.0	0.0	0.1	1.0
97	THR	0.0	0.0	0.0	0.0	1.7	0.0	0.1	0.0
98	GLN	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.6
99	SER	0.0	0.0	0.0	0.0	0.2	0.0	0.4	0.7
100	ALA	0.0	0.0	0.0	0.0	0.2	0.0	0.2	0.1
101	GLN	0.0	0.0	0.0	0.0	20.0	0.0	0.2	0.1
102	LYS	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.1
103	ASN	0.0	0.0	0.0	0.0	0.2	0.0	0.4	0.1
104	VAL	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.5
105	GLY	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
111	MET	0.0	0.0	0.0	0.0	0.0	2.5	0.0	0.0
112	ALA	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
113	SER	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0
114	ASP	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
115	THR	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
116	THR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
132	ASP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
135	GLN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
136	LEU	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
137	PRO	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7
138	CYS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4
139	GLY	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
140	LEU	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0
155	VAL	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
161	ASN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
162	THR	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
163	ALA	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
166	LYS	0.0	0.0	0.0	0.0	0.0	40.0	0.0	0.0
167	TYR	0.0	0.0	0.0	0.0	0.0	40.0	0.0	0.0
175	GLN	6.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0
178	ARG	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
181	LYS	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
193	GLU	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
195	SER	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
196	SER	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
197	ASN	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1
198	ASN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3
199	ALA	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

200	ASN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2
201	THR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2
202	GLY	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
214	ASP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
217	GLY	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
222	SER	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
223	GLU	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
224	ALA	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0
225	LEU	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0
226	THR	0.0	0.0	3.3	0.0	0.0	0.0	0.0	0.0
228	HIS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
245	GLY	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
246	THR	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
247	TYR	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
248	SER	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0
249	ASP	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
250	ASN	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
251	ARG	1.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0
252	TYR	0.4	0.0	0.1	0.0	0.0	0.0	0.0	0.0
257	ASP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
258	PRO	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
259	ASP	0.5	5.7	1.0	0.0	0.0	0.0	0.0	0.0
260	GLY	0.0	0.3	0.3	0.0	0.0	0.0	0.0	0.0
261	CYS	0.0	1.0	0.4	0.0	0.0	0.0	0.0	0.0
262	ASP	0.2	0.9	0.1	0.0	0.0	0.0	0.0	0.0
263	TRP	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
264	ASN	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
265	PRO	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
266	TYR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
267	ARG	0.4	0.2	0.2	0.0	0.0	0.0	0.0	0.0
268	LEU	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
287	LYS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
317	GLU	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
318	LEU	0.0	0.2	0.3	0.0	0.0	0.0	0.0	0.0
319	GLY	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
320	SER	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
321	TYR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
330	TYR	0.0	2.5	0.0	0.0	0.0	0.0	0.0	0.0
331	CYS	0.0	2.8	3.1	0.0	0.0	0.0	0.0	0.0
332	THR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

333	ALA	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
334	GLU	0.0	0.1	1.0	0.0	0.0	0.0	0.0	0.0
335	GLU	0.0	1.0	20.0	0.0	0.0	0.0	0.0	0.0
336	ALA	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
337	GLU	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0
338	PHE	0.0	6.7	10.0	0.0	0.0	0.0	0.0	0.0
339	GLY	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0
340	GLY	0.1	0.6	0.0	0.0	0.0	0.0	0.0	0.0
341	SER	0.1	0.7	0.0	0.0	0.0	0.0	0.0	0.0
342	SER	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
343	PHE	0.0	0.5	40.0	0.0	0.0	0.0	0.0	0.0
344	SER	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
367	TRP	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
368	ASP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
369	ASP	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.1
370	TYR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8
371	TYR	4.4	0.0	0.0	0.0	0.0	0.0	0.0	0.2
372	ALA	2.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
373	ASN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
374	MET	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
375	LEU	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
376	TRP	20.0	0.0	0.8	0.0	0.0	0.0	0.0	0.0
377	LEU	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0
378	ASP	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
379	SER	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
380	THR	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
381	TYR	0.4	1.1	20.0	0.0	0.0	0.0	0.0	0.0
382	PRO	1.1	0.2	0.2	0.0	0.0	0.0	0.0	0.0
383	THR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
384	ASN	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0
385	GLU	0.0	0.4	0.2	0.0	0.0	0.0	0.0	0.0
386	THR	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
387	SER	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
388	SER	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
389	THR	0.7	0.3	0.1	0.0	0.0	0.0	0.0	0.0
390	PRO	0.2	0.2	1.1	0.0	0.0	0.0	0.0	0.0
391	GLY	0.1	1.9	0.2	0.0	0.0	0.0	0.0	0.0
392	ALA	0.5	1.7	0.4	0.0	0.0	0.0	0.0	0.0
393	VAL	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
394	ARG	0.1	0.0	10.0	0.0	0.0	0.0	0.0	0.0

395	GLY	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
401	SER	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4
402	GLY	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8
403	VAL	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
404	PRO	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3
405	ALA	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
406	GLN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
407	VAL	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2
408	GLU	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9
409	SER	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
410	GLN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
411	SER	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
412	PRO	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
413	ASN	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.6
414	ALA	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
415	LYS	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0
416	VAL	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
417	THR	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
418	PHE	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
419	SER	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
420	ASN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 4-S3: Residue binding times for 1CEL binding sites. Omitted residues have a value of 0.

		Binding Site
Residue #	Residue Type	17
8	CYS	0.0
9	GLY	0.6
10	GLY	1.1
11	ILE	0.6
12	GLY	0.4
13	TYR	0.7
14	SER	0.1
15	GLY	0.2
16	PRO	0.9
17	THR	0.1
18	VAL	0.1
19	CYS	0.0
25	CYS	0.0
26	GLN	0.2
27	VAL	0.1
28	LEU	0.2
29	ASN	0.0
30	PRO	0.1
31	TYR	1.0
32	TYR	0.5
33	SER	0.1
34	GLN	0.2

Table 4-S4: Residence times for the binding site on 1CBH. Omitted residues have a value of 0.

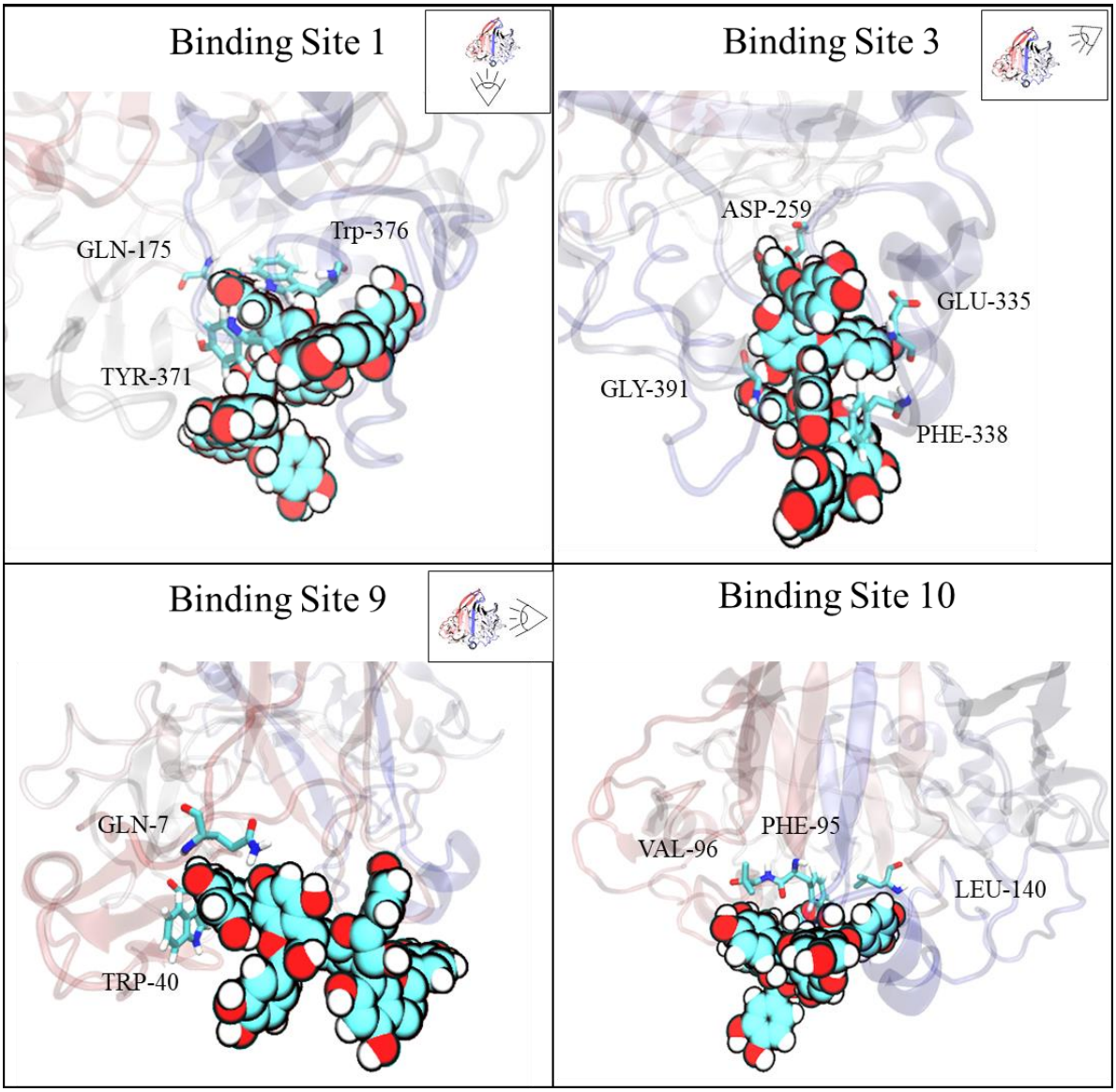


Figure 4-S9: Illustrations of the strongest interactions of the remaining confirmed binding site interactions seen in Figure 4-5.

Chapter 5: Concluding Remarks

5.1 The case for further studying extractives

NARA's biofuel production process is not yet economical and needs to incorporate more high value secondary products and find process improvements. The oil industry makes 10 co-products during refining (constituting at least 1% of the raw crude resource)[208] while NARA's current process focuses on 4 secondary products (jet fuel, activated carbon, lignosulfonates, and electricity from burning extra unusable material). The fastest way to make these fermentation based biofuel routes economical would be to add further revenue streams to the process. Products derived from the most abundant non-carbohydrate components, lignin and extractives, are of particular interest because they would not harm biofuel yields. Production of materials such as vanillin, syringaldehyde, and syringol[209] from lignin could be promising as well as copolymerization of lignin with currently available petroleum-based polymers[210]. However, 50 million tons of technical lignin from lignocellulose is already produced annually by pulping industries and most of it has only been found economical to be used as in-house fuel[211]. Extractives have not received the same amount of attention, and there could be money to be made with these diverse phytochemicals.

Tannins and phenols could be possible secondary products, and their high water solubility allows them to be extracted without organic solvents. Their removal would also benefit saccharification and fermentation. This would require exploration of current tannin markets since much of our economics and knowledges is 40+ years old[79]. Investigating modern uses of tannins could also spur implementation of biofuel production. Further exploration into the market price of extractives was not carried out in this work, but may be very helpful for

feedstocks with extractive-rich bark. Phytohormones such as the phytosterols could be of interest to the health and pharmaceutical industry, but would require the use of organic extraction solvents. However, the health benefits of using foods enriched with phytosterols was reported to have the possibly of saving the Canadian health care system up to $2.45 billion annually[212].

Regardless, it is only a matter of time until the scarcity of oil resources make biomass based biofuels cost competitive. Process optimizations around extractives could speed up biofuel incorporation into the marketplace. The pulp and paper industry has long known that extractives can impact biomass processing. They are even included in NREL biofuel process models of corn stover, but these models assume all extractives have the properties of glucose[81, 213]. While that may be appropriate for corn stover, assuming Douglas-fir extractives act like glucose is a very bad assumption. Understanding what inhibits a process could provide new ways to improve the process. Studies on how extractives are affected in actual biofuel pretreatments and fermentation have either not been carried out or have not been made available to the public. We have begun to explore how extractives affect saccharification and believe that further research is warranted based on how much inhibition we see can occur.

5.2 Summary of implications of extractives in biofuel production

Extractives are both possible revenue sources and inhibitors at different points in chemical processing. The phytochemical extractive category make up 5% to 25% of the dry weight for different tissues of Douglas-fir and the non-carbohydrate extractives are expected to make up ~8% of the forestry residue. Extractives can undergo a variety of reaction in pretreatment, and there is a need to investigate which ones predominantly occur. Dihydroquercetin and condensed tannins have been specifically studied for their reactivity in sulfite processing [30, 44] and it can be

expected that other components such as waxes will also be reactive to some extent, depending on details of the pretreatment conditions [144]. Treatment by sulfuric acid further polymerizes condensed tannins, while treatment by sulfite breaks them up into smaller sulfonated units [42-44]. With sulfite/bisulfite biofuel treatments using chemicals such as sulfuric acid, sulfite, and bisulfite [80, 137], it is unclear which preferentially happens.

Depending on the amount of bark in the slash and processing, saccharification can expect from 10-100 g of proanthocyanidins and phlobaphenes/kg o.d. dry sugars, 10-100g/kg (o.d. dry sugars) of flavonoids, 10-100g/kg of waxes, 1-10 g/kg of diterpenoids and monoterpenoids, and 1-10 g/kg of phytosterols. Extractives that show up in significant quantities in the feed for fermentation include waxes, proanthocyanidins and phlobaphenes, hydrophobic flavonoids, diterpenoids, and phytosterols. Diterpenoids are of particular interest because of their known toxicity [105] as well as phenols for their known antifungal properties [45, 46]. Extractives have been found to be some of the most toxic components of mechanically and sulfite pretreated pulp waste streams and will also likely be a concern for biofuel waste. This toxicity is due to compounds such as the diterpenoids whose parts per million aqueous solubility are toxic for fish [105] and phytosterols which are possible endocrine disruptors [120].

Tannins of Douglas-fir have been extensively studied for use in industries such as leather tanning, drilling muds, ore flotation, ceramics, and cement. However, competition with South American and European tannin production has limited its use in these areas. It is commonly believed that harvesting tannins in a multi-product process could be possible, but an industry based on extracting the tannins alone is not economical [79]. Considering that tannins were found to be inhibitory to saccharification, extractions of tannins could be a way to simultaneously remove an inhibitor and collect another secondary product.

5.3 Summary of findings on tannin inhibition of cellulases

Tannins are the most potent extractive saccharification inhibitors and show up primarily in the bark. Douglas-fir bark extracts were shown to cause between 15-25% inhibition of *Trichoderma reesei* exoglucanases, endoglucanases, and β-glucosidases at biofuel relevant conditions. The inhibition stems from tannin binding to the cellulase as opposed to the cellulose and Lineweaver-Burk plots were used to show that bark extract inhibits β-glucosidase via an uncompetitive binding mechanism. UV-Vis absorption at 280nm was also shown to be correlated with inhibitors and could be developed into a useful tool for characterizing a solutions resistance to saccharification.

Molecular dynamics simulations studying how a Douglas-fir procyanidin tetramer (condensed tannin) inhibits the exoglucanase Cel7A showed that inhibition occurs on the catalytic module of the enzyme near the entrance or in the exit of the catalytic tunnel. Sites were seen to consist of a minimum of 6 amino acid-tannin interactions and as many as 19. These multi-residue interactions cooperate to stabilize the binding structure. Residues in the binding sites were most commonly hydrophobic or polar types and belonging to unstructured or loop regions. Binding sites were primarily comprised of loops around the catalytic tunnel where the enzyme has some flexibility to find optimal alignments between the amino acids and tannin aromatic rings. Inhibition thus appears to result primarily from either clogging the catalytic tunnel or interfering with the loops responsible for guiding the cellulose onto catalytic residues. This would imply a competitive or mixed inhibition mechanism. Creating loops that are more highly structured along the catalytic tunnel could reduce the ability of these areas to optimally align with the tannin rings. In addition, removal or modification of aromatic and charged residues could also be investigated.

5.4 Resultant Publications List

1. Oleson, K.R., Schwartz, D.T., Extractives in Douglas-fir forestry residue and considerations for biofuel production. *Phytochemistry Reviews*. 2016

2. Oleson, K.R., Schwartz, D.T., Inhibition of *Trichoderma reesei* cellulases by aqueous Douglas-fir bark extractives. Submitted to *Bioresource Technology* 2018.

3. Oleson, K.R. Sprenger K.G., Pfaendtner, J.P., Schwartz, D.T., Inhibition of Exoglucanase CEL7A by a Douglas-fir tannin. Preparing for submission to *Journal of Physical Chemistry B*. 2018

Chapter 6: Preliminary Results for Suggested Future Studies

There is still much work to be done in order to comprehensively understand the possible impacts extractives can have on biofuel production. The work here could be further expanded on for a clearer understanding of how Douglas-fir extractives inhibit saccharification. Of particular interest are how extractives react in pretreatment or saccharification buffer and how these molecules affect saccharification. We observed the evolution of yellow particles in solutions containing taxifolin over several days, indicating that the taxifolin oxidized to quercetin in our saccharification buffer. Correspondingly the extract solutions changed to a darker color over the same time span. Explorative experiments on how reacted extractives affect saccharification are presented in section 6.1 and show that the reactive extractives may be more inhibitory than their unreacted counterparts. Assessing what other molecules react in the buffer, how fast this occurs, and a better understanding of the inhibition mechanisms would be of interest This would be an excellent complementary study since these would be relevant molecules in the saccharification reactor.

Tannins are also known for their ability to aggregate and even precipitate various enzymes. Dynamics light scattering experiments showed that aggregation was not occurring over the time scale of these experiments and these concentrations, but higher concentrations of tannin could cause this to happen. While this means enzyme aggregation is not occurring, it is possible that mechanical processing in actual biofuel processes could allow for better tannins extraction and could release more tannins into solution. It could be useful to determine what concentration of tannin in the extractives cause the onset of aggregation as this could mean a shift in inhibition mechanism. Dynamic light scattering experiments of what happens when cellulases are

introduced into an extractive solution or a concentrated tannin solution are presented in Section 6.2.

Further work into the mechanism of endoglucanase inhibition would also help bioengineers create a tannin-resistant endoglucanase. Binding sites proposed from Patchdock for the tannin-endoglucanase system are shown in Section 6.3. Further work needs to be done to confirm and characterized these binding sites similar to the exoglucanase study.

6.1 Preliminary results on inhibition caused by reacted extractives

One difficulty in studying the effects of extractives on saccharification is that some of them are known antioxidants and can react in the citrate buffer. Reactions such as the oxidation of taxifolin (dihydroquercetin) to quercetin are known and have been reported in literature at high acid concentrations[30]. The citrate buffers in experiments with taxifolin developed yellow particles over time that were visible to the eye. Upon vacuum filtration of 3 day old taxifolin-buffer solution and subsequent dissolution of the solids in ethanol, the new solution had a new UV-VIS absorption at 375nm which is characteristic of quercetin[214]. Reacted solutions were tested to see if this could lead to inhibition as well. Saccharification experiments with extractives that reacted in the buffer for 72 hours are shown in Figure 6-1.

Saccharification experiments with bark extract, tannin supplement, and taxifolin were carried out in which cellulose-buffer-extractive solutions were shaken and heated at 50°C for 72 hours before cellulase addition. Experimental conditions were otherwise kept the same as previous experiments.

As can be seen, even with oxidation reactions occurring in the buffer, the extractives have a very similar inhibition to the non-reacted case for the case of crystalline cellulose. The lower

solubility of quercetin causes it to form visible particles in solution and thus could cause inhibition through non-productive solid phase binding similar to observed for solid lignin (Source). However there is much more solid cellulose in solution (250g/L) in these experiments than the 1g/L maximum of quercetin particles. This means that the exoglucanases' catalytic binding module will rarely see the surface of a quercetin particle compared to a cellulose surface. There would need to be very strong binding to the quercetin surface compared to the cellulose surface in order to see inhibition. Since there is no statistically relevant inhibition seen for the crystalline cellulose experiments, it indicates that this is not the case.

An approximately 9% inhibition is seen for the case of amorphous cellulose with quercetin however. The quercetin present a new solid phase in solution for the carbohydrate binding modules to bind onto, and the amorphous cellulose used here does not have surfaces like the crystalline cellulose. Thus a carbohydrate binding module is much more likely to find a quercetin surface and cause the catalytic module to be restricted when there are no cellulose surfaces more readily available. It is also notable that the reacted tannin results appear to be stronger than the inhibition from the reacted bark extracts. This should be double checked to make sure that this is actually happening as opposed to some experimental error. If this is true though it may imply that tannins from different sources react differently in solution and result in different inhibition percent.

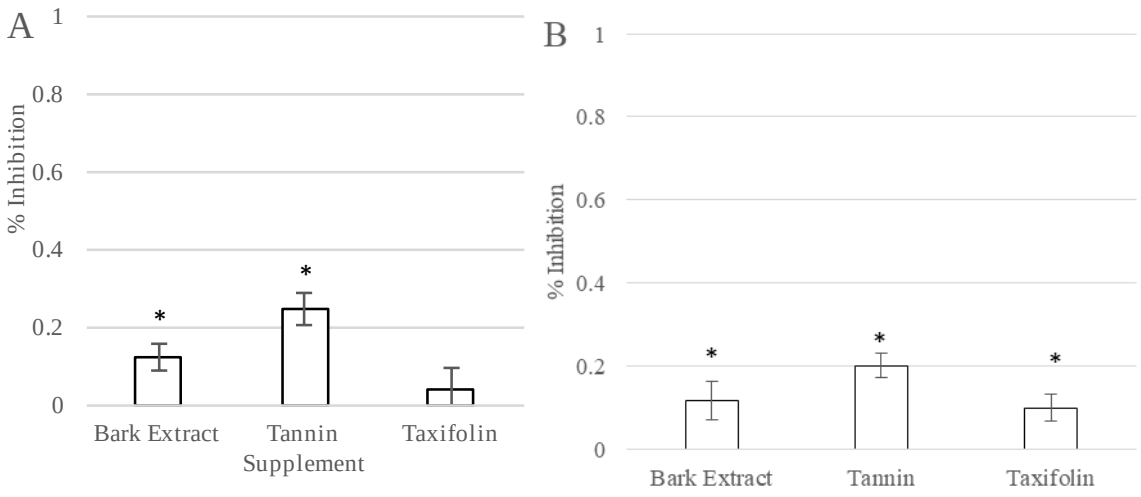


Figure 6-1: Saccharification yields from amorphous and crystalline cellulose with or without bark extractives compounds that had been given 72 hours to react. A) Equivalent cellobiose yield after 1 hour from crystalline cellulose. B) Equivalent glucose yield after 6 hours from amorphous cellulose. * Indicates inhibition at a 95% confidence level.

6.2 Preliminary results on aggregation of cellulases by tannins

Tannins are well known from their ability to aggregate proteins and peptides under appropriate conditions. Tests were undertaken to determine if tannin induced aggregation was occurring in our system during our experiments. Solutions of the bark extract and tannin extract were prepared as previously described except that the tannin supplement solution was prepared so that a larger amount (3g/L of tannins to match the initial biofuel simulations) was added. Cellulase solutions were added to these buffer solutions and filtered through a 0.02 micrometer filter. The resultant bark solutions contained particles with hydrodynamic radius around 40nm. Solutions were filtered into a disposable cuvette added to a Malvern Zetasizer for dynamic light scattering. Individual hydrodynamic radius measurements were tracked over time at 25°C with

general protein parameters in the instrument. This information for cellulase in the bark solution and elevated tannin supplement solution are presented in Figure 6-2.

An increase in hydrodynamic radius indicates that aggregation is occurring. As can be seen, the hydrodynamic radius of the bark solution remains unchanged while the solution with the elevated tannin concentration increases. This shows that tannins are capable of causing aggregation but our experiments are likely not at a high enough concentration for this to occur.

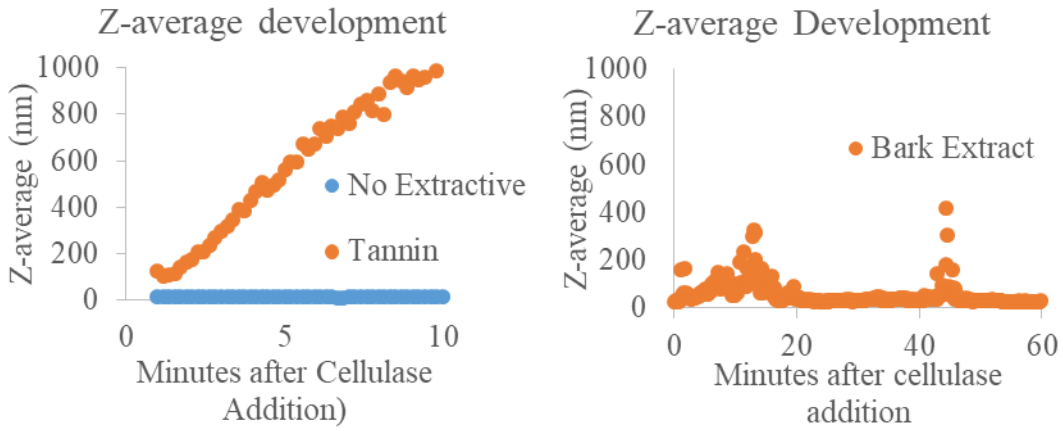


Figure 6-2: Hydrodynamic radius changes when cellulases are added into different solutions.

6.3 Preliminary results on binding of tannin-endoglucanase binding

The endoglucanase Cel7B is one of the major endoglucanases produced by *Trichoderma reesei* and accounts for 5-10% of the total cellulases[215, 216]. The crystal structure of the endoglucanase shows that it has a substrate-binding cleft[217] and the carbohydrate binding module has minimal impact on pretreated substrates[218]. Binding sites Patchdock suggested on this enzyme are presented in Figure 6-3. The PCA residue (residue 401) and ligands were removed from the PDB file. The tannins are placed almost in a line and this line goes along the catalytic

cleft. This would make competitive type inhibition the most likely inhibition mechanism, since the catalytic cleft would be blocked by these molecules.

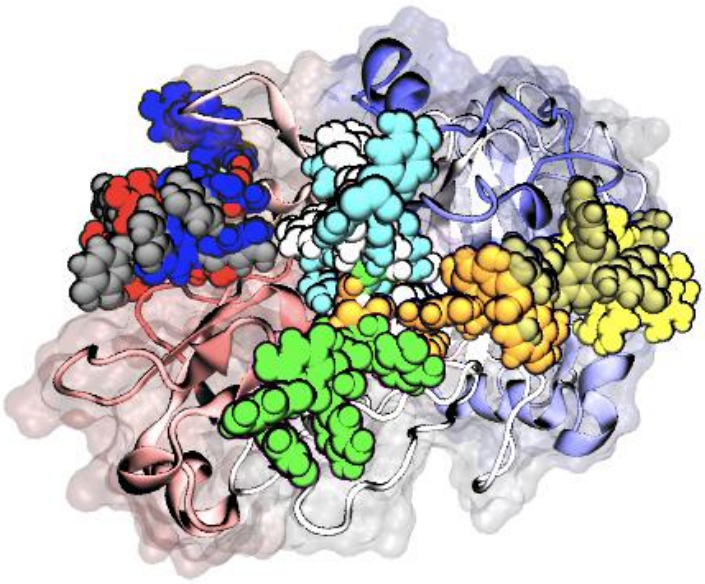


Figure 6-3: Patchdock suggested binding sites on Endoglucanase I from *Trichoderma reesei.*

Chapter 7: References

1. USDA, N. A. R. A. *Wood to Wing | The economics of a wood-based biorefinery in Longview, Washington*; 2017; pp 1-2.
2. Searchinger, T.; Heimlich, R.; Houghton, R. A.; Dong, F. X.; Elobeid, A.; Fabiosa, J.; Tokgoz, S.; Hayes, D.; Yu, T.-H., Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change. *Science* **2008,** 319, (5867), 1238-1240.
3. Fargione, J.; Hill, J.; Tilman, D.; Polasky, S.; Hawthorne, P., Land clearing and the biofuel carbon debt. *Science* **2008,** 319, (5867), 1235-1238.
4. NARA *Wood to wing: Residual biomass harvesting and carbon emissions*; USDA: 2017.
5. Perlack, R.; Stokes, B. *U.S. Billion-Ton Update: Biomass Supply for a Bioenergy and Bioproducts Industry*; Oakridge National Laboratory: Oak Ridge (TN), 2011; pp 16-51.
6. Alliance, N. A. R.; USDA *NARA Supply Chain*; 2014.
7. Davidson, C.; Newes, E.; Schwab, A.; Vimmerstedt, L. *An overview of aviation fuel markets for biofuel stakeholders*; NREL: 2014; pp 1-44.
8. Kamin, R.; Rudy, M., From seed to supersonic: How camelina powered the Navy's premier fighter jet. *Currents* 2011.
9. Tao, L.; Milbrandt, A.; Zhang, Y. A.; Wang, W. C., Techno-economic and resource analysis of hydroprocessed renewable jet fuel. *Biotechnol Biofuels* **2017,** 10, 16.
10. Zhu, J. Y.; Chandra, M. S.; Gu, F.; Gleisner, R.; Reiner, R.; Sessions, J.; Marrs, G.; Gao, J.; Anderson, D., Using sulfite chemistry for robust bioconversion of Douglas-fir forest residue to bioethanol at high titer and lignosulfonate: A pilot-scale evaluation. *Bioresour Technol* **2015,** 179, 390-397.
11. Zhou, H. F.; Zhu, J. Y.; Gleisner, R.; Qiu, X. Q.; Horn, E.; Negron, J., Pilot-scale demonstration of SPORL for bioconversion of lodgepole pine to bioethanol and lignosulfonate. *Holzforschung* **2016,** 70, (1), 21-30.
12. Zhou, H. F.; Zhu, J. Y.; Gleisner, R.; Qiu, X. Q.; Horn, E., High titer ethanol and lignosulfonate production from SPORL pretreated poplar at pilot scale. *Front Energy Res* **2015**, 9.
13. News, W. S. U. *Forest-powered biofuel flight heads to Washington, D.C.*; 2016.
14. Marrs, G.; Spink, T.; Gao, A. *Process design and economics for biochemical conversion of softwood lignocellulosic biomass to isoparaffinic kerosene and lignin co-products*; 2016; pp 1-76.
15. Airnav, L. Jet A prices within 50 miles of Seattle, WA 98103. www.airnav.com/fuel/local.html
16. Zhu, J. Y.; Pan, X. J.; Zalesny, R. S., Pretreatment of woody biomass for biofuel production: energy efficiency, technologies, and recalcitrance. *Applied Microbiology and Biotechnology* **2010,** 87, (3), 847-857.
17. Kim, K. H.; Tucker, M.; Nguyen, Q., Conversion of bark-rich biomass mixture into fermentable sugar by two-stage dilute acid-catalyzed hydrolysis. *Bioresource Technology* **2005,** 96, (11), 1249-1255.
18. Zhang, C.; Lei, X. C.; Scott, C. T.; Zhu, J. Y.; Li, K. C., Comparison of Dilute Acid and Sulfite Pretreatment for enzymatic saccharification of earlywood and latewood of Douglas fir. *Bioenerg Res* **2014,** 7, (1), 362-370.

19. Gao, J.; Anderson, D.; Levie, B., Saccharification of recalcitrant biomass and integration options for lignocellulosic sugars from Catchlight Energy's sugar process (CLE Sugar). *Biotechnol Biofuels* **2013,** 6, 7.
20. Zhao, X. B.; Cheng, K. K.; Liu, D. H., Organosolv pretreatment of lignocellulosic biomass for enzymatic hydrolysis. *Appl Microbiol and Biot* **2009,** 82, (5), 815-827.
21. Kurabi, A.; Berlin, A.; Gilkes, N.; Kilburn, D.; Bura, R.; Robinson, J.; Markov, A.; Skomarovsky, A.; Gusakov, A.; Okunev, O.; Sinitsyn, A.; Gregg, D.; Xie, D.; Saddler, J., Enzymatic hydrolysis of steam-exploded and ethanol organosolv-pretreated Douglas-fir by novel and commercial fungal-cellulases. *Appl Biochem Biotech* **2005,** 121, 219-230.
22. Zhu, J. Y.; Pan, X. J.; Wang, G. S.; Gleisner, R., Sulfite pretreatment (SPORL) for robust enzymatic saccharification of spruce and red pine. *Bioresource Technol* **2009,** 100, (8), 2411-2418.
23. Wang, Z. J.; Lan, T. Q.; Zhu, J. Y., Lignosulfonate and elevated pH can enhance enzymatic saccharification of lignocelluloses. *Biotechnol Biofuels* **2013,** 6, 10.
24. Zhu, J. Y.; Zhu, W. Y.; Obryan, P.; Dien, B. S.; Tian, S.; Gleisner, R.; Pan, X. J., Ethanol production from SPORL-pretreated lodgepole pine: preliminary evaluation of mass balance and process energy efficiency. *Appl Microbiol Biot* **2010,** 86, (5), 1355-1365.
25. Luo, X.; Gleisner, R.; Tian, S.; Negron, J.; Zhu, W.; Horn, E.; Pan, X. J.; Zhu, J. Y., Evaluation of Mountain Beetle-infested Lodgepole Pine for cellulosic ethanol production by sulfite pretreatment to overcome recalcitrance of lignocellulose. *Ind Eng Chem* **2010,** 49, (17), 8258-8266.
26. Zhu, J. Y., Physical pretreatment - woody biomass size reduction - for forest biorefinery. In *Sustainable Production of Fuels, Chemicals, and Fibers from Forest Biomass*, 2011; pp 89-107.
27. Brandt, A.; Grasvik, J.; Hallett, J. P.; Welton, T., Deconstruction of lignocellulosic biomass with ionic liquids. *Green Chem* **2013,** 15, (3), 550-583.
28. Kumar, A. K.; Sharma, S., Recent updates on different methods of pretreatment of lignocellulosic feedstocks: a review. *Bioresour Bioprocess* **2017,** 4, 1-19.
29. U.S.D.A., *U.S. Forest Facts and Historical Trends*. 2001; p 20.
30. Hoge, W. H., The Resistance of Douglas-Fir to Sulphite Pulping. *TAPPI* **1954,** 37, (9), 369-376.
31. Zhang, C.; Houtman, C. J.; Zhu, J. Y., Using low temperature to balance enzymatic saccharification and furan formation during SPORL pretreatment of Douglas-fir. *Process Biochem* **2014,** 49, (3), 466-473.
32. Zhou, H. F.; Zhu, J. Y.; Gleisner, R.; Qiu, X. Q.; Horn, E., High titer ethanol and lignosulfonate production from SPORL pretreated poplar at pilot scale. *Frontiers in Energy Research* **2015**, 9.
33. Novozymes, Cellulosic ethanol: Novozymes Cellic CTec3 - secure your plant's lowest total cost. In 2012; pp -6.
34. NARA *NARA Goal One*; 2014-2015; pp 1-63.
35. Papari, S.; Hawboldt, K., A review on the pyrolysis of woody biomass to bio-oil: Focus on kinetic models. *Renew Sust Energy Rev* **2015,** 52, 1580-1595.
36. Mohan, D.; Pittman, C. U.; Steele, P. H., Pyrolysis of wood/biomass for bio-oil: A critical review. *Energ Fuel* **2006,** 20, (3), 848-889.
37. Roberts, A. F., Review of kinetics data for pyrolysis of wood and related substances. *Combust Flame* **1970,** 14, (1-3), 261-&.

38. Mahmood-Khan, Z.; Hall, E. R., Occurrence and removal of plant sterols in pulp and paper mill effluents. *J Environ Eng Sci* **2003,** 2, (1), 17-26.
39. Hubbard, J. K. The distribution and properties of the tannin in Douglas fir bark (Pseudotsuga taxifolia, Britt.). Oregon State College, 1949.
40. Oleson, K. R.; Schwartz, D. T., Extractives in Douglas-fir forestry residue and considerations for biofuel production. *Phytochem Rev* **2016,** 15, (5), 985-1008.
41. Cala, O.; Dufourc, E. J.; Fouquet, E.; Manigand, C.; Laguerre, M.; Pianet, I., The colloidal state of tannins impacts the nature of their interaction with proteins: The case of salivary proline-rich protein/procyanidins binding. *Langmuir* **2012,** 28, (50), 17410-17418.
42. Graham, H. M.; Kurth, E. F., Constituents of extractives from Douglas fir. *Ind Eng Chem* **1949,** 41, (2), 409-414.
43. Bae, Y. S.; Malan, J. C. S.; Karchesy, J. J., Sulfonation of procyanidin polymers: evidence of intramolecular rearrangement and aromatic ring substitutions *Holzforschung* **1994,** 48, (2), 119-123.
44. Foo, L. Y.; McGraw, G. W.; Hemingway, R. W., Condensed tannins: preferential substitution at the interflavanoid bond by sulphite ion. *J Chem Soc Chem Comm* **1983**, (12), 672-673.
45. Scalbert, A., Antimicrobial properties of tannins. *Phytochemistry* **1991,** 30, (12), 3875-3883.
46. Kennedy, R. W. Fungicidal toxicity of certain extraneous components of Douglas-fir heartwood. University of British Columbia, 1955.
47. Miron, D.; Battisti, F.; Silva, F. K.; Lana, A. D.; Pippi, B.; Casanova, B.; Gnoatto, S.; Fuentefria, A.; Mayorga, P.; Schapoval, E. E. S., Antifungal activity and mechanism of action of monoterpenes against dermatophytes and yeasts. *Rev Bras Farmacogn* **2014,** 24, (6), 660-667.
48. Tesevic, V.; Milosavljevic, S.; Vajs, V.; Dordevic, I.; Sokovic, M.; Lavadinovic, V.; Novakovic, M., Chemical composition and antifungal activity of the essential oil of Douglas fir (Pseudosuga menziesii Mirb. Franco) from Serbia. *J Serb Chem Soc* **2009,** 74, (10), 1035-1040.
49. McRae, J. M.; Kennedy, J. A., Wine and grape tannin interactions with salivary proteins and their impact on astringency: a review of current research. *Molecules* **2011,** 16, (3), 2348-2364.
50. Kato, C.; Goncalves, G.; Peralta, R.; Seixas, F.; Sa-Nakanishi, A.; Bracht, L.; Comar, J.; Bracht, A.; Peralta, R., Inhibition of α-amylases by condensed and hydrolysable tannins: focus on kinetics and hypoglycemic actions. *Enzyme Res* **2017,** 2017, 12.
51. Duarte, G. C.; Moreira, L. R. S.; Jaramillo, P. M. D.; Filho, E. X. F., Biomass-derived inhibitors of holocellulases. *Bioenerg Res* **2012,** 5, (3), 768-777.
52. Peng, G. M.; Roberts, J. C., Solubility and toxicity of resin acids. *Water Res* **2000,** 34, (10), 2779-2785.
53. Ali, M.; Sreekrishnan, T. R., Aquatic toxicity from pulp and paper mill effluents: a review. *Adv Environ Res* **2001,** 5, (2), 175-196.
54. Lehtinen, K. J.; Mattsson, K.; Tana, J.; Engstrom, C.; Lerche, O.; Hemming, J., Effects of wood-related sterols on the reproduction, egg survival, and offspring of brown trout (Salmo trutta lacustris L.). *Ecotox and Environ Safe* **1999,** 42, (1), 40-49.
55. Munkittrick, K. R.; Vanderkraak, G. J.; McMaster, M. E.; Portt, C. B.; Vandenheuvel, M. R.; Servos, M. R., Survey of Receiving-water environmental impacts associated with discharges from pulp-mills. 2. Gonad size, liver size, hepatic erod activity and plasma sex steroid-levels in white sucker. *Environ Toxicol Chem* **1994,** 13, (7), 1089-1101.

56. Lambertz, C.; Garvey, M.; Klinger, J.; Heesel, D.; Klose, H.; Fischer, R.; Commandeur, U., Challenges and advances in the heterologous expression of cellulolytic enzymes: a review. *Biotechnol Biofuels* **2014,** 7, (135), 15.
57. Gao, D. H.; Uppugundla, N.; Chundawat, S. P. S.; Yu, X. R.; Hermanson, S.; Gowda, K.; Brumm, P.; Mead, D.; Balan, V.; Dale, B. E., Hemicellulases and auxiliary enzymes for improved conversion of lignocellulosic biomass to monosaccharides. *Biotechnol Biofuels* **2011,** 4, (5), 11.
58. Girio, F. M.; Fonseca, C.; Carvalheiro, F.; Duarte, L. C.; Marques, S.; Bogel-Lukasik, R., Hemicelluloses for fuel ethanol: A review. *Bioresource Technol* **2010,** 101, (13), 4775-4800.
59. Saha, B. C., Hemicellulose bioconversion. *J Ind Microbiol Biot* **2003,** 30, (5), 279-291.
60. Pollegioni, L.; Tonin, F.; Rosini, E., Lignin-degrading enzymes. *Febs J* **2015,** 282, (7), 1190-1213.
61. Frandsen, K. E. H.; Simmons, T. J.; Dupree, P.; Poulsen, J. C. N.; Hemsworth, G. R.; Ciano, L.; Johnston, E. M.; Tovborg, M.; Johansen, K. S.; von Freiesleben, P.; Marmuse, L.; Fort, S.; Cottaz, S.; Driguez, H.; Henrissat, B.; Lenfant, N.; Tuna, F.; Baldansuren, A.; Davies, G. J.; Lo Leggio, L.; Walton, P. H., The molecular basis of polysaccharide cleavage by lytic polysaccharide monooxygenases. *Nat Chem Biol* **2016,** 12, (4), 298-303.
62. Kovalenko, V. I., Crystalline cellulose: structure and hydrogen bonds. *Russ Chem Rev* **2010,** 79, (3), 231-241.
63. Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P., Crystal structure and hydrogen bonding system in cellulose 1(alpha), from synchrotron X-ray and neutron fiber diffraction. *J Am Chem Soc* **2003,** 125, (47), 14300-14306.
64. Nishiyama, Y.; Langan, P.; Chanzy, H., Crystal structure and hydrogen-bonding system in cellulose 1 beta from synchrotron X-ray and neutron fiber diffraction. *J Am Chem Soc* **2002,** 124, (31), 9074-9082.
65. Medve, J.; Karlsson, J.; Lee, D.; Tjerneld, F., Hydrolysis of microcrystalline cellulose by cellobiohydrolase I and endoglucanase II from Trichoderma reesei: adsorption, sugar production pattern, and synergism of the enzymes. *Biotechnol Bioeng* **1998,** 59, (5), 621-634.
66. Shibafuji, Y.; Nakamura, A.; Uchihashi, T.; Sugimoto, N.; Fukuda, S.; Watanabe, H.; Samejima, M.; Ando, T.; Noji, H.; Koivula, A.; Igarashi, K.; Iino, R., Single-molecule Imaging Analysis of Elementary Reaction Steps of Trichoderma reesei Cellobiohydrolase I (Cel7A) Hydrolyzing Crystalline Cellulose I (alpha) and III (I). *J Biol Chem* **2014,** 289, (20), 14056-14065.
67. Arola, S.; Linder, M. B., Binding of cellulose binding modules reveal differences between cellulose substrates. *Sci Rep* **2016,** 6, 9.
68. Alekozai, E. M.; GhattyVenkataKrishna, P. K.; Uberbacher, E. C.; Crowley, M. F.; Smith, J. C.; Cheng, X. L., Simulation analysis of the cellulase Cel7A carbohydrate binding module on the surface of the cellulose I beta. *Cellulose* **2014,** 21, (2), 951-971.
69. Tejirian, A.; Xu, F., Inhibition of enzymatic cellulolysis by phenolic compounds. *Enzyme Microb Tech* **2011,** 48, (3), 239-247.
70. Hsieh, C. W. C.; Cannella, D.; Jorgensen, H.; Felby, C.; Thygesen, L. G., Cellulase inhibition by high concentrations of monosaccharides. *J Agr Food Chem* **2014,** 62, (17), 3800-3805.
71. Wojtusik, M.; Villar, J. C.; Zurita, M.; Ladero, M.; Garcia-Ochoa, F., Study of the enzymatic activity inhibition on the saccharification of acid pretreated corn stover. *Biomass Bioenerg* **2017,** 98, 1-7.

72. Tejirian, A.; Xu, F., Inhibition of cellulase-catalyzed lignocellulosic hydrolysis by iron and oxidative metal ions and complexes. *Appl Environ Microb* **2010,** 76, (23), 7673-7682.
73. Ximenes, E.; Kim, Y.; Mosier, N.; Dien, B.; Ladisch, M., Deactivation of cellulases by phenols. *Enzyme Microb Tech* **2011,** 48, (1), 54-60.
74. Bae, H. D.; McAllister, T. A.; Yanke, J.; Cheng, K. J.; Muir, A. D., Effects of condensed tannins on endoglucanase activity and filter paper digestion by *Fibrobacter succinogenes* S85. *Appl Environ Microb* **1993,** 59, (7), 2132-2138.
75. Ximenes, E.; Kim, Y.; Mosier, N.; Dien, B.; Ladisch, M., Inhibition of cellulases by phenols. *Enzyme Microb Tech* **2010,** 46, (3-4), 170-176.
76. Mullins, J. T.; Lee, J. H., Interactions of tannins with enzymes: A potential role in the reduced rate of ethanol fermentation from high-tannin biomass. *Biomass Bioenerg* **1991,** 1, (6), 355-361.
77. Jonsson, L. J.; Alriksson, B.; Nilvebrant, N. O., Bioconversion of lignocellulose: inhibitors and detoxification. *Biotechnol Biofuels* **2013,** 6, 10.
78. Jonsson, L. J.; Martin, C., Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects. *Bioresource Technol* **2016,** 199, 103-112.
79. Hall, A., *Utilization of Douglas-Fir bark*. Northwest Forest and Range Experiment Station, Forest Service, U.S. Department of Agriculture: Portland, Oregon, 1971.
80. Zhu, J. Y.; Chandra, M. S.; Gu, F.; Gleisner, R.; Reiner, R.; Sessions, J.; Marrs, G.; Gao, J.; Anderson, D., Using sulfite chemistry for robust bioconversion of Douglas-fir forest residue to bioethanol at high titer and lignosulfonate: A pilot-scale evaluation. *Bioresour. Technol.* **2015,** 179, 390-397.
81. Humbird, D.; Davis, R.; Tao, L.; Kinchin, C.; Hsu, D.; Aden, A.; Schoen, J.; Lukas, J.; Olthof, B.; Worley, M.; Sexton, D.; Dudgeon, D. *Process Design and Economics for Biochemical Conversion of Lignocellulosic Biomass to Ethanol*; National Renewable Energy Laboratory: Golden, Colorado, 2011; pp 1-136.
82. Leach, J. M.; Thakore, A. N., Toxic constituents in mechanical pulping effluents. *TAPPI* **1976,** 59, (2), 129-132.
83. Foo, L. Y.; McGraw, G. W.; Hemingway, R. W., Condensed Tannins: Preferential Substitution at the Interflavanoid Bond by Sulphite ion. *J. Chem. Soc.-Chem. Commun.* **1983**, (12), 672-673.
84. Kaar, W. E.; Brink, D. L., Summative analysis of 9 common North American woods. *J Wood Chem Technol* **1991,** 11, (4), 479-494.
85. Kurth, E. F., Chemicals from Douglas-fir bark. *TAPPI* **1953,** 36, (7), 119A-122A.
86. Turley, D. B.; Chaudhry, Q.; Watkins, R. W.; Clark, J. H.; Deswarte, F. E. I., Chemical products from temperate forest tree species - Developing strategies for exploitation. *Industrial Crops and Products* **2006,** 24, (3), 238-243.
87. Laver, M. L.; Loveland, P. M.; Chen, C. H.; Fang, H. H. L.; Zerrudo, J. V.; Liu, Y. C. L., Chemical Constituents of Douglas-Fir Bark: A Review of More Recent Literature. *Wood Sci.* **1977,** 10, (2), 85-92.
88. Rudloff, E. v., Chemosystematic studies in the genus *Pseudotsuga*. *I*. Leaf oil analysis of the coastal and Rocky Mountain varieties of the Douglas fir. *Canadian Journal of Botany* **1972,** 50(5), 1025-1040.
89. Foo, L. Y.; Karchesy, J. J., Procyanidin polymers of Douglas-fir bark - structure from degradation with phloroglucinol. *Phytochemistry* **1989,** 28, (11), 3185-3190.

90. Foo, L. Y.; Karchesy, J. J., PROCYANIDIN TETRAMERS AND PENTAMERS FROM DOUGLAS-FIR BARK. *Phytochemistry* **1991,** 30, (2), 667-670.
91. Kaundun, S. S.; Lebreton, P.; Bailly, A., Needle flavonoid variation in coastal Douglas-fir (Pseudotsuga menziesii var. menziesii) populations. *Canadian Journal of Botany-Revue Canadienne De Botanique* **1998,** 76, (12), 2076-2083.
92. Yeap Foo, L.; Karchesy, J. J., Chemical Nature of Phlobaphenes. In *Chemistry and Significance of Condensed Tannins*, Hemingway, R. W., Ed. Plenum Press: New York, 1989; pp 109-118.
93. Laver, M. L.; Fang, H. H. L., Ferulic Acid Esters from Bark of Pseudotsuga menziesii. *J. Agric. Food Chem.* **1989,** 37, (1), 114-116.
94. Pan, S. B.; Pu, Y. Q.; Foston, M.; Ragauskas, A. J., Compositional Characterization and Pyrolysis of Loblolly Pine and Douglas-fir Bark. *Bioenerg. Res.* **2013,** 6, (1), 24-34.
95. Foster, D. O.; Zinkel, D. F.; Conner, A. H., Tall oil percursors of Douglas fir. *Tappi* **1980,** 63, (12), 103-105.
96. Kurth, E. F., The composition of the wax in Douglas-fir bark. **1950,** J Am Chem Soc, (72(4)), 1685-1686.
97. Laver, M. L.; Fang, H. H. L., Ferulic acid esters from bark of Pseudotsuga menziesii. *J Agric Food Chem* **1989,** 37, (1), 114-116.
98. Foo, L. Y.; Helm, R.; Karchesy, J., 5',5' -bisdihydroquercetin - a B-ring linked biflavonoid from *Pseudotsuga menziesii Phytochem* **1992,** 31, (4), 1444-1445.
99. Foo, L. Y.; Karchesy, J., Pseudotsuganol, a biphenyl-linked pinoresinol dihydroquercetin from Douglas-fir bark - isolation of the 1st true flavonolignan. *J Chem Soc Chem Comm* **1989**, (4), 217-219.
100. Krauze-Baranowska, M.; Sowinski, P.; Kawiak, A.; Sparzak, B., Flavonoids from Pseudotsuga menziesii. *Zeitschrift Fur Naturforschung Section C-a Journal of Biosciences* **2013,** 68, (3-4), 87-96.
101. Dellus, V.; Mila, I.; Scalbert, A.; Menard, C.; Michon, V.; duPenhoat, C., Douglas-fir polyphenols and heartwood formation. *Phytochem.* **1997,** 45, (8), 1573-1578.
102. Hergert, H. L., Chemical Composition of Tannins and Polyphenols from Conifer Wood and Bark. *For. Prod. J.* **1960,** 10, (11), 610-617.
103. Pew, J. C., A flavonone from Douglas-fir heartwood. *J Am Chem Soc* **1948,** 70, (9), 3031-3034.
104. Foo, L. Y.; Karchesy, J. J., Polyphenolic glycosides from Douglas-fir inner bark. *Phytochem* **1989,** 28, (4), 1237-1240.
105. Peng, G. M.; Roberts, J. C., Solubility and toxicity of resin acids. *Water Res.* **2000,** 34, (10), 2779-2785.
106. Ali, M.; Sreekrishnan, T. R., Aquatic toxicity from pulp and paper mill effluents: a review. *Advances in Environmental Research* **2001,** 5, (2), 175-196.
107. Buchbauer, G.; Jirovetz, L.; Wasicky, M.; Nikiforov, A., Comparative investigation of Douglas-fir headspace samples, essential oils, and extracts (needles and twigs) Using GC-FID and GC-FTIR-MS. *J Agr Food Chem* **1994,** 42, (12), 2852-2854.
108. Tesevic, V.; Milosavljevic, S.; Vajs, V.; Dordevic, I.; Sokovic, M.; Lavadinovic, V.; Novakovic, M., Chemical composition and antifungal activity of the essential oil of Douglas fir (Pseudosuga menziesii Mirb. Franco) from Serbia. *Journal of the Serbian Chemical Society* **2009,** 74, (10), 1035-1040.

109. Erdtman, H.; Kimland, B.; Norin, T.; Daniels, P. J. L., The Constituents of the "Pocket Resin" from Douglas Fir Pseudotsuga menziesii (Mirb.) Franco. *Acta Chem. Scand.* **1968,** 22, (3), 938-942.
110. Jirovetz, L.; Puschmann, C.; Stojanova, A.; Metodiev, S.; Buchbauer, G., Analysis of the essential oil volatiles of Douglas fir (Pseudotsuga menziesii) from Bulgaria. *Flavour and Fragrance Journal* **2000,** 15, (6), 434-437.
111. Wagner, M. R.; Clancy, K. M.; Tinus, R. W., Maturational variation in needle essential oils from Pseudotsuga-menziesii, Abies-concolor and Picea-engelmannii. *Phytochemistry* **1989,** 28, (3), 765-770.
112. Sakai, T.; Maarse, H.; Kepner, R. E.; Jennings, W. G.; Longhurs.Wm, Volatile components of Douglas fir needles. *J Agr Food Chem* **1967,** 15, (6), 1070-&.
113. Kebbi-Benkeder, Z.; Colin, F.; Dumarcay, S.; Gerardin, P., Quantification and characterization of knotwood extractives of 12 European softwood and hardwood species. *Annals of Forest Science* **2015,** 72, (2), 277-284.
114. Holmbom, T.; Reunanen, M.; Fardim, P., Composition of callus resin of Norway spruce, Scots pine, European larch and Douglas fir. *Holzforschung* **2008,** 62, (4), 417-422.
115. Conner, A. H.; Foster, D. O., Triterpenes from Douglas fir sapwood. *Phytochem* **1981,** 20, (11), 2543-2546.
116. Laver, M. L.; Fang, H. H.-L.; Aft, H., The n-Hexane-Soluble Components of *Pseudotsuga Menziesii* Bark. *Phytochemical Reports* **1971,** 10, 3292-3294.
117. Fischer, F.; Koch, H.; Borchers, B.; Hontsch, R.; Pruzina, K. D., Preparation and use of phytosterols from wood. *Pharm* **1981,** 36, (7), 456-462.
118. De Smet, E.; Mensink, R. P.; Plat, J., Effects of plant sterols and stanols on intestinal cholesterol metabolism: Suggested mechanisms from past to present. *Molecular Nutrition & Food Research* **2012,** 56, (7), 1058-1072.
119. Lehtinen, K. J.; Mattsson, K.; Tana, J.; Engstrom, C.; Lerche, O.; Hemming, J., Effects of wood-related sterols on the reproduction, egg survival, and offspring of brown trout (Salmo trutta lacustris L.). *Ecotoxicology and Environmental Safety* **1999,** 42, (1), 40-49.
120. Denton, T. E.; Howell, W. M.; Allison, J. J.; McCollum, J.; Marks, B., Masculinization of Female Mosquitofish By Exposure to Plant Sterols and Mycobacterium smegmatis. *Bull. Environ. Contam. Toxicol.* **1985,** 35, (5), 627-632.
121. Heinonen, S.; Nurmi, T.; Liukkonen, K.; Poutanen, K.; Wahala, K.; Deyama, T.; Nishibe, S.; Adlercreutz, H., In vitro metabolism of plant lignans: New precursors of mammalian lignans enterolactone and enterodiol. *Journal of Agricultural and Food Chemistry* **2001,** 49, (7), 3178-3186.
122. Kong, L. S.; Abrams, S. R.; Owen, S. J.; Van Niejenhuis, A.; Von Aderkas, P., Dynamic changes in concentrations of auxin, cytokinin, ABA and selected metabolites in multiple genotypes of Douglas-fir (Pseudotsuga menziesii) during a growing season. *Tree Physiology* **2009,** 29, (2), 183-190.
123. Kong, L. S.; von Aderkas, P.; Owen, S. J.; Jaquish, B.; Woods, J.; Abrams, S. R., Effects of stem girdling on cone yield and endogenous phytohormones and metabolites in developing long shoots of Douglas-fir (Pseudotsuga menziesii). *New Forests* **2012,** 43, (4), 491-503.
124. Redemann Extracting 4-p-Tolylvaleric Acid from Douglas Fir. 1971.
125. Graham, H. M.; Kurth, E. F., Constituents of Extractives from Douglas Fir. *Ind. Eng. Chem.* **1949,** 41, (2), 409-414.

126. Porter, L. J.; Hrstich, L. N.; Chan, B. G., The conversion of procyanidins and prodelphinidins to cyanidin and delphinidin. *Phytochemistry* **1986,** 25, (1), 223-230.
127. Folin, O.; Ciocalteu, V., On tyrosine and tryptophane determinations in proteins. *Journal of Biological Chemistry* **1927,** 73, (2), 627-650.
128. Kurth, E. F., The Composition of the Wax in Douglas-fir Bark. **1950,** J. Am. Chem. Soc., (72(4)), 1685-1686.
129. Abraham, M. H.; Acree, W. E., On the solubility of quercetin. *J. Mol. Liq.* **2014,** 197, 157-159.
130. Auriol, D.; Nalin, R.; Robe, P.; Lefevre, F. Water soluble and activable phenolics derivatives with dermocosmetic and therapeutic applications and process for preparing said derivatives. 2007.
131. Zou, J. P.; Cates, R. G., Foliage constituents of Douglas-fir (Pseudotsuga-menziesii (mirb) Franco (pinaceae)) - their seasonal-variation and potential role in Douglas-fir resistance and silviculture management. *J Chem Ecol* **1995,** 21, (4), 387-402.
132. Hai-Loong, F. H. Douglas-fir bark: n-hexane soluble and volatile materials. Oregon State Universtiy, 1974.
133. Zhang, C.; Zhu, J. Y.; Gleisner, R.; Sessions, J., Fractionation of Forest Residues of Douglas-fir for Fermentable Sugar Production by SPORL Pretreatment. *Bioenerg. Res.* **2012,** 5, (4), 978-988.
134. Leu, S. Y.; Zhu, J. Y.; Gleisner, R.; Sessions, J.; Marrs, G., Robust enzymatic saccharification of a Douglas-fir forest harvest residue by SPORL. *Biomass Bioenerg.* **2013,** 59, 393-401.
135. Hergert, H. L.; Goldschmid, O., Biogenesis of Heartwood and Bark Constituents. I. A New Taxifolin Glucoside. *J. Org. Chem.* **1958,** 23, (5), 700-704.
136. Back, E. L., The Location and Morphology of Resin Components in the Wood. In *Pitch Control, Wood Resin and Deresination*, TAPPI PRESS: Atlanta, 2000; pp 1-35.
137. Gao, J.; Anderson, D.; Levie, B., Saccharification of recalcitrant biomass and integration options for lignocellulosic sugars from Catchlight Energy's sugar process (CLE Sugar). *Biotechnol. Biofuels* **2013,** 6, 7.
138. Dziedzic, J. A.; McDonald, A. G., A comparative survey of proteins from recalcitrant tissues of a non-model gymnosperm, Douglas-fir. *Electrophor.* **2012,** 33, (7), 1102-1112.
139. Robinson, J.; Keating, J. D.; Mansfield, S. D.; Saddler, J. N., The fermentability of concentrated softwood-derived hemicellulose fractions with and without supplemental cellulose hydrolysates. *Enzym. Microb. Technol.* **2003,** 33, (6), 757-765.
140. Stromvall, A. M.; Petersson, G., Terpenes Emitted to Air from TMP and Sulphite Pulp Mills. *Holzforschung* **1992,** 46, (2), 99-102.
141. Auriol, D.; Nalin, R.; Robe, P.; Lefevre, F. Water soluble and activable phenolics derivatives with dermocosmetic and therapeutic applications and process for preparing said derivatives. 2007.
142. Mota, F. L.; Queimada, A. J.; Pinho, S. P.; Macedo, E. A., Aqueous Solubility of Some Natural Phenolic Compounds. *Ind. Eng. Chem. Res.* **2008,** 47, (15), 5182-5189.
143. Mahmood-Khan, Z.; Hall, E. R., Occurrence and removal of plant sterols in pulp and paper mill effluents. *J. Environ. Eng. Sci.* **2003,** 2, (1), 17-26.
144. Ekman, R.; Holmbom, B., The Chemistry of Wood Resin. In *Pitch Control, Wood Resin and Deresination*, Back, E. L.; Allen, L. H., Eds. TAPPI PRESS: Atlanta, G.A., 2000; pp 37-76.

145. Kumar, L.; Arantes, V.; Chandra, R.; Saddler, J., The lignin present in steam pretreated softwood binds enzymes and limits cellulose accessibility. *Bioresource Technology* **2012,** 103, (1), 201-208.
146. Petersen, J. C.; Hill, N. S., Enzyme Inhibition by Sericea Lespedeza Tannins and the Use of Supplements to Restore Activity. *Crop. Sci.* **1991,** 31, (May-June), 827-831.
147. Wang, Z. J.; Lan, T. Q.; Zhu, J. Y., Lignosulfonate and elevated pH can enhance enzymatic saccharification of lignocelluloses. *Biotechnology for Biofuels* **2013,** 6, 10.
148. Sithole, B.; Shirin, S.; Zhang, X.; Lapierre, L.; Pimentel, J.; Paice, M., Deresination optons in sulphite pulping *Bioresour* **2010,** 5, (1), 187-205.
149. Nwaneshiudu, I. C.; Schwartz, D. T., Rational design of polymer-based absorbents: application to the fermentation inhibitor furfural. *Biotechnology for Biofuels* **2015,** 8, 8.
150. Willfor, S.; Sundberg, A.; Hemming, J.; Holmbom, B., Polysaccharides in some industrially important softwood species. *Wood Sci. Technol.* **2005,** 39, (4), 245-258.
151. Medve, J.; Karlsson, J.; Lee, D.; Tjerneld, F., Hydrolysis of microcrystalline cellulose by cellobiohydrolase I and endoglucanase II from Trichoderma reesei: adsorption, sugar production pattern, and synergism of the enzymes. *Biotechnol and Bioeng* **1998,** 59, (5), 621-634.
152. Ximenes, E.; Kim, Y.; Mosier, N.; Dien, B.; Ladisch, M., Deactivation of cellulases by phenols. *Enzyme and Microbial Technology* **2011,** 48, (1), 54-60.
153. Boussaid, A. L.; Esteghlalian, A. R.; Gregg, D. J.; Lee, K. H.; Saddler, J. N., Steam pretreatment of Douglas-fir wood chips - Can conditions for optimum hemicellulose recovery still provide adequate access for efficient enzymatic hydrolysis? *Appl Biochem and Biotech* **2000,** 84-6, 693-705.
154. McAllister, T. A.; Bae, H. D.; Yanke, L. J.; Cheng, K. J.; Muir, A., Effect of condensed tannins from Birdsfoot-Trefoil on endoglucanase activity and the digestion of cellulose filter-paper by ruminal fungi. *Can J Microbiol* **1994,** 40, (4), 298-305.
155. Karlsson, J.; Momcilovic, D.; Wittgren, B.; Schulein, M.; Tjerneld, F.; Brinkmalm, G., Enzymatic degradation of carboxymethyl cellulose hydrolyzed by the endoglucanases Cel5A, Cel7B, and Cel45A from Humicola insolens and Cel7B, Cel12A and Cel45Acore from Trichoderma reesei. *Biopolymers* **2002,** 63, (1), 32-40.
156. Claeyssens, M.; Van tilbeurgh, H.; Tomme, P.; Wood, T. M.; McRae, S. I., Fungal cellulase systems. Comparison of the specificities of the cellobiohydrolases isolated from *Penicillium pinophilum* and *Trichoderma reesei*. *Biochem J* **1989,** 261, (3), 819-825.
157. Biely, P.; Vrsanska, M.; Claeyssens, M., The endo-1,4-beta-glucanase I from Trichoderma reesei. Action on beta-1,4-oligomers and polymers derived from D-glucose and D-xylose. *Eur J Biochem* **1991,** 200, (1), 157-163.
158. Fogler, S. H., Uncompetitive inhibition. In *Essentials of Chemical Reaction Engineering*, Prentice Hall: 2011; pp 367-377.
159. Prieur, C.; Rigaud, J.; Cheynier, V.; Moutounet, M., Oligomeric and polymeric procyanidins from grape seeds. *Phytochemistry* **1994,** 36, (3), 781-784.
160. Zhang, C.; Zhu, J. Y.; Gleisner, R.; Sessions, J., Fractionation of forest residues of Douglas-fir for fermentable sugar production by SPORL pretreatment. *Bioenerg Res* **2012,** 5, (4), 978-988.
161. Lorrain, B.; Ky, I.; Pechamat, L.; Teissedre, P. L., Evolution of analysis of polyhenols from grapes, wines, and extracts. *Molecules* **2013,** 18, (1), 1076-1100.

162. Auriol, D.; Nalin, R.; Robe, P.; Lefevre, F., Water soluble and activable phenolics derivatives with dermocosmetic and therapeutic applications and process for preparing said derivatives. In Google Patents: 2009.
163. Fichan, I.; Larroche, C.; Gros, J. B., Water solubility, vapor pressure, and activity coefficients of terpenes and terpenoids. *J Chem Eng Data* **1999,** 44, (1), 56-62.
164. Phillips, M. L.; White, R. L., Dependence of chromatogram peak areas obtained by curve-fitting on the choice of peak shape function. *J Chromatogr Sci* **1997,** 35, (2), 75-81.
165. Korczak, M. K.; Koziarski, S.; Komorowska, B., Anaerobic treatment of pulp mill effluents. *Water Sci Technol* **1991,** 24, (7), 203-206.
166. Bhat, T. K.; Singh, B.; Sharma, O. P., Microbial degradation of tannins - A current perspective. *Biodegradation* **1998,** 9, (5), 343-357.
167. Bommarius, A. S.; Sohn, M.; Kang, Y. Z.; Lee, J. H.; Realff, M. J., Protein engineering of cellulases. *Curr Opin Biotech* **2014,** 29, 139-145.
168. Damborsky, J.; Brezovsky, J., Computational tools for designing and engineering enzymes. *Curr Opin Chem Biol* **2014,** 19, 8-16.
169. Biedermannova, L.; Prokop, Z.; Gora, A.; Chovancova, E.; Kovacs, M.; Damborsky, J.; Wade, R. C., A single mutation in a tunnel to the active site changes the mechanism and kinetics of product release in haloalkane dehalogenase LinB. *J Biol Chem* **2012,** 287, (34), 29062-29074.
170. Sprenger, K. G.; Choudhury, A.; Kaar, J. L.; Pfaendtner, J., Lytic polysaccharide monooxygenases ScLPMO10B and ScLPMO10C are stable in ionic liquids as determined by molecular simulations. *J Phys Chem B* **2016,** 120, (16), 3863-3872.
171. Pratter, S. M.; Konstantinovics, C.; Cristiana, M. L.; Di Giuro, C. M. L.; Leitner, E.; Kumar, D.; de Visser, S. P.; Grogan, G.; Straganz, G. D., Inversion of enantioselectivity of a mononuclear non-heme iron(II)-dependent hydroxylase by tuning the interplay of metal-center geometry and protein structure. *Angew Chem Int Edit* **2013,** 52, (37), 9677-9681.
172. Nakamura, A.; Tsukada, T.; Auer, S.; Furuta, T.; Wada, M.; Koivula, A.; Igarashi, K.; Samejima, M., The tryptophan residue at the active site tunnel entrance of *Trichoderma reesei* cellobiohydrolase Cel7A is important for initiation of degradation of crystalline cellulose. *J Biol Chem* **2013,** 288, (19), 13503-13510.
173. Payne, C. M.; Resch, M. G.; Chen, L. Q.; Crowley, M. F.; Himmel, M. E.; Taylor, L. E.; Sandgren, M.; Stahlberg, J.; Stals, I.; Tan, Z. P.; Beckham, G. T., Glycosylated linkers in multimodular lignocellulose-degrading enzymes dynamically bind to cellulose. *P Natl Acad Sci USA* **2013,** 110, (36), 14646-14651.
174. Gritzali, M.; Brown, R. D. J., The cellulase system of Trichoderma: Relationships between purified extracellular enzymes from induced or cellulose-grown cells. In *Hydrolysis of Cellulose: Mechanisms of Enzymatic and Acid Catalysis*, 1979; Vol. 181, pp 237-260.
175. Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J., PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res* **2005,** 33, W363-W367.
176. Duhovny, D.; Nussinov, R.; Wolfson, H. J., Efficient unbound docking of rigid molecules. *Lect N Bioinformat* **2002,** 2452, 185-200.
177. Dennington, R.; Keith, T. A.; Millam, J. M. *Gaussview 6*, Semichem Inc: Shawnee Mission, KS, 2016.
178. Schuttelkopf, A. W.; van Aalten, D. M. F., PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D* **2004,** 60, 1355-1363.
179. Kraulis, P. J.; Clore, G. M.; Nilges, M.; Jones, T. A.; Pettersson, G.; Knowles, J.; Gronenborn, A. M., Determination of the three-dimensional solution structure of the C-terminal

domain of the cellobiohydrolase I from *Trichoderma reesei*. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. *Biochemistry-US* **1989,** 28, (18), 7241-7257.
180. Divne, C.; Stahlberg, J.; Reinikainen, T.; Ruohonen, L.; Pettersson, G.; Knowles, J. K. C.; Teeri, T. T.; Jones, T. A., The 3-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science* **1994,** 265, (5171), 524-528.
181. Humphrey, W.; Dalke, A.; Schulten, K., VMD - Visual molecular dynamics. *J Mol Graph Model* **1996,** 14, (1), 33-38.
182. Chen, Y. C., Beware of docking! *Trends Pharmacol Sci* **2015,** 36, (2), 78-95.
183. Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E., GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013,** 29, (7), 845-854.
184. Berendsen, H. J. C.; Postma, J. P. M.; van Gunstere, W. F.; Hermans, J., Interaction models for water in relation to protein hydration. In *Intermolecular Forces*, 1981; pp 331-342.
185. van Gunsteren, W. F.; Berendsen, H. J. C., GROMOS-87 Manual BIOMOS BV. In AG Groningen, Nijenborgh: 1987.
186. Gueroux, M.; Pinaud-Szlosek, M.; Fouquet, E.; De Freitas, V.; Laguerre, M.; Pianet, I., How wine polyphenols can fight Alzheimer disease progression: towards a molecular explanation. *Tetrahedron* **2015,** 71, (20), 3163-3170.
187. Goncalves, R.; Mateus, N.; Pianet, I.; Laguerre, M.; de Freitas, V., Mechanisms of tannin-induced trypsin inhibition: A molecular approach. *Langmuir* **2011,** 27, (21), 13122-13129.
188. Cala, O.; Pinaud, N.; Simon, C.; Fouquet, E.; Laguerre, M.; Dufourc, E. J.; Pianet, I., NMR and molecular modeling of wine tannins binding to saliva proteins: revisiting astringency from molecular and colloidal prospects. *Faseb Journal* **2010,** 24, (11), 4281-4290.
189. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *J Chem Phys* **2007,** 126, (1), 7.
190. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *J Chem Phys* **1984,** 81, (8), 3684-3690.
191. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys* **1981,** 52, (12), 7182-7190.
192. Zucker, W. V., Tannins: Does structure determine function? An ecological perspective. *Am Nat* **1983,** 121, (3), 335-365.
193. Haslam, E., Polyphenol-protein interactions. *Biochem J* **1974,** 139, (1), 285-288.
194. McManus, J. P.; Davis, K. G.; Lilley, T. H.; Haslam, E., The association of proteins with polyphenols. *J Chem Soc Chem Comm* **1981**, (7), 309-311.
195. McQuarrie, D. A., *Statistical Mechanics*. Harper & Row: New York, 1973.
196. Frishman, D.; Argos, P., Knowledge-based protein secondary structure assignment. *Proteins* **1995,** 23, (4), 566-579.
197. Bras, N. F.; Goncalves, R.; Fernandes, P. A.; Mateus, N.; Ramos, M. J.; de Freitas, V., Understanding the binding of procyanidins to pancreatic elastase by experimental and computational methods. *Biochemistry* **2010,** 49, (25), 5097-5108.
198. Smith, J.; Sprenger, K. G.; Liao, R.; Joseph, A.; Nance, E.; Pfaendtner, J., Determining dominant driving forces affecting controlled protein release from polymeric nanoparticles. *Biointerphases* **2017,** 12, (2), 02D412-1 - 9.

199. Smith, J.; Sprenger, K. G.; Liao, R.; Joseph, A.; Nance, E.; Pfaendtner, J., Determining dominant driving forces affecting controlled protein release from polymeric nanoparticles. *Biointerphases* **2017,** 12, (2), 9.
200. Charlton, A. J.; Baxter, N. J.; Khan, M. L.; Moir, A. J. G.; Haslam, E.; Davies, A. P.; Williamson, M. P., Polyphenol/peptide binding and precipitation. *J Agr Food Chem* **2002,** 50, (6), 1593-1601.
201. Hagerman, A. E.; Butler, L. G., The specificity of proanthocyanidin-protein interactions. *J Biol Chem* **1981,** 256, (9), 4494-4497.
202. Adamczyk, B.; Adamczyk, S.; Smolander, A.; Kitunen, V., Tannic acid and Norway spruce condensed tannins can precipitate various organic nitrogen compounds. *Soil Biol Biochem* **2011,** 43, (3), 628-637.
203. Luck, G.; Liao, H.; Murray, N. J.; Grimmer, H. R.; Warminski, E. E.; Williamson, M. P.; Lilley, T. H.; Haslam, E., Polyphenols, astringency, and proline-rich proteins. *Phytochemistry* **1994,** 37, (2), 357-371.
204. Baxter, N. J.; Lilley, T. H.; Haslam, E.; Williamson, M. P., Multiple interactions between polyphenols and a salivary proline-rich protein repeat result in complexation and precipitation. *Biochemistry* **1997,** 36, (18), 5566-5577.
205. Bodenheimer, A. M.; Meilleur, F., Crystal structures of wild-type Trichoderma reesei Cel7A catalytic domain in open and closed states. *Febs Lett* **2016,** 590, (23), 4429-4438.
206. Divne, C.; Stahlberg, J.; Teeri, T. T.; Jones, T. A., High-resolution crystal structures reveal how a cellulose chain is bound in the 50 angstrom long tunnel of cellobiohydrolase I from Trichoderma reesei. *J Mol Biol* **1998,** 275, (2), 309-325.
207. Granum, D. M.; Schutt, T. C.; Maupin, C. M., Computational evaluation of the dynamic fluctuations of peripheral loops enclosing the catalytic tunnel of a family 7 cellobiohydrolase. *J Phys Chem B* **2014,** 118, (20), 5340-5349.
208. EIA, Petroleum & other liquids: refinery yields (U.S.). In eia.gov, 2017.
209. Eckert, C.; Liotta, C.; Ragauskas, A.; Hallett, J.; Kitchens, C.; Hill, E.; Draucker, L., Tunable solvents for fine chemicals from the biorefinery. *Green Chemistry* **2007,** 9, (6), 545-548.
210. Chung, H. Y.; Washburn, N. R., Chemistry of lignin-based materials. *Green Materials* **2013,** 1, (3), 137-160.
211. Zhang, Y. H. P., Reviving the carbohydrate economy via multi-product lignocellulose biorefineries. *J Ind Microbiol Biot* **2008,** 35, (5), 367-375.
212. Gyles, C. L.; Carlberg, J. G.; Gustafson, J.; Davlut, D. A.; Jones, P. J. H., Economic valuation of the potential health benefits from foods enriched with plant sterols in Canada. *Food Nutr Res* **2010,** 54, 7.
213. Aden, A.; Ruth, M.; Ibsen, K.; Jechura, J.; Neeves, K.; Sheehan, J.; Wallace, B.; Montague, L.; Slayton, A.; Lukas, J. *Lignocellulosic Biomass to Ethanol Process Design and Economics Utilizing Co-Current Dilute Acid Prehydrolysis and Enzymatic Hydrolysis for Corn Stover*; National Renewable Energy Laboratory: Golden Colorado, 2002; pp 1-154.
214. Razmara, R. S.; Daneshfar, A.; Sahraei, R., Solubility of Quercetin in Water plus Methanol and Water plus Ethanol from (292.8 to 333.8) K. *Journal of Chemical and Engineering Data* **2010,** 55, (9), 3934-3936.
215. Song, X. F.; Zhang, S. J.; Wang, Y. F.; Li, J. W.; He, C. Y.; Yao, L. S., A kinetic study of Trichoderma reesei Cel7B catalyzed cellulose hydrolysis. *Enzym Microb Technol* **2016,** 87-88, 9-16.

216. Kolbe, J.; Kubicek, C. P., Quantification and identification of the main components of the Trichoderma cellulase complex with monoclonal-antibodies using an enzyme-linked-immunosorbent-assay (ELISA). *Appl Microb Biotech* **1990,** 34, (1), 26-30.
217. Kleywegt, G. J.; Zou, J. Y.; Divne, C.; Davies, G. J.; Sinning, I.; Stahlberg, J.; Reinikainen, T.; Srisodsuk, M.; Teeri, T. T.; Jones, T. A., The crystal structure of the catalytic core domain of endoglucanase I from Trichoderma reesei at 3.6 angstrom resolution, and a comparison with related enzymes. *J Molec Biol* **1997,** 272, (3), 383-397.
218. Dienes, D.; Borjesson, J.; Stalbrand, H.; Reczey, K., Production of Trichoderma reesei Cel7B and its catalytic core on glucose medium and its application for the treatment of secondary fibers. *Process Biochem* **2006,** 41, (9), 2092-2096.