Enzymes and feedstocks for sustainable biomass utilisation

by

Yaseen Mottiar

A thesis submitted in conformity with the requirements for the degree of Master of Science

Cell & Systems Biology
University of Toronto

© Copyright by Yaseen Mottiar 2012
ENZYMES AND FEEDSTOCKS FOR SUSTAINABLE BIOMASS UTILISATION

Yaseen Mottiar

Master of Science

Cell & Systems Biology
University of Toronto

2012

ABSTRACT

Modern biorefineries provide a framework for the sustainable conversion of biomass to biofuels and biochemicals. In light of the recalcitrance of lignin in woody feedstocks, the native shrub eastern leatherwood is proposed as a model hypolignified species. Xylem tissue of this low-lignin plant contained syringyl-rich lignin that was more easily hydrolysed and did not appear to be localised in the middle lamellae. Also, leatherwood cellulose was less crystalline and the xylan was highly acetylated. While viable low-lignin plants will enable the sustainable utilisation of woody feedstocks, high-value bioproducts are needed to economise future biorefineries. The carbohydrate oxidoreductases galactose oxidase and glucoooligosaccharide oxidase were studied for use in the oxidation and derivatisation of plant-derived polysaccharides for the production of such high-value bioproducts. The carbohydrate-binding module of galactose oxidase was necessary for recombinant protein production. Also, a mutant library of glucoooligosaccharide oxidase variants was produced to generate enzymes with novel activity.
Pamuditha Silva assisted in cloning the CBM sequences and with the random mutagenesis work. Dr. Dragica Jeremic helped with the leatherwood project and in particular the HPLC analysis. Finally, Dr. Harry Brumer generously supplied the codon-optimised galactose oxidase gene.

Thanks to Professor Emma Master for all her kind support and expert guidance. Thanks also to all the past and present members of the Master lab who have helped make this a rewarding experience that I will not soon forget. In particular, thanks to Alex Tsai, Matt Doering, Mohammed Sherif, Dr. Robyn Goacher and Dr. Thu Vuong who have helped me in countless ways over the past two years. I am also very grateful to all the members of BioZone, especially Susie Susilawati.

To all the friends I gained in Toronto, especially Marie Manchester and Ilan Adler, I offer many thanks for your support and encouragement. Finally to my parents and the rest of my family who have given me endless love and support, I could never thank you enough…

“from the Northwest Passage,

to the Great Divide,

everybody’s looking for the other side…”

Josh Ritter, 2001
# TABLE OF CONTENTS

Acknowledgments ..................................................................................................................... iii  
Table of Contents ...................................................................................................................... iv  
List of Tables ............................................................................................................................. vi  
List of Figures .......................................................................................................................... vii  
List of Abbreviations ............................................................................................................... viii  
Chapter 1: Introduction .............................................................................................................. 1  
  1.1 Motivation ................................................................................................................ 1  
  1.2 Research Objectives ................................................................................................. 2  
  1.2.2 Low-lignin Feedstocks.................................................................................. 2  
  1.2.1 Carbohydrate Oxidases ................................................................................. 3  
Chapter 2: Literature Review ..................................................................................................... 4  
  2.1 Constituents of Lignocellulose................................................................................. 4  
  2.1.1 Cellulose ....................................................................................................... 4  
  2.1.2 Hemicellulose ............................................................................................... 6  
  2.1.3 Lignin ............................................................................................................ 7  
  2.2 The Quest for Low-lignin Feedstocks................................................................. 10  
  2.2.1 The Case for Low-lignin Plants .................................................................. 10  
  2.2.2 Natural Variability in Lignin ...................................................................... 11  
  2.2.3 Lignin Engineering ..................................................................................... 12  
  2.2.4 Towards Viable Low-lignin Plants ............................................................. 14  
  2.3 Enzymatic Modification of Plant Polysaccharides................................................. 15  
  2.3.1 Carbohydrate-active Enzymes .................................................................... 15  
  2.3.2 Carbohydrates Oxidases for Carbohydrate Derivitisation .......................... 15  
  2.3.3 Galactose Oxidase....................................................................................... 17  
  2.3.4 Glucooligosaccharide Oxidase.................................................................... 19  
  2.3.6 Carbohydrate-binding Modules .................................................................. 21  
Chapter 3: Eastern Leatherwood is a Low-lignin Plant with a Story to Tell ........................... 23  
  3.1 Introduction ............................................................................................................ 23  
  3.2 Materials and Methods ........................................................................................... 24  
  3.2.1 Wood Samples ............................................................................................ 24  
  3.2.2 Lignin Content ............................................................................................ 25  
  3.2.3 Fourier-transform Infrared Spectroscopy ................................................... 25  
  3.2.4 Time-of-flight Secondary Ion Mass Spectrometry ..................................... 26  
  3.2.5 X-ray Diffraction ........................................................................................ 27  
  3.4 Results and Discussion........................................................................................... 27  
  3.4.1 Lignin Content ............................................................................................ 28  
  3.4.2 Evaluation of Lignocellulose Composition by FTIR .................................... 30  
  3.4.3 Evaluation of Lignocellulose Composition by ToF-SIMS .............................. 34  
  3.4.4 Estimation of Lignin Composition.............................................................. 36  
  3.4.5 Evaluation of Lignin Localisation by ToF-SIMS ....................................... 37  
  3.4.6 Cellulose Crystallinity ................................................................................ 38  
  3.4.7 Hypolignification in Leatherwood.............................................................. 40
Chapter 4: Carbohydrate Oxidases to Enable the Derivatisation of Wood Carbohydrates

4.1 Swapping the GalOx CBM

  4.1.1 Introduction
  4.1.2 Design of GalOx Variants
  4.1.3 Construction of GalOx Variants
  4.1.4 Transformation and Expression in P. pastoris
  4.1.5 Screening Transformants for GalOx activity
  4.1.6 The Native CBM is Necessary for GalOx Production

4.2 Production of GalOx

  4.2.1 Introduction
  4.2.2 Recombinant Production of Wild-type GalOx

4.3 Attempts to Engineer a Cellulose-6-oxidase

  4.3.1 Introduction
  4.3.2 Design and Production of GalOx Variants
  4.3.3 Screening Transformants and Understanding Activity Results

4.4 Random Mutagenesis of GOOX

  4.4.1 Introduction
  4.4.2 Error-prone Mutagenesis Strategy
  4.4.3 Production of GOOX Mutant Library
  4.4.4 Pichia Transformation for Future Screening

Chapter 5: Summary and Recommendations

  5.1 Low-lignin Feedstocks
  5.2 Carbohydrate Oxidases

References

Appendix 1: Peak List for FTIR
Appendix 2: Peak List for ToF-SIMS
Appendix 3: Holocellulose Composition Evaluated by HPLC
  A1.1 Methodology
  A1.2 Results and Discussion
Appendix 4: Cellulose Crystallinity Evaluated by FTIR
  A2.1 Methodology
  A2.1 Results and Discussion
Appendix 5: Primers and PCR Conditions
Appendix 6: Recipes
Appendix 7: Sequences
LIST OF TABLES

Table 1. Lignin content of several North American woods .......................................................... 12
Table 2. Lignin content of xylem tissues ...................................................................................... 29
Table 3. Lignin composition estimates ......................................................................................... 37
Table 4. Crystallinity of xylem tissues ........................................................................................ 40
LIST OF FIGURES

Figure 1. Red pine harvest in Algonquin Provincial Park ............................................................ 1
Figure 2. Schematic representation of cellulose ................................................................. 5
Figure 3. Some common hemicelluloses ........................................................................... 6
Figure 4. Biosynthesis pathway for monolignols .............................................................. 8
Figure 5. Galactose oxidase ................................................................................................... 17
Figure 6. Close-up view of the galactose oxidase active site .............................................. 18
Figure 7. Glucooligosaccharide oxidase ............................................................................. 20
Figure 8. Close-up view of the glucooligosaccharide oxidase active site ............................. 21
Figure 9. Watercolour of eastern leatherwood ................................................................ 23
Figure 10. PCA analysis of FTIR spectra .............................................................................. 31
Figure 11. PCA analysis of ToF-SIMS spectra .................................................................... 35
Figure 12. Peaks fitted to representative FTIR spectra of leatherwood lignin ...................... 36
Figure 13. TOF-SIMS Ion Images ......................................................................................... 38
Figure 14. X-ray diffractograms ......................................................................................... 39
Figure 15. Vector map of pPICZα ......................................................................................... 43
Figure 16. Workflow for production of GalOx variants ......................................................... 44
Figure 17. GalOx expression constructs for CBM swapping project .................................... 45
Figure 18. Triple digest of GalOx expression constructs for CBM swapping project ............. 46
Figure 19. Troubleshooting the plate-based assay for GalOx activity ................................... 47
Figure 20. Plate-based activity screen of GalOx transformants for variants #1-3 ................. 48
Figure 21. SDS-PAGE gel of culture supernatants of selected GalOx transformants ............ 49
Figure 22. Colony blot of variant #2 transformants .............................................................. 49
Figure 23. SDS-PAGE gel of *Pichia* supernatants .............................................................. 51
Figure 24. Protein desalting .................................................................................................. 52
Figure 25. Affinity-based protein purification ....................................................................... 53
Figure 26. Western blot of purified galOx ............................................................................. 54
Figure 27. Workflow for site-directed mutagenesis .............................................................. 55
Figure 28. GalOx expression constructs for cellulose-6-oxidase project .............................. 56
Figure 29. Triple digest of GalOx expression constructs for cellulose-6-oxidase project ....... 57
Figure 30. Plate-based activity screen of GalOx transformants for variants #4-8 .................. 58
Figure 31. Workflow for error-prone PCR ........................................................................... 60
LIST OF ABBREVIATIONS

4CL ......................... 4-coumarate-CoA ligase
bp.............................. base pairs
C3H.......................... coumarate 3-hydroxylase
C4H.......................... cinnamate 4-hydroxylase
CAD.......................... cinnamyl alcohol dehydrogenase
CBM.......................... carbohydrate-binding module
CCoAOMT ............. caffeoyl-CoA 3-O-methyltransferase
CCR........................ cinnamoyl-CoA reductase
CD............................ catalytic domain
COMT........................ caffeic acid/5-hydroxyferulic acid O-methyltransferase
FAD.......................... flavin adenine dinucleotide
F5H.......................... ferulate 5-hydroxylase
FTIR........................ Fourier-transform infrared spectroscopy
G lignin...................... guaiacyl lignin
GalOx........................ galactose oxidase
GGOX........................ glucooligosaccharide oxidase
H lignin...................... p-hydroxyphenyl lignin
HCT.......................... shikimate hydroxycinnamoyl transferase
HPLC........................ high-performance liquid chromatograph
PAL.......................... phenylalanine ammonia lyase
PCA.......................... principal component analysis
PCR.......................... polymerase chain reaction
S lignin...................... syringyl lignin
SDS-PAGE .................. sodium dodecyl sulphate polyacrylamide gel electrophoresis
ToF-SIMS.................. time-of-flight secondary ion mass spectrometry
XRD.......................... X-ray diffraction
CHAPTER 1: INTRODUCTION

1.1 Motivation

The sustainable forest bioproducts industry in Canada is potentially worth up to 100 billion dollars annually (Wetzel et al., 2006). Being a large northern nation, Canada possesses vast quantities of high-quality harvestable wood fibre in the form of biomass from the boreal and northern temperate forests (Figure 1). As technology improves and the traditional markets for our slow-growing harvestable forests wane, next-generation bioproducts are poised to become the pillars of the emergent bioeconomy in Canada. Harvestable biomass available for bioproducts production is estimated to be as much as 561 million tonnes per annum (Gronowska et al., 2009).

Second-generation biofuel produced from lignocellulose is being heralded as a practical and sustainable alternative to displace traditional petroleum-derived fuel. Widespread research efforts are presently underway to translate lab-scale experiments to industrial processes. While advances in downstream processing technologies will undoubtedly improve overall efficiency in the bioconversion of cellulose and hemicellulose into fuel alcohol, the removal of lignin from woody feedstocks remains a key challenge (Himmel et al., 2007). The development of low-lignin plants was originally motivated by the financial realities of the pulp and paper industry, but it is even more critical now for the growth of the nascent biofuels sector.

Amid concerns about the long-term economic feasibility of biofuels production, there is also a drive to develop high-value products from wood fibre that would economise the production of cheap fuels in biorefineries much as is the case in the petroleum industry (Ragauskas et al., 2006). It is imperative to look beyond the traditional use of woody biomass as fuel towards the myriad useful products waiting to be fashioned from the impressive and underexploited
molecular complexity contained in plant fibres. High-value bioproducts such as biopharmaceuticals and bioplastics will be required to ensure economic viability in biorefinery operations. Such advanced bioproducts could be derived from lignin, polysaccharides or the wide array of secondary metabolites in wood (Bruce and Palfreyman, 1998). While chemical routes to derivitisation and functionalisation of such molecules are possible, enzyme-based approaches are being sought to ensure sustainability in production.

I believe that novel uses for lignocellulose will preserve and promote the value of our forests and in turn encourage sustainable forest and land management. Having grown up in the Upper Ottawa Valley, I am well-aware of the innumerable consequences of the weakening forestry industry in rural Canada. Through this project, I aim to make significant incremental contributions towards the development of technologies for sustainable biomass utilisation. This project aligns well with my ultimate professional goal to establish myself at the interface of key discoveries in the biological sciences and the development of sustainable technologies for the better of humankind and our environment.

1.2 Research Objectives

1.2.2 Low-lignin Feedstocks

In light of the economic and environmental cost of separating lignin from wood polysaccharides, widespread research efforts are currently underway to engineer commercial tree genotypes with low lignin content. Whereas natural low-lignin species could inspire such efforts, I have pursued the following four research objectives in studying the xylem tissue of the native shrub eastern leatherwood:

1. To determine the lignin content.

2. To evaluate the composition of lignin.

3. To assess the distribution of lignin across the cell wall.

4. To measure the cellulose crystallinity.
Chapter 2 presents a literature review pursuant to these objectives. The research results and discussion are presented in Chapter 3. Finally, the key conclusions and recommendations are summarized in Chapter 5.

**1.2.1 Carbohydrate Oxidases**

Wood polysaccharides could be used in a wide range of next-generation bioproducts and the oxidation of hydroxyl groups could open the door to derivitisation and functionalisation. Whereas carbohydrate oxidases are unique biocatalysts capable of directing this chemistry, I have pursued the following four research objectives in enzyme engineering:

1. To explore the potential of expanding the activity of galactose oxidase on galactose-containing polysaccharides through carbohydrate-binding module swapping.

2. To evaluate the role of the native carbohydrate-binding module of galactose oxidase.

3. To explore the potential of expanding the activity of galactose oxidase to cellulose through a combination of site-directed mutagenesis and carbohydrate-binding module swapping.

4. To explore the potential of expanding the activity of glucooligosaccharide oxidase through random mutagenesis.

Chapter 2 presents a review of the relevant literature. The research results and discussion are presented in Chapter 4. Finally, the key conclusions and recommendations are summarized in Chapter 5.
CHAPTER 2: LITERATURE REVIEW

Secondary cell walls in woody plant tissues are composed of lignocellulose - a complex matrix of cellulose, hemicellulose and lignin. While this thesis takes two distinct paths – enzymes and feedstocks – the unifying theme is lignocellulosic biomass. This chapter will first provide an overview of lignocellulose before delving into subject-specific literature reviews for low-lignin plants and carbohydrate oxidases.

2.1 Constituents of Lignocellulose

2.1.1 Cellulose

Cellulose is the most abundant biopolymer on the planet and it plays an integral role in cell wall structure (Robyt, 1998). It is a polysaccharide composed of β-(1→4)-linked D-glucopyranose with cellobiose as the repeating unit. In the lignocellulosic matrix of secondary plant cell walls, cellulose exists in the form of microfibrils which are strongly associated with hemicellulose and lignin. It is a remarkably simple biomolecule with myriad uses ranging from food for cellulolytic bacteria and fungi to the paper on which this thesis is/may be printed.

Cellulose is synthesized by transmembrane rosette complexes each composed of six individual rosette units (Doblin et al., 2002). Rosette subunits themselves each comprise six cellulose synthase enzymes. In this manner, each rosette complex generates cellulose as 36-parallel-chain units called elementary fibrils. Cellulose synthases add UDP-glucose from the cytoplasm to the growing glucan chain which is simultaneously released into the cell wall (Delmer and Amor, 1995). Since rosettes cluster on the cell membrane, many elementary fibrils may initially agglomerate as macrofibrils (Ding and Himmel, 2006). Shortly thereafter, macrofibrils disintegrate into microfibrils upon the integration of hemicellulose and lignin. A depiction of the ultrastructure of cellulose in this context is presented in Figure 2.

The remarkable strength properties of wood result from the close integration of cellulose, lignin and hemicellulose and also from the tightly-packed nature of parallel cellulose chains (Klemm et al., 1998a). The interior chains are the most crystalline and crystallinity likely
decreases radially resulting from different cellulose synthase activities and the local cell wall environment into which cellulose strands are released (Doblin et al., 2002). While the exterior chains of cellulose microfibrils are amorphous and integrated with hemicellulose, the microfibril interior is highly crystalline due to extensive hydrogen bonding. Of the two different hydrogen-bonding patterns found in native cellulose, cellulose Iα and Iβ, the second allomorph is more prominent in woody plants (Atalla and VanderHart, 1984).

Cellulose is interwoven into the history of humanity and has been used in textiles and paper products for more than 5000 years (Hon, 1994). Besides its stalwart use in such traditional materials as cotton fabrics and newsprint, cellulose has found utility in a great number of advanced products from cellophane films to fillers in medications (Klemm et al., 2005). Many modern applications require that cellulose be chemically derivatised to modify its physical properties or to impart functionality. For example, nanocrystalline cellulose can be used in bioplastic composites if it is first derivatised to reduce hydrophilicity (Hubbe et al., 2008; Hasani et al., 2008). In addition to advanced bioproducts, cellulose is increasingly being pursued as a
source of fuel alcohols. Once the lignin and other interfering components are removed, hydrolysis of cellulose provides glucose which fermentative bacteria transform into ethanol (Carroll and Somerville, 2009).

### 2.1.2 Hemicellulose

Like cellulose, hemicelluloses are polysaccharides composed of $\beta-(1\rightarrow4)$-linked backbones. But unlike cellulose, hemicelluloses can contain backbone sugars other than glucose (Robyt, 1998). In addition, hemicelluloses are shorter and may contain branching sugars. Common hemicelluloses in woody plants include xylan, arabinoxylan, galactomannan, glucomannan and galactoglucomannan (Figure 3). Other hemicelluloses that are less important in secondary cell walls include xyloglucan and $\beta$-glucan. Although it is a minor constituent of secondary walls, pectin is often confused with hemicellulose. For example, galactans, arabinans and arabinogalactans are variously classified as pectins or hemicelluloses (Scheller and Ulvskov, 2010). Pectins are complex and highly branched polysaccharides which are more easily extracted than hemicellulose. While hemicelluloses differ in composition and structure, all support cell wall integrity through interactions with cellulose and lignin (Keegstra, 2010).

Xylans are the most abundant hemicellulose in broadleaf species typically comprising 15-25% compared to mannans which constitute less than 5% (Timmel, 1957). On the other hand, mannans are the more prevalent hemicellulose in conifers comprising 5-15% while xylans represent 5-10%.

Xylans are composed of a $\beta-(1\rightarrow4)$-linked D-xylopyranose backbone with $\alpha-(1\rightarrow2)$-linked D-glucuronosyl and 4-O-methyl D-glucuronosyl units (Pettersen, 1984; Ebringerová et al., 2005). Most
xylans also contain O-2 or O-3-linked acetyl groups. In some species – especially conifers, xylan contains both 4-O-methyl D-glucuronosyl and O-3-linked L-arabinofuranosyl units and is called arbinglucuronoxylan. Mannans contain β-(1→4)-linked backbones of D-mannopyranose and D-glucopyranose which may be acetylated at O-2 or O-3 and, in conifers, they may contain α-(1→6)-linked D-galactopyranose (Pettersen, 1984). Some common types of mannan include glucomannan, galactomannan, and galactoglucomannan (Ebringerová et al., 2005).

The biosynthesis of hemicellulose occurs in the Golgi apparatus and is largely driven by glycosyltransferases (Sandhu et al., 2009). Nucleotide sugars are first imported from the cytoplasm into the Golgi. Subsequently, glycosyltransferases build hemicellulose chains by forming glycosidic bonds between the growing chain and activated nucleotide sugars. Glycosyltransferases constitute a large class of enzymes that are described by over 90 different families with specific and varied substrate preferences (Scheller and Ulvskov, 2010; CAZy Database). As such, diversity in hemicellulose composition across cell types and species is due largely to different glycosyltransferase activities and the diversity in nucleotide sugar pools. After synthesis, hemicellulose chains are exported to the cytoplasm in vesicles and are subsequently released to the cell wall where hemicellulose deposition occurs.

Hemicelluloses have found use in a wide array of products from low-calorific food additives to advanced bioproducts such as biodegradable films and biocompatible gels for drug delivery (Hansen and Plackett, 2008). Its highly branched structures are particularly useful as binding agents. But as with cellulose, chemical derivatisation of hemicellulose could facilitate its use in countless other advanced bioproducts (Ebringerová et al., 2005; Heinze and Daus, 2011).

2.1.3 Lignin

Lignin is a highly cross-linked polyphenolic macromolecule found in secondary plant cell walls. It is produced by the dehydrogenative polymerisation of three constituent monolignols: p-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol (Sarkanen and Ludwig, 1971). Once incorporated into a lignin chain, these components are known as p-hydroxyphenyl (H lignin), guaiacyl (G lignin) and syringyl (S lignin) units. Besides these core 4-hydroxyphenylpropanoid
constituents, some other minor components present in lignin include the following:
hydroxycinnamyl aldehydes, acetates, \( p \)-coumarates, \( p \)-hydroxybenzoates, and tyramine ferulate (Boerjan et al., 2003). Lignin is a critical element of secondary cell walls providing structural integrity, water retention and defence against microbial attack (Sarkanen and Ludwig, 1971).

Monolignols are synthesized in the cytoplasm from phenylalanine through the phenylpropanoid biosynthesis pathway (Vanholme et al., 2010). Phenylalanine is shuttled from the shikimate biosynthesis pathway and the first dedicated step in phenylpropanoid synthesis is the deamination of phenylalanine to produce cinnamic acid. What follows is the successive hydroxylation of the aromatic ring and subsequent O-methylation and reduction of the propenyl side chain (Freudenberg and Neish, 1968; Boerjan et al., 2003). As shown in Figure 4, the production of monolignols from phenylalanine is mediated by at least ten different enzymes: phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (C4H), 4-coumarate-CoA ligase

![Figure 4. Biosynthesis pathway for monolignols from L-phenylalanine. Carbon atom numbering is shown for H-lignin. Adapted from Vanholme et al., 2010. The carbon numbering is shown in red for H-lignin.](image-url)
(4CL), shikimate hydroxycinnamoyl transferase (HCT), coumarate 3-hydroxylase (C3H), caffeoyl-CoA 3-O-methyltransferase (CCoAOMT), cinnamoyl-CoA reductase (CCR), cinnamyl alcohol dehydrogenase (CAD), ferulate 5-hydroxylase (F5H) and caffeic acid/5-hydroxyferulic acid O-methyltransferase (COMT). Monolignol biosynthesis is a tightly-controlled process and its regulation is an active area of research. Following biosynthesis, monolignols are exported to the cell wall where polymerisation and lignin deposition occurs. The export process is thought to occur through the 4-O-glucosylated form of monolignols produced by glucosyltransferases although this mechanism is poorly understood and no dedicated membrane transporters have yet been identified (Vanholme et al., 2010). The subsequent polymerisation probably occurs through a radical-coupling process whereby oxidation produces monolignol radicals that readily couple together in a combinatorial manner. The dehydrogenation is likely mediated by plant peroxidases, laccases or other polyphenol oxidases which are present in the cell wall (Baucher et al., 1998). Two monomer radicals may first couple together to quench unpaired electrons and form dimers. Radicals are stabilised by electron delocalisation and may be transferred to a lignin dimer or polymer to facilitate the endwise coupling reactions that build the growing chain (Sarkanen, 1971). As an alternative to the radical-coupling process, a protein-mediated approach has also been proposed but remains unproven (Davin and Lewis, 2000; Hatfield and Vermerris, 2001). Monolignol radicals preferentially couple at the β position and so the predominant linkages in lignin are β-β’, β-O-4’ and β-5’ (Freudenberg and Neish, 1968). The 5-5’ and 5-O-4’ linkages are not possible between S units but do occur with G lignin units. As a result, lignins that are lacking in S units are more cross-linked and contain more C-C bonds (Sarkanen, 1971). Indeed, the greater resilience of C-C bonds in comparison to C-O-C bonds largely explains why such lignins are more recalcitrant.
2.2 The Quest for Low-lignin Feedstocks

2.2.1 The Case for Low-lignin Plants

While humanity has exploited wood as a building material for many thousands of years, the use of wood fibres in the production of paper is a much more recent development. The modern pulp and paper industry dates back to the 1844 invention of mechanical pulping by the Canadian Charles Fenerty (Fenerty, 1844). In the decades following, its growing popularity led wood pulp to displace cotton and hemp in paper production and led to the establishment of the new discipline of wood science. Considerable research was undertaken to elucidate the structure, composition and chemistry of wood and to find uses for the many different classes of compounds found within. In fact, it was already evident by 1876 that wood could be used for the production of alcohols (Fremy, 1876).

As commercial-scale wood pulping operations were established, it quickly became clear that efficient delignification was a key determinant for plant efficiency and prosperity. Although some small markets for lignin have emerged over the years and it is advantageous in fuel wood and lumber, lignin is generally undesirable to the pulp and paper industry as it reduces pulp quality. Indeed, one of the primary goals of wood pulping in the production of high-quality paper products is to break the chemical bonds that bind lignin to polysaccharides (Smook, 2002). Vast amounts of energy and chemicals are expended in this effort. Although improvements in delignification technologies have certainly increased the overall efficiency since the 19th century, the development of less recalcitrant feedstocks is an attractive approach towards reducing the environmental footprint of biomass utilisation even further (Herrera, 2005).

While the economic realities of the pulp and paper sector provided the initial inspiration, the production of low-lignin feedstocks is even more critical now for the success of the biofuels industry. The inevitable carbon footprint associated with traditional petroleum-derived fuels will drive the development of second generation biofuels (Ragauskas et al., 2006). As in the pulp and paper industry, lignin is seen as an unwanted impurity in biofuels production. Lignin that is not removed by pre-treatment will severely inhibit the downstream fermentation process that
converts sugar to fuel alcohol (Klinke et al., 2004). The biofuels industry has learned a great deal from the pulp and paper sector and modern pre-treatment systems are built on the legacy of efficient wood processing technologies. Nonetheless, the limits of process efficiency are still largely determined by the lignin content of the feedstock.

Viable commercial low-lignin trees would obviate the need for excessive investment of energy and chemicals to delignify woody feedstocks (Pilate et al., 2005). In this age where environmental stewardship is a prerequisite rather than an afterthought, modern bioproducts derived from lignocellulose must be produced via sustainable processes. The economic and moral incentives to reduce the ecological footprint of wood processing justify efforts to develop low-lignin commercial plants.

2.2.2 Natural Variability in Lignin

There exists remarkable variability in lignin among secondary cell walls in the plant kingdom (Table 1). Besides differing in lignin content, the composition of H, G and S lignin varies among plant species and across cell types within an individual organism (Campbell and Sederoff, 1996). Lignin content even varies across the different layers of the cell wall. The highest lignin concentration is typically found in the middle lamellae and the cell corners (Fromm et al., 2003). While lignin in gymnosperms is predominantly of the G type, angiosperms may have varying amounts of G, S and H lignin. Although one must be cautious in making generalisations, it is largely true that conifers have more lignin than broadleaf species. In general, coniferous plants have lignin content in the range of 25-35% whereas broadleaf lignin content typically ranges from 20-30% (Pettersen, 1984). Not only does the amount of each monolignol component vary, but the lignin structure is also highly variable (Monties, 2005). The representation of lignin linkages appears to be primarily a function of monolignol composition but it is not possible to entirely rule out another level of regulation in bond making – be it protein-mediated or merely a function of the cell wall environment during lignification (Hatfield and Vermerris, 2001).

Lignin biosynthesis has been relatively well-studied over the past 30 years. And while we now have a very good understanding of which genes and enzymes are involved in monolignol
biosynthesis, we are only beginning to understand the dynamic equilibrium that governs monolignol deposition. It is now recognized that lignin variability across different species and tissue types is largely due to differences in gene expression mediated by a regulatory cascade of transcription factors (Rogers and Campbell, 2004; Vanholme et al., 2010). However, it is also becoming apparent that other mechanisms such as micro-RNA-mediated regulation may play an important role in tightly-regulating lignin biosynthesis (Neutelings, 2011). Moreover, the secretion and polymerisation of monolignols in the cell wall is a very poorly understood process. It is entirely conceivable that lignification may also be regulated at the levels of export and dehydrogenation-mediated polymerisation. In short, while our understanding of the genetic basis lignification has improved dramatically in recent years, there is a great deal more that we must learn about these dynamics in order to produce viable high-yielding hypolignified plants (Li and Chapple, 2010).

Table 1. Lignin content of several North American woods.

<table>
<thead>
<tr>
<th>Species</th>
<th>Lignin Content (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sugar maple (Acer saccharum)</td>
<td>23</td>
</tr>
<tr>
<td>American elm (Ulmus americana)</td>
<td>24</td>
</tr>
<tr>
<td>Trembling aspen (Populus tremuloides)</td>
<td>21</td>
</tr>
<tr>
<td>Balsam fir (Abies balsamea)</td>
<td>29</td>
</tr>
<tr>
<td>White spruce (Picea glauca)</td>
<td>27</td>
</tr>
<tr>
<td>White cedar (Thuja occidentalis)</td>
<td>31</td>
</tr>
</tbody>
</table>


2.2.3 Lignin Engineering

The first attempts to engineer low-lignin plants followed shortly after the development of plant transformation methods in the 1970s (Neutelings, 2011). While this work was initially done using models such as Arabidopsis and Nicotiana spp., more recent work has sought to engineer lignin in relevant feedstocks such as Populus spp. Although there have been great successes in engineering lignin, the phenotypic result is not always as expected or desired. For example, down-regulating biosynthesis of one monolignol may inadvertently boost the production of
another (Li et al., 2003). The overall result may be a zero net decrease in lignin content. And in other cases, the primary result has been a surge in the production of cellulose (Hu, 1999). But in the end, the most disconcerting effects of tampering with monolignol biosynthesis are the gross plant defects such as dwarfism, delayed development or even lethality caused by an accumulation of harmful intermediates and an imbalance of metabolites needed for normal plant development (Bonawitz and Chappel, 2010). Clearly lignification is a carefully controlled biochemical process in nature.

Most efforts in lignin engineering have targeted the suite of genes encoding monolignol biosynthetic enzymes. On the whole, there have been tremendous successes in controlling monolignol biosynthesis. In summary of this large body of work, lignin content can be reduced by varying degrees through downregulation of any of the following genes in the monolignol biosynthetic pathway (see Figure 4): PAL, C4H, 4CL, HCT, C3H, CCoAOMT, CCR and CAD (Vanholme et al., 2008; Simmons et al., 2010). For example, lignin content in *Pinus radiata* was reduced from 29% to as little as 17% by downregulating the expression of HCT (Wagner et al., 2007). Besides such drastic changes in lignin, the composition can also be affected by altering the expression of the following genes: HCT, C3H, F5H, COMT and CAD (Vanholme et al., 2008). For instance, the S/G lignin ratio in *Populus* lignin was increased more than 15 fold by overexpressing F5H (Stewart et al., 2009).

While monolignol biosynthesis is clearly an important part of the equation, it is not the sum total of lignification. Comparatively fewer researchers have tried to engineer lignin by targeting the glucosyltransferases that may export monolignols from the cell or the secreted peroxidases and laccases that may be involved in the radical-coupling polymerisation of monolignols (Vanholme et al., 2008). Generally speaking, these two processes are still poorly understood. Nonetheless, controlling lignification through these mechanisms holds tremendous promise. For instance, down-regulation of a peroxidase in *Populus* has been shown to significantly affect lignin composition (Li et al., 2003b). Constraining lignification at the level of the cell wall is appealing since it would not necessitate drastic changes to important metabolic processes. Moreover, these enzymes show remarkable substrate specificity and drastic changes in lignin
structure could be effected by directing the population of such enzymes (Bonawitz and Chappel, 2010). For example, it might be possible to engineer less cross-linked lignin that would be more easily degraded (Vanholme et al., 2010). However, the main challenge in engineering lignin by targeting monolignol dehydration is that there is considerable redundancy in these enzymes and most isoforms still remain undiscovered and uncharacterised.

### 2.2.4 Towards Viable Low-lignin Plants

There are a great number of questions remaining in the quest for low-lignin trees. A careful study of plant species that are naturally low in lignin will reveal which route to hypolignification is preferred through evolution. By taking a systems biology approach to studying such species, it will be possible to discern whether nature has achieved low lignin levels by carefully regulating the expression of biosynthetic genes, by regulating the export of monolignols, by controlling the polymerisation process during lignin deposition, or through a combination of all these mechanisms.

The study of low-lignin species could also reveal how plants survive with less lignin. Cell wall lignin plays an important role in structural stability and also in plant defence. As a result, engineered hypolignified plants invariably have reduced strength and resistance to attack by insects and fungi (Cummins, 2005). In addition, low-lignin transgenic plants produced to date typically show signs of water transport problems and vessel collapse (Li and Chappel, 2010). Plant engineers must overcome such issues of plant vitality in order to produce agronomically-viable genotypes. Woody plants that have evolved low levels of lignin must have developed such compensatory traits and theses species could inspire future approaches in plant transgenics.

In short, the dynamic equilibrium that governs monolignol biosynthesis, export and polymerisation remains elusive to plant engineers and new model species with atypical lignification are now needed to demystify this cryptic biochemistry (Neutelings, 2011).
2.3 Enzymatic Modification of Plant Polysaccharides

2.3.1 Carbohydrate-active Enzymes

Carbohydrates are one of the most abundant classes of chemical compounds on the planet. Nearly all forms of life produce or metabolise polysaccharides. Therefore, evolution has produced an immense array of enzymes that act on these biopolymers (Cantarel et al., 2009; Gilbert et al., 2008). Such carbohydrate-active enzymes are involved in everything from the production of starch granules in maize plants to the destruction of cellulose by wood-decaying fungi. For example, glycosyltransferases catalyse the formation of new glycosidic bonds while glycoside hydrolases cause hydrolytic cleavage of such linkages. Similarly, carbohydrate esterases sever ester linkages and transglycosidases cause the substitution of one glycoside with another.

In recent years, carbohydrate-active enzymes have been successfully deployed for various biotechnological purposes. For example, amylases, cellulases and xylanases have all been used in the industrial processing of plant polysaccharides. These hydrolytic enzymes have been used in applications as diverse as animal feed, wine production and wood pulping (Bhat, 2000). In some cases, purified enzymes are used while enzyme cocktails are preferred in other processes. Another successful deployment of carbohydrate-active enzymes has been the use of glycosyltransferases in polysaccharide modification (Palcic, 2011). For instance, xyloglucan endo-transglycosylase has been used to derivatise cellulose by modifying xyloglucan chains which are then assembled onto cellulose surfaces (Brumer et al., 2004). Similarly, glycoside hydrolyases have been engineered for use in the synthesis of custom oligosaccharides (Mackenzie et al., 1998). A large number of these so-called glycosynthases have been developed to produce wide-ranging synthetic carbohydrates.

2.3.2 Carbohydrates Oxidases for Carbohydrate Derivitisation

Carbohydrate oxidases are another class of carbohydrate-active enzyme which has recently found biotechnological purpose in the derivatisation of plant-derived carbohydrates. These oxidoreductase enzymes convert hydroxyl groups to carbonyl or carboxylic acid species (van
Hellemont et al., 2006). In doing so, oxygen is used as an electron acceptor and hydrogen peroxide is released. In nature, such enzymes are thought to provide hydrogen peroxide substrate to peroxidases involved in cell-wall attack and pathogenesis (Ander and Marzullo, 1997). From a technology perspective, oxidation of hydroxyl groups is a means of increasing the reactivity of polysaccharides and facilitating further chemical derivatization (Allen and Cuculo, 1973). By generating reactive handles, carbohydrate oxidases could facilitate the further derivitisation of plant-derived carbohydrates (Turner, 2011). For example, subsequent esterification or amination could produce a wide range of valuable bioproducts (Yalpani and Hall, 1982).

Much like in the petroleum industry where refineries must produce high-value commodities such as waxes, lubricants, and other petrochemicals to ensure economic viability, successful biorefineries will produce high-value coproducts from carbohydrate processing streams. For example, native cellulose can be carefully processed to generate nanocellulose particles with impressive strength properties (Ye, 2007). Upon oxidation of hydroxyl groups and further chemical derivatisation, nanocellulose can be used in biocomposite plastics to displace petroleum-derived products (Hubbe et al., 2008). Similarly, enzymatically-modified hemicellulose preparations could be formed into bioactive films or gels (Ebringerová et al., 2005).

While there are several chemical approaches to the oxidation of carbohydrates, these typically involve harsh solvents and complex multi-step reactions (Klemm et al., 1998b). The most common reagent for polysaccharide oxidation is (2,2,6,6-tetramethylpiperidin-1-yl)oxyl (Isogai et al., 2010). While chemical approaches are feasible, reductions in chain length and crystallinity are often unavoidable (Kim et al., 2000; Saito et al., 2010b). On the other hand, enzyme-based approaches can provide specificity in fewer reaction steps with mild reactions conditions. Since enzyme-assisted processes are applied examples of green chemistry, the use of carbohydrate oxidases in polysaccharide derivatisation may enable sustainability in the production of high-value bioproducts.

Glucose oxidase and pyranose-2-oxidase are the most well-known carbohydrate oxidases. Unfortunately, both only accept monomeric substrates and are thereby limited in usefulness for
derivatising plant-derived polysaccharides (Spadiut et al., 2009; Griffhorn, 2000). Recently, attention has fallen on the small number of oxidases that accept oligosaccharide and polysaccharide substrates. Successful deployment of these enzymes in carbohydrate derivatisation will require a thorough understanding of substrate binding and catalytic mechanisms. Furthermore, protein engineering could be used to expand the usefulness of these enzymes (van den Heuvel et al., 2000).

### 2.3.3 Galactose Oxidase

Galactose oxidase (GalOx, EC 1.1.3.9) is a fungal oxidoreductase from *Gibberella zeae* belonging to the radical copper oxidase family (Figure 5A; McPherson et al., 1992). GalOx cleaves the pro-S hydrogen and converts the C6 hydroxyl of D-galactosyls in oligosaccharides to the corresponding aldehyde (Figure 5B; Avigad et al., 1962; Whittaker, 2002; Parikka et al., 2010). This reaction consumes molecular oxygen and generates a hydrogen peroxide co-product.

GalOx is especially notable as a metalloenzyme that catalyses two-electron redox chemistry with a single mononuclear metal ion site (Whittaker and Whittaker, 1998). A unique post-translational modification generates a catalytic thioether bridge between Cys^{228} and Tyr^{272} that coordinates with the single Cu atom prosthetic group. GalOx is unique among carbohydrate oxidases as it targets the C6 hydroxyl group of oligomeric and polymeric substrates.

![Figure 5](image-url)

Figure 5. Panel A: The crystal structure of GalOx (PDB 1GOF) with the active site copper atom (green) and thioether bridge (blue) shown. Panel B: The oxidation of D-galactopyranose catalysed by GalOx. All protein structure images were generated using Swiss PDB Viewer version 4.0 (Swiss Institute of Bioinformatics).
While homologues have been identified in *Gibberella fujikuroi, Stigmatella aurantiaca* and several other species, the enzyme that is secreted from *G. zeae* has been the primary focus of study (van Hellemond et al., 2006). GalOx is encoded by a 2043 bp coding sequence which contains no introns and features an N-terminal export signal sequence and a pro-sequence (McPherson et al., 1992). The native enzyme is glycosylated and boasts considerable stability owing to its β-sheet dominated structure. The mature protein comprises 639 amino acids and has a total molecular weight of approximately 69 kDa. GalOx contains a 480-residue catalytic domain and a 147-residue substrate-binding domain which is connected by a 12-residue linker sequence (Ito et al., 1994). The production of functional enzyme requires the cleavage of the secretion signal sequence, the release of the 17-amino acid pro-sequence and the formation of the thioester bridge in the active site (Firbank et al., 2001). The catalytic domain is composed of 28 β-strands arranged in a pseudo 7-fold symmetry with an additional seven-strand β-sheet whereas the binding domain consists of an eight-strand β-sandwich (Ito et al., 1994). The binding domain is a family 32 carbohydrate-binding module (CBM, see section 2.3.6 below) which is thought to have affinity towards galactose-containing oligosaccharide substrates.

A stepwise reaction mechanism has been proposed for GalOx (Baron et al., 1994; Himo et al., 2000). His496, His581, and Tyr272 are involved in copper coordination and the formation of the thioether bridge generates a radicalised Tyr272 (Figure 6). The proposed mechanism begins with the enzyme in the fully oxidized state (Tyr272 radical and Cu(II)). A proton is

![Figure 6. Close-up view of the galactose oxidase active site. Residues shown in red are involved the thioether bridge while those shown in blue coordinate the copper ion, which is shown in green.](image-url)
transferred from the substrate to Tyr^{495} and a C2 hydrogen atom is transferred from the substrate to Tyr^{272}. At this point, the enzyme is in the intermediate redox state (non-radical Tyr^{272} and Cu(II)). An electron transfer from Tyr^{272} to the copper atom brings the enzyme to its fully reduced state (non-radical Tyr^{272} and Cu(I)). Upon release of substrate, molecular oxygen reoxidises the enzyme (van Hellemond et al., 2006 and Baron et al., 1994).

GalOx has been used for the oxidation of a wide array of galactose-containing oligosaccharides and polysaccharides. Compared to other carbohydrate oxidases, GalOx features a surface-exposed active site so that bulky substrates are permissible (Ito et al., 1994). For example, GalOx accepts hemicelluloses such as galactomannan and galactoglucomannan (Parikka et al., 2010). The post-translational modification required for thioether bond formation is an autocatalytic, copper-mediated process such that heterologous expression is feasible (Rogers et al., 2000). GalOx has been successfully recombinantly produced in *Pichia pastoris*, *Aspergillus nidulans* and *Escherichia coli* (van Hellemond et al., 2006). While the wild-type enzyme is useful in its own right, protein engineering efforts could expand the usefulness of GalOx for the oxidation of plant-derived polysaccharides (Deacon and McPherson, 2011).

### 2.3.4 Glucooligosaccharide Oxidase

Glucooligosaccharide oxidase (GOOX, EC 1.1.3.X) is an oxidoreductase from *Acremonium strictum* that belongs to the vanillyl alcohol oxidase family (Figure 7A; Lin et al., 1991). GOOX catalyses the oxidation of the anomeric hydroxyl on the terminal glucosyl residue of glucooligosaccharides while concurrently reducing molecular oxygen to produce hydrogen peroxide (Figure 7B). It is a monomeric glycoprotein that features a double covalently-bound flavin adenine dinucleotide (FAD) cofactor.

Homologues of GOOX include chitooligosaccharide oxidase from *Gibberella zeae*, celloooligosaccharide oxidase from *Paraconiothyrium* sp. and maltooligosaccharide oxidase from *Microdochium nivale* (van Hellemond et al., 2006). GOOX is encoded by a 1500 bp coding sequence which contains one intron and an N-terminal signal sequence. It contains at least two N-glycosylation sites (Asn^{365}, Asn^{341} confirmed, and Asn^{394} putative) and has an apparent
molecular weight of about 61 kDa (Huang et al., 2005). GOOX features an FAD-binding domain and a substrate binding domain with a comparatively open carbohydrate binding groove. The FAD-binding domain comprises two subdomains. The first subdomain contains four \( \beta \) strands sandwiched by three \( \alpha \) helixes. The second subdomain contains five antiparallel \( \beta \)-strands covered by five \( \alpha \) helixes. The FAD cofactor is covalently bound to the apoenzyme at His\(^{70} \) (8\( \alpha \)-N1-histidyl) and Cys\(^{130} \) (6-S-cysteiny1) (Huang et al., 2005). The doubly covalently-bound FAD serves to modulate the redox potential needed for the complete oxidation of substrate (Huang et al., 2008). The substrate-binding domain is composed of a seven-strand antiparallel \( \beta \)-sheet covered by five \( \alpha \) helixes. While it is not a true carbohydrate-binding domain (see Chapter 2.3.6), the broad open carbohydrate-binding groove permits oligomeric substrates to align head-first into the active site.

The GOOX-catalysed reaction is thought to proceed by way of two half-reactions: i) oxidation of the substrate glycosyl to its corresponding lactone, and ii) the water-mediated hydrolysis of the lactone to generate the corresponding carboxylic acid (van Hellemond et al., 2006). Tyr\(^{429} \) is thought to be responsible for abstracting the proton from the C1 hydroxyl and Asp\(^{355} \) may help lower the pKa through water-mediated hydrogen bonding (Figure 8; Huang et al., 2005).
al., 2005). Also, Tyr$^{300}$ and Trp$^{351}$ form a stacking pair characteristic of many carbohydrate-protein interactions (Huang et al., 2005).

GOOX possess a broad substrate tolerance on a range of α- and β-linked oligosaccharides which contain a reducing-end glucosyl such as lactose, maltooligosaccharides and cellooligosaccharides (Lin et al., 1991). While the active site is located in an open carbohydrate-binding groove, it is comparatively shallow and requires that substrates align head-on such that bulky substrates are inadmissible. On the other hand, GOOX is amenable to protein engineering efforts to broaden substrate tolerance (Lee et al., 2005; Foumani et al., 2011). Further mutagenesis may produce GOOX variants that are useful in derivatising hemicelluloses.

2.3.6 Carbohydrate-binding Modules

Many carbohydrate-active enzymes feature non-catalytic binding modules that improve overall reaction rates by increasing the local concentration of substrate (Bolam et al., 1998). These carbohydrate-binding modules (CBMs) are autonomously-folding domains which bind carbohydrates through hydrophobic interactions. CBMs may exist as part of a large carbohydrate-active enzyme or as independent proteins (Guillén et al., 2010). In addition, the large multi-protein cellulosome complexes produced by cellulolytic bacteria invariably feature CBMs.
At present, there are 64 recognised families of CBMs which are organised by amino acid sequence similarity (CAZy database). While all CBMs are β-sheet-dominated structures, there is considerable diversity in structure (Hashimoto, 2006; Blake et al., 2006). CBMs have been classified into the following three types based primarily on the ligand-binding site topology: type A CBMs feature a flat hydrophobic surface containing aromatic residues which interact with surfaces of crystalline polysaccharides; type B CBMs contain a ligand-binding cleft which contains individual hydrophobic residues that interact with free polysaccharide chains through stacking; and type C CBMs bind mono-, di- or trisaccharides (Boraston et al., 2004; Guillén et al., 2010). In addition to the sequence-based classification, CBMs are also classified based on seven fold families: β-sandwich, β-trefoil, cysteine knot, unique, cellulose-binding fold, hevein fold and unique containing hevein-like fold (Boraston et al., 2004).

Inspired by nature, enzyme engineers have used CBMs for a variety of biotechnological purposes (Shoseyov et al., 2006). Chimeric enzymes featuring CBMs have been used in applications as wide-ranging as laundry detergent additives, recombinant protein production and bioremediation. CBMs have been appended to xylanases, cellulases, arabinofuranosidases and pectin lyases to improve cell wall desconstruction (Carrard et al., 2000; Herve et al., 2010). Also, a CBM was successfully added to GalOx to target the oxidation of glucans in dental plaque and thereby prevent bacterial growth (Lis and Kuramitsu, 1997). In other cases, endogenous CBMs were replaced to improve activity on target substrates. For instance, the native CBM of a bacterial endoglucanase was replaced with a CBM from a fungal exoglucanase and resulted in improved activity on microcrystalline cellulose (Kim et al., 1998). CBMs are a valuable tool in the toolbox of protein engineers working on carbohydrate-active enzymes (Levy and Shoseyov, 2002).
CHAPTER 3: EASTERN LEATHERWOOD IS A LOW-LIGNIN PLANT WITH A STORY TO TELL

In light of the widespread efforts to generate viable low-lignin trees, I set about to study the peculiar wood chemistry of this uncommon native shrub. This chapter presents a brief introduction followed by materials and methods and then results and discussion. Conclusions and recommendations are summarised in Chapter 5.

3.1 Introduction

Eastern leatherwood (Dirca palustris L. also known as moosewood, wicopy and in French: bios de plomb), is a small understory shrub found sporadically throughout eastern North America in rich mesic soils of hardwood and mixed conifer forests (Figure 9). The Dirca genus comprises four species of slow-growing shrubs: D. occidentalis Gray, which is found only in the San Francisco bay region of California; D. mexicana Nesom and Mayfield, which is endemic to a single valley in Mexico’s Sierra Madre Oriental Mountains; D. decipiens Floden, which is a recently-identified species found in Kansas and Arkansas; and D. palustris, which is widespread but uncommon throughout eastern North America where it grows alongside shady woodland streams and in damp upland locales (Schrader and Graves, 2004; Floden et al., 2009).

Leatherwood bark was traditionally sought by First Nations peoples and early settlers for natural cordage and naturopathic medicines (Bigelow, 1818). Indeed, more recent work has identified pharmaceutically-active compounds to treat cancer and diabetes from Dirca tissues (Badawi, 1983; McCune, 2002). Other recent scientific studies of this species have examined its
reproductive mechanisms and ecological role since the ephemeral flowers emerge in early spring when the forest floor remains snow-covered and when pollinator activity is limited (Williams, 2004; Peterson and Graves, 2011).

It has long been known that xylem tissues of *Dirca* are exceptionally pliable (Thoreau, 1856). The flexibility of the wood is so great that one can even tie a knot in a twig without it snapping (Anderson, 1933). Some have previously suggested that low lignification may be a likely cause of this peculiarity (Nevling, 1962). Since lignin is the bane of the pulp and paper industry and the burgeoning cellulosic biofuels sector which together spend great amounts of money, energy and chemicals to delignify woody feedstocks, natural low-lignin plant species are eminently worthy research subjects. A careful study of leatherwood cell wall chemistry may inform widespread efforts that are underway to engineer trees that are low in lignin.

In this study, I employed traditional wood chemistry methods as well as the following techniques: Fourier-transform infrared spectroscopy (FTIR), time-of-flight secondary ion mass spectrometry (ToF-SIMS) and X-ray diffraction (XRD). In addition, principal component analysis (PCA) was used to compare the large data sets acquired with FTIR and ToF-SIMS.

3.2 Materials and Methods

3.2.1 Wood Samples

An upright-growing branch was harvested from a mature leatherwood plant (*D. palustris*) near Pembroke, Ontario (45° 49' N, 77° 6' W) in June 2010. Xylem samples for chemical analysis were debarked, air-dried and disintegrated using a Wiley mill to pass a 0.6 mm sieve. Similarly, samples for FTIR and XRD were ground to pass a 50 μm sieve. Reference samples consisted of similarly processed sapwood xylem from trembling aspen (*Populus tremuloides*) and white spruce (*Picea glauca*). Wood samples for ToF-SIMS analysis were similarly harvested in September 2010 but were immediately flash-frozen in liquid nitrogen and stored at -80ºC until use. To preserve the cellular structure, thawed samples were solvent-exchange-dried with ethanol from 20% to 100%. A defect-free flat radial cross-sectional surface was obtained using a microtome equipped with a diamond knife.
3.2.2 Lignin Content

The acid-insoluble lignin content was evaluated using the standard methodology (TAPPI, 2002a). Briefly, wood powder (0.5 g) was Soxhlet-extracted using 95% ethanol for four hours, ethanol:toluene (70:30) for six hours and water for three hours. Extractive-free wood was oven-dried and weighed prior to holocellulose hydrolysis using 72% H$_2$SO$_4$ for two hours at 20ºC and 4% H$_2$SO$_4$ for one hour at 121ºC in an autoclave. Acid-insoluble lignin was determined as the dry weight of residue after washing. Triplicate samples were processed for each species.

Acid-soluble lignin was evaluated in triplicate using UV-spectroscopy of the acid-hydrolysis filtrate (TAPPI, 2002b). An extinction coefficient of 110 L g$^{-1}$ cm$^{-1}$ at 205 nm was assumed for all species (Swan, 1965). The filtrates (145 mL) were diluted five fold for the broadleaf samples and were measured undiluted for the conifer species.

3.2.3 Fourier-transform Infrared Spectroscopy

Finely ground samples (2 mg) of wood, acid-insoluble lignin and chlorite holocellulose were homogenised with FTIR-grade potassium bromide (200 mg) in a mortar and pestle and dried overnight at 105ºC prior to being formed into compressed discs using a Carver hydraulic press and a 13-mm die mold. Chlorite holocellulose was prepared using extractive-free wood according to the standard procedure described by Browning (1967). Spectra were recorded using a Tensor 27 FTIR spectrometer (Bruker Optics, Ettlingen, Germany) at a resolution of 4 cm$^{-1}$ with 150 scans from 4000 to 400 cm$^{-1}$. The spectra were blanked with a KBr-only sample, corrected for atmospheric contributions, cropped to the fingerprint region (1800 to 800 cm$^{-1}$) and baseline-adjusted by the rubberband algorithm with 64 baseline points using OPUS version 5.1 (Bruker Optics, Ettlingen, Germany). Triplicate samples were processed for each species and each sample type.

PCA was performed for all three sample types using the PLS Toolbox version 5.8 (Eigenvector Research Incorporated, Wenatchee, WA, USA) in Matlab version 7.8 (Mathworks Incorporated, Natick, MA, USA). For each of the three sample types, three spectra from each species were mean-centred and normalised by unit area prior to analysis.
Peak deconvolution was performed for lignin spectra. Peaks were identified using the second
derivative scanning mode of OPUS version 5.1 with 9 smoothing points and curve fitting was
performed with these peaks modelled using a pseudo-Voigt shape. The Levenberg-Maquardt
algorithm was performed with the peak positions fixed and peak intensity, width and shape
unfixed. The ratio of S/G lignin was estimated using the intensity of the peaks at 1325 cm\(^{-1}\) (S
lignin, broadleafs), 1311 cm\(^{-1}\) (S lignin, conifer) and 1270 (G lignin) cm\(^{-1}\) (Rana et al., 2010).

### 3.2.4 Time-of-flight Secondary Ion Mass Spectrometry

Spectra were recorded by Surface Interface Ontario (University of Toronto) using a TOF-
SIMS IV instrument (ION-TOF GmbH, Münster, Germany) with the scan parameters previously
reported (Goacher et al., 2011). Since wood is an electrically-insulated material, low-energy
electron flooding was used to facilitate the collection of positively-charged secondary ions. High
mass resolution (bunched mode) spectra were collected with a pulsed current of ~0.3 pA and a
256×256 pixel random raster pattern over an area of 300×300 μm\(^2\). Similarly, nominal mass
resolution (burst alignment mode) spectra used to produce ion images were obtained using a
pulsed current of ~0.1 pA over an area of 62.5×62.5 μm\(^2\). Spectra were calibrated to the C\(^+\),
CH\(_3\)\(^+\), C\(_2\)H\(_3\)\(^+\), C\(_2\)H\(_5\)\(^+\), and C\(_7\)H\(_7\)\(^+\) ions using the Surface Lab version 6.1 (ION-TOF GmbH,
Münster, Germany). Xylem wood samples were Soxhlet-extracted as described above and
triplicate samples were processed for each species.

A PCA model was generated using the high mass resolution spectra and the PLS Toolbox for
Matlab. Prior to the analysis, the data were mean-centred and normalised by unit area. High mass
resolution spectra were also used to evaluate the ratio of S/G lignin using the total corrected
intensities of the peaks at m/z 167.07 and 181.05 for syringyl (S) lignin and 137 and 151.05 for
guaiacyl (G) lignin (Saito et al., 2010a).

Ion images obtained using the nominal mass resolution spectra were used to assess ion
localisation. Lignin peaks were those reported previously as “Group 1” assignments (Goacher et
al., 2011) with the addition of m/z 167 and 181 for S lignin and 107 and 121 for H lignin (Saito
et al., 2005a). Similarly, polysaccharide peaks were those reported as “Group 2” assignments (Goacher et al., 2011) with the addition of m/z 115 and 133 for xylan (Fardim and Durán, 2003).

### 3.2.5 X-ray Diffraction

X-ray diffraction of wood powder (~0.5 g) was performed using reflection mode with a Philips powder X-ray diffractometer composed of a PW1830 HT generator, a PW1050 goniometer and a PW3710 control unit alongside X’Pert Quantify data capture software version 1.0 (PANalytical B.V., Almelo, The Netherlands). The CuKα radiation generated at 40 kV and 40 mA was primarily monochromatic upon passing through a 15 μm nickel filter (PW1385/00). Diffraction profiles were recorded from 5 to 55 degrees 2θ with steps of 0.02 degrees every 2.5 seconds. An amorphous profile was obtained using stable amorphous cellulose prepared by dissolving Avicel PH-101 (50 μm particle size, Sigma-Aldrich, St. Louis, Missouri, U.S.A.) in sulphur dioxide-diethylamine-dimethylsulfoxide and regenerating in water before freeze-drying (Isogai and Atalla, 1991). Triplicate samples were processed for each species.

The amorphous cellulose curve was scaled to match each experimental spectrum at only one point (Thygesen, 2005). The crystallinity index was then evaluated as the fraction of total area above the amorphous profile which is attributable to crystallinity. To enable comparison to literature values, the crystallinity index was also calculated using the common, but unreliable, peak height method which involves the ratio between the intensity at the 002 reflection (I200) and the height of the curve at the minimum between the 002 and 101 peaks (Iam) (Park et al., 2010).

### 3.4 Results and Discussion

The first description of leatherwood appears in 1700 in the form of a letter from the early Canadian naturalist Michel Sarrazin to the French botanist Joseph Pitton de Tournefort in which he describes the many uses of leatherwood gleaned from the First Nations (Valliant, 1704; Mottiar, 2011). In 1739, the Dutch botanist Jan Frederik Gronovius described the twigs and bark as very strong and flexible (Gronovius, 1739). But it was the French botanist Van Tiegham who first reported that leatherwood tissues are “weakly lignified” (Van Tiegham, 1893). Since this early histological work, others have also observed poor staining using lignin-specific dyes
(Holm, 1921; Choquette, 1925). To my knowledge, the following is the first careful analytical study of the lignin in leatherwood xylem tissue.

### 3.4.1 Lignin Content

Vertically-growing branches were chosen to avoid any tension or compression wood which could have atypical lignification. Similarly, the samples used were devoid of any signs of attack or damage. All samples were harvested from specimens growing under similar light and soil conditions. While many techniques are available for the determination of total lignin content, the standard methodology was chosen to enable comparison to lignin content values in the literature (Hatfield and Fukushima, 2005). Acid-insoluble lignin (also known as Klason lignin) results from a two-step acid hydrolysis of polymeric carbohydrates. The insoluble lignin is measured gravimetrically upon filtering and drying after soluble residues have been removed by washing. Acid-soluble lignin is measured spectrophotometrically as the lignin component which accrues in the acid hydrolysate. To minimise the effect of waxes, tannins and other wood extractives, it is very important to thoroughly remove these components first. In the present study, an exhaustive three-step Soxhlet extraction was performed using ethanol, ethanol-toluene and water. The total lignin content was then assessed as the sum of acid-insoluble and acid-soluble lignin fractions.

The xylem lignin content was measured for leatherwood and for two reference species. These data show that leatherwood contains considerably less total lignin than the other species (Table 2). Leatherwood contained about half as much acid-insoluble lignin as spruce, and 30% less than aspen. On the other hand, leatherwood xylem had eight times as much acid-soluble lignin as spruce and 1.4 times more than aspen. Taken together, the total lignin content was found to be 15.5% for leatherwood, 21.2% for aspen and 25.4% for spruce.
Table 2. Lignin content of xylem tissues.

<table>
<thead>
<tr>
<th>Species</th>
<th>Acid-insoluble Lignin</th>
<th>Acid-soluble Lignin</th>
<th>Total Lignin Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern leatherwood</td>
<td>13.8 ± 0.20</td>
<td>1.7 ± 0.01</td>
<td>15.5 ± 0.21</td>
</tr>
<tr>
<td>Trembling aspen</td>
<td>20.0 ± 1.16</td>
<td>1.2 ± 0.05</td>
<td>21.2 ± 1.21</td>
</tr>
<tr>
<td>White spruce</td>
<td>25.2 ± 0.23</td>
<td>0.2 ± 0.01</td>
<td>25.4 ± 0.24</td>
</tr>
</tbody>
</table>

Three replicates were analysed and standard deviations are shown alongside mean values. All values are mass % of dry, extractive-free wood.

In comparison to other wild-type shrubs and trees which have been reported, the lignin content in leatherwood xylem is remarkably low (Petterson, 1984). These results provide analytical support to the histological observations of poor lignification (Choquette, 1925). In this study, sapwood xylem samples from two other species were also included as technical controls. The results for these reference samples correspond reasonably well with the literature values: 21% total lignin for trembling aspen and 27% for white spruce (Petterson, 1984). Although there are no other known reports for lignin content in leatherwood xylem, Ferrari reported acid-insoluble lignin content in autumn leaves to be 13.4% (Ferrari, 1999). Lignin content in leaves typically corresponds well to the levels in xylem and this value affirms the measurement of 13.8%.

In contrast to the low acid-insoluble lignin from leatherwood, the amount of acid soluble lignin was remarkably high. While it was not unexpected that leatherwood – a broadleaf species, produced more acid-soluble lignin than spruce – a conifer, it was remarkable that the acid-soluble lignin content in leatherwood was significantly greater than that of aspen. Since the analysis was done in parallel, these results revealed that the lignin in leatherwood was more easily hydrolysed. Given that the hydrolysability of lignin is directly correlated to structure and composition (Grabber, 2005; Baucher et al., 2003), the wood chemistry of leatherwood clearly merits further study.
3.4.2 Evaluation of Lignocellulose Composition by FTIR

FTIR is a non-invasive analytical technique which is increasingly being used in a semi-quantitative approach to the chemical characterisation of wood (Zhou et al., 2011b). By monitoring the absorbance of infrared light across a range of wavelengths, FTIR provides a measure of the chemical bond vibrations within a sample. Sample preparation involves drying to remove all traces of water and embedding in an alkali halide media such as KBr which has no interfering absorption bands. While sample preparation is simple, data analysis can be cumbersome because of the large data sets that are produced. Statistical techniques such as PCA facilitate comparative analyses by generating lower order linear models which describe and reveal patterns in large multivariate data sets. PCA analysis of FTIR spectra can reveal a great deal about the chemical makeup of wood.

Since FTIR spectra of powdered wood are notoriously complex because of overlapping peaks from lignin and polysaccharides, spectra of chlorite holocellulose and acid-insoluble lignin were also studied. Figure 10 presents representative baseline-corrected FTIR spectra and the corresponding PCA analyses. For all three sample types, the PCA models captured over 99% of the cumulative variance in two principal components. Broadly speaking, the fingerprint regions of the spectra (1800 to 800 cm\(^{-1}\)) were typical of those previously published in the literature (Hergert, 1971). Some key signatures of the spectra include the following: C=O stretch in the region from about 1800 to 1600 cm\(^{-1}\); aromatic skeletal vibrations in the range of 1600 to 1500 cm\(^{-1}\); C-H deformations between 1470 and 1370 cm\(^{-1}\); C-O stretching in the region from 1160 to 1270 cm\(^{-1}\); glycosyl ring and bond vibrations in the region of 1200 to 1000 cm\(^{-1}\); and, C-O deformations below 1000 cm\(^{-1}\). A complete and fully-referenced list of literature assignments for lignocellulose FTIR peaks is included as Appendix 1.

Figure 10 presents representative scans for powdered wood and the corresponding two-component PCA model. The first principal component differentiated aspen, and leatherwood to a much lesser extent, from spruce. Compared to spruce, aspen was rich in the regions around 1730 and 1240 cm\(^{-1}\). On the other hand, the spruce spectra were more prominent in the signals at 1510, 1385 and the bands near 1085 and 1015 cm\(^{-1}\). While many of these regions are confounded by
Figure 10. PCA analysis of FTIR spectra. One representative spectrum is shown for each species and sample type in the first column of plots. The second column shows PCA scores plots while the final two columns show the loadings for the first two principal components. Three spectra for each species were used to generate two-component PCA models for each sample type which captured over 99% of the cumulative variance.
both lignin and polysaccharide signatures, the peak at 1510 cm⁻¹ undoubtedly corresponds to
generic lignin aromatic skeletal vibration confirming that spruce contains more total lignin
(Hergert, 1971). Also, the peak at 1730 cm⁻¹ in aspen corresponds to the stretching of C=O bonds
which are typically associated with acetyl groups in polysaccharides (Marchessault, 1962; Faix,
1991). The second principal component of the wood spectra PCA model distinguished spruce
and aspen from leatherwood. Although the PC2 scores for all three species were quite low,
spruce and aspen contained more of the bonds described by 1725, 1510, 1275, 1160, 1120, 1060
and 1030 cm⁻¹ whereas leatherwood featured comparatively more of signals at 1755, 1625 and
975 cm⁻¹. The bands at 1275 and 1030 cm⁻¹ most likely represent lignin (Hergert, 1971) whereas
the signals at 1160 and 1060 probably correspond to carbohydrates (Fengel and Ludwig, 1991).
Due to the many overlapping peaks from lignin and carbohydrates in these spectra, it was
necessary to consider lignin and holocellulose in isolation to fully realise the potential of FTIR.

By comparing the spectra for the three different samples type in Figure 9, it is immediately
apparent that the lignin spectra contained much lower signals in the carbohydrate-characteristic
region of 1200 to 900 cm⁻¹ than the powdered wood spectra. Also, the holocellulose spectra
contained many broad overlapping peaks in the region from 1600 to 1200 cm⁻¹. The acid-
insoluble lignin analysed in the present study likely contained very little if any carbohydrates
after the extensive hydrolysis. On the other hand, the chlorite holocellulose preparation may still
contain as much as 4% lignin (Browning, 1967).

The first principal component of the holocellulose spectra PCA model captured 94% of the
total variance and distinguished leatherwood, and aspen to a lesser extent, from spruce.
Leatherwood and aspen contained more of the bonds associated with 1740, 1245 and a wide
region near 1040 cm⁻¹. The presence of the 1740 cm⁻¹ band in the holocellulose PCA model
confirms that leatherwood, and aspen to a lesser extent, were rich in the C=O bond stretch
associated with acetylated carbohydrates (Marchessault, 1962). This was not unexpected since
xylan, which occurs in more abundance in broadleafs, is often highly acetylated. Previous
chemical analyses have shown that trembling aspen contains nearly twice as many acetyl groups
as white spruce (Timmel, 1957). The band at 1245 cm⁻¹ and the region near 1040 cm⁻¹ are also
both associated with increased acetylation (Saikia et al., 1995; Xu et al., 2010). The second principle component, which captured less than 6% of the total variance, separated aspen from spruce and leatherwood. The PCA scores associated with this principal component were very low. Nonetheless, it is worth noting that holocellulose from leatherwood was apparently rich in the bonds associated with 1605, 1510 and 1270 cm\(^{-1}\) while it was lacking in the signals at 1165, 1110 and 980 cm\(^{-1}\). The peaks at 1165 and 980 cm\(^{-1}\) are related to carbohydrates (Maréchal and Chanzy, 2000; Kačuráková et al., 2000) and the bonds at 1605, 1510 and 1270 cm\(^{-1}\) are associated with lignin (Hergert, 1971) that evidently remains as a minor impurity in the holocellulose preparations. The signal at 1110 cm\(^{-1}\) could be related to carbohydrates or lignin (Pandey, 1999; Maréchal and Chanzy, 2000).

The first component of the lignin spectra PCA model separated aspen from spruce and leatherwood. It revealed that lignin from aspen contained more of the signal at 1110 cm\(^{-1}\) which corresponds to aromatic C-H in-plane deformation in lignin, whereas lignin from spruce and leatherwood contained more of peak at 1510 cm\(^{-1}\) which is also characteristic of total lignin (Pandey, 1999). The second principal component captured over 34% of the total variance and distinguished spruce, and aspen to a lesser extent, from leatherwood. Lignin from spruce and aspen were rich in the peaks at 1510, 1270, 1140 and 1030 cm\(^{-1}\) whereas leatherwood contained more of the bonds related to 1715, 1665 and 1110 cm\(^{-1}\). The absorption signals at 1510 and 1140 cm\(^{-1}\) represent total lignin whereas the bands at 1270 and 1030 cm\(^{-1}\) are related to G-lignin (Pandey, 1999; Faix, 1991). These results suggest that, in comparison to spruce and aspen, leatherwood contained less total lignin and less G-lignin in particular. It is also very interesting to note that leatherwood appeared rich in the C-H stretching bonds at 1110 cm\(^{-1}\) and the C=O stretching bonds located near 1715 and 1665 cm\(^{-1}\) (Hergert, 1971; Faix, 1991). During acid hydrolysis, free methoxy groups in lignin may be converted to carbonyls (Hergert, 1971). Therefore, the abundance of carbonyl groups in the leatherwood spectra may reveal that leatherwood lignin is less cross-linked and retains more free methoxy groups.
3.4.3 Evaluation of Lignocellulose Composition by ToF-SIMS

In addition to FTIR, ToF-SIMS was also employed to evaluate the composition of leatherwood xylem. ToF-SIMS is a powerful analytical technique for measuring the chemical constitution of a wide range of surfaces (Sodhi, 2006). In essence, primary ions bombard the sample surface and secondary ions produced in the collision are captured and identified using a time-of-flight mass spectrometer. The use of this technology for characterizing lignocellulosic samples has recently been bolstered by efforts to produce reliable fingerprint peak assignments for lignin and carbohydrates (Goacher et al., 2011; Appendix 2). In the present study, ultra-flat cross-sections of xylem, which had been exhaustively solvent-extracted, were examined by ToF-SIMS to complement the data obtained using FTIR spectroscopy.

A two-component PCA model was generated which captured over 96% of the cumulative variance. These results are shown in Figure 11 alongside representative spectra for each species. The first principal component distinguished aspen from spruce. It was dominated by the presence of the m/z 43.02 peak present in the broadleaf species. This signal corresponds to acetyl ions (C$_2$H$_3$O$_+$) and its presence can be attributed to acetylated xylan. As expected, aspen was also rich in the S-lignin peaks of m/z 167.07 and 181.05 and the H-lignin characteristic peak at m/z 121.03 whereas spruce contained more of the G-lignin peak at m/z 137.06 (Saito et al., 2005a). Also more prevalent in aspen were the carbohydrate-related peaks at m/z 57.03, 69.04, 85.03, 97.03 and 115.04 (Goacher et al., 2011). The second principal component captured over 11% of the variance and differentiated aspen and spruce from leatherwood. Again this principal component was dominated by the presence of the acetyl ion m/z 43.02 indicating that leatherwood xylan is highly acetylated. All of the remaining notable peaks were more abundant in spruce and aspen. These included the G-lignin peaks at m/z 137.06 and 151.05, the H-lignin peak at m/z 121.03, the other lignin-related peaks at m/z 29.04 and 43.06, the cellulose-related peaks at m/z 127.04 and 145.05, the xylan-specific peaks at m/z 115.04 and 133.05, and the other carbohydrate-related peaks at m/z 41.04, 45.03, 55.02, 57.03, 69.04, 73.03, 85.03, 97.03 (Saito et al., 2005a; Goacher et al., 2011). The scarcity of lignin-related signals from leatherwood confirmed that it contained less lignin. Moreover, since ToF-SIMS liberates lignin-characteristic
ions with no bias towards linkage type, it is not feasible to extrapolate these results further to infer the extent of cross-linking or condensation in lignin (Saito et al, 2005b).

![Spectra and PCA Scores](image)

Figure 11. PCA analysis of ToF-SIMS spectra. One representative spectrum is shown for each species alongside the PCA scores plot. The final two plots show the loadings for the first two principal components. Three spectra for each species were used to generate a two-component PCA model which captured over 96% of the cumulative variance.

In summary, the PCA analyses of FTIR and ToF-SIMS spectra revealed that, of the three species examined, spruce xylem contained the most lignin and the most G-lignin-specific signals. Similarly, aspen provided more S-lignin-specific peaks. Leatherwood, and to a lesser extent aspen, provided strong signals related to acetyl groups suggesting that xylan was highly acetylated. A more focused approach was required in order to appreciate why leatherwood lignin was more easily degraded by acid hydrolysis.
3.4.4 Estimation of Lignin Composition

In order to further explore the composition of lignin in leatherwood, a semi-quantitative analysis was performed by deconvoluting the FTIR spectra of lignin (Rana et al., 2010) and by carefully examining the ToF-SIMS spectra of whole wood (Saito et al., 2010a). Peak picking and curve fitting are computational methods that must be used judiciously (Faix, 1992). The results must be readily acknowledged as semi-quantitative and should only be viewed in light of appropriate controls. In the present study, the S/G lignin ratio was estimated using triplicate lignin FTIR spectra of each species by a systematic and reproducible approach. Figure 12 shows a representative spectrum which has been deconvoluted. To complement this analysis, ToF-SIMS high mass resolution data from triplicate spectra were also used to estimate the ratio of S/G lignin from solid wood samples. The results obtained from both methods are presented in Table 3.

The S/G lignin ratios determined by FTIR and ToF-SIMS show similar trends. In both cases, leatherwood was found to have the highest S/G ratio. Using the FTIR method, the S/G lignin ratio was more than twofold that of aspen and over three times that of spruce. On the other hand, the ToF-SIMS approach showed that the leatherwood S/G ratio was only 25% greater than aspen but over four times greater than spruce. Although the trend is more important than the absolute values, it is worth mentioning that the S/G lignin ratios for aspen correspond reasonably well with previously published data. For example, the S/G ratio in *Populus trichocarpa* has been measured to be as low as 0.5 and as high as 2 using a variety of techniques (Zhou et al., 2011a; Sannigrahi et al., 2010).
Table 3. Lignin composition estimates.

<table>
<thead>
<tr>
<th>Species</th>
<th>S/G lignin estimated by FTIR</th>
<th>S/G lignin estimated by ToF-SIMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern leatherwood</td>
<td>2.47 ± 0.18</td>
<td>1.03 ± 0.09</td>
</tr>
<tr>
<td>Trembling aspen</td>
<td>1.23 ± 0.38</td>
<td>0.81 ± 0.02</td>
</tr>
<tr>
<td>White spruce</td>
<td>0.71 ± 0.03</td>
<td>0.24 ± 0.01</td>
</tr>
</tbody>
</table>

Three replicates were analysed and standard deviations are shown alongside mean values.

The S/G ratio is increasingly recognised as an important metric in efforts to reduce biomass recalcitrance, perhaps being even more important than the absolute lignin content (Studer et al., 2011). Several recent studies have shown improved digestibility was obtained in feedstocks with elevated S/G lignin ratios (Li et al., 2010; Huntley et al., 2003). This may be due to the greater susceptibility of β-O-4’ linkages to cleavage compared to C-C linkages which are less prevalent with decreasing G-lignin content (Russell et al., 2000). Also, since S lignin is not capable of forming C₅ linkages, S-rich lignins are less cross-linked and therefore less recalcitrant (Walker, 2006; Studer et al., 2011). Several studies have previously shown that the S/G lignin ratio can be modified by augmenting monolignol biosynthesis (Vanholme, 2010). These results suggest that natural hypolignified broadleaf species such as leatherwood may produce more S-lignin.

3.4.5 Evaluation of Lignin Localisation by ToF-SIMS

Another important factor in biomass recalcitrance is the resistance of fibre cells to defibrillation caused by lignin in the middle lamellae (Li et al., 2006). In order to assess the localisation of lignin in the leatherwood cell wall, ion images were produced using nominal-mass resolution ToF-SIMS scans across an area of several xylem cells. Since ToF-SIMS can raster primary ions across a surface, it is possible to capture images that depict the distribution of ions in a cross-section of wood (Tokareva et al., 2011). Figure 13 shows total ion images alongside ion-coloured images where green represents ions that are characteristic of carbohydrates and red represents ions typical of lignin.
As denoted by the white arrows, middle lamella was clearly visible in all three species. In poplar and spruce, the middle lamellae appeared to contain much higher levels of lignin. On the contrary, the leatherwood image was far less definitive and it was difficult to discern whether there was preferential localisation of lignin or carbohydrates.

Several previous reports have confirmed the localisation of lignin in the middle lamella (Wi et al., 2005; Fromm et al., 2003). After biosynthesis, monolignols are exported to the middle lamella where polymerisation occurs and lignin deposition follows. Since leatherwood xylem contains lower overall lignin levels, it is not surprising that the middle lamella would also be lacking. Still, one must be cautious in over-interpreting individual images as representative of all xylem cells in the species. Broadleaf species are known to show considerable heterogeneity in the micro-distribution of lignin in the middle lamella (Daniel et al., 1991). Also, leatherwood samples routinely produced low total ion counts compared to the other species, further complicating the analysis. In short, while it appears that lignin may not be localised in the middle lamella of leatherwood, such a phenomenon cannot be excluded.

### 3.4.6 Cellulose Crystallinity

In light of the comparatively low lignin content and noteworthy composition of lignin in leatherwood, it was sensible to consider whether cellulose chemistry might be affected to compensate. Cellulose crystallinity is routinely measured using X-ray diffraction techniques.
(Garvey et al., 2005). In reflection mode, the X-rays diffracted from a sample of powdered wood are captured by a detector mounted to a goniometer that is slowly rotated through a range of angles. In the present study, very-fine wood powder was used which had been dried to ambient humidity and formed into a flat, packed bed on a quartz slide. The crystallinity index was then evaluated using two approaches: i) the peak height method and ii) the amorphous subtraction method.

The peak height method provides a crude estimate of the crystallinity index (Segal et al., 1959). Although it was only intended to provide a quick estimate of crystallinity, this technique is widely popular in the field (Thygesen et al., 2005; Park et al., 2010). On the other hand, the amorphous subtraction method is a robust empirical approach which considers that a large portion of the diffractogram pattern is derived from a broad amorphous halo. The most important factor in this approach is the choice of amorphous standard. Typically, lignin or hemicellulose is used to represent the amorphous fraction of cellulose (Thygesen et al., 2005). In the present study, a chemically derivatised form of cellulose was used (Isogai and Atalla, 1991). This is apparently the first report of using such stable amorphous cellulose to represent the amorphous fraction of wood.

Figure 14 shows representative diffractograms for each species including the reflections used in the peak-height method and the scaled amorphous profiles used in the area subtraction method. Using the peak height method, the crystallinity index of leatherwood was...
20% lower than both aspen and spruce (Table 4). Using the amorphous subtraction approach, the difference was less drastic. Nonetheless, both methods revealed that leatherwood cellulose was less crystalline than the other species. Using the amorphous subtraction method, the crystallinity index of leatherwood cellulose was found to be 0.306 compared to 0.345 for aspen and 0.373 for spruce. Since X-ray diffraction results are highly sensitive to experimental setup, Avicel was also measured as a technical control. Using both the peak height method and the amorphous subtraction method, the results for Avicel align well with those in the literature (Park et al., 2010).

Table 4. Crystallinity of xylem tissues.

<table>
<thead>
<tr>
<th>Species</th>
<th>Crystallinity Index (peak height method)</th>
<th>Crystallinity Index (amorphous subtraction method)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern leatherwood</td>
<td>0.477 ± 0.017</td>
<td>0.306 ± 0.010</td>
</tr>
<tr>
<td>Trembling aspen</td>
<td>0.605 ± 0.013</td>
<td>0.345 ± 0.017</td>
</tr>
<tr>
<td>White spruce</td>
<td>0.608 ± 0.003</td>
<td>0.373 ± 0.022</td>
</tr>
<tr>
<td>Avicel PH-101</td>
<td>0.882 ± 0.002</td>
<td>0.667 ± 0.008</td>
</tr>
</tbody>
</table>

Three replicates were analysed and standard deviations are shown alongside mean values.

3.4.7 Hypolignification in Leatherwood

In summary, this study has found that leatherwood contains remarkably low levels of lignin. While leatherwood was comparatively low in both G and S lignin, the ratio of S/G lignin was notably higher than in the reference species. Also, the lignin was more easily degraded during acid hydrolysis. This may be due to a low degree of cross-linking revealed by an abundance of free methoxy groups or simply due to a greater abundance of S-lignin. Finally, leatherwood xylan appeared to be highly acetylated and leatherwood cellulose was less crystalline in comparison to the reference species.
While cell wall chemistry is clearly an important determinant in the gross properties of wood, the interplay between biological, chemical and physical characteristics is also quite relevant. In comparison to three other broadleaf species, the extreme pliability of leatherwood was previously found to be a function of its significantly thinner cell walls, larger lumen and shorter fibres (Potzger and Geisler, 1937). Leatherwood has narrow, usually uniserate, xylem rays and a unique arrangement of interconnected vessel cells. Leatherwood xylem also contains peculiar intraxylem parenchyma cells (Choquette, 1925). It is tempting to speculate that the reduced density of cells in leatherwood could be an adaptation to hypolignification. Similarly, many cells within leatherwood cross-sections contain crystals of calcium oxalate (Holm, 1921). It is known that calcium can sometimes play an important structural role in cell walls (Demarty et al., 1984).

In addition to cell-level traits, leatherwood has adapted other secondary, or spandrel, traits that have co-evolved with hypolignification. For example, leatherwood compensates for reduced structural support in the wood by producing very strong and flexible bark. And when a branch does break, leatherwood has a remarkable propensity towards rapid wound healing and the plant subsequently grows even more vigorously (Loew, 1945). Besides providing structural integrity, lignin plays a key role in plant defence and it is often formed at the site of attack or wounding (Vance et al., 1980). Given that basal levels of lignification in leatherwood are so low, it is intriguing to consider if and how lignification is temporarily unregulated at the site of attack.

Leatherwood is a viable hypolignified species because evolution has brought it to the fulcrum of wood chemistry and compensatory secondary adaptations. Such a balance will likely be sine qua non for any successful commercial trees engineered for low lignin levels. A careful systems biology approach to the study of leatherwood could uncover a great deal about hypolignification in woody plants.
CHAPTER 4: CARBOHYDRATE OXIDASES TO ENABLE THE DERIVATISATION OF WOOD CARBOHYDRATES

With a goal of deploying carbohydrate oxidases in efforts to produce a wide range of high-value products from wood carbohydrates, I have studied the fungal enzymes galactose oxidase (GalOx) and glucooligosaccharide oxidase (GOOX). The following four sub-chapters describe my various efforts in enzyme engineering. Conclusions and recommendations are summarised in Chapter 5.

4.1 Swapping the GalOx CBM

GalOx has previously been used in the oxidation of a wide array of polysaccharides (Parikka et al., 2010). Since it contains a carbohydrate-binding module (CBM), I attempted to improve native GalOx activity on hemicellulose by swapping the native CBM for one with known binding affinity towards hemicellulose.

4.1.1 Introduction

The GalOx CBM is a member of CBM family 32. It is also a member of CBM fold family 1 as it features a β-sandwich structure (Abbott et al., 2008). And due to its flat hydrophobic binding surfaces, it is a type C CBM. Many members of this family originate from bacteria and GalOx is the only eukaryotic enzyme whose structure has been solved (CAZy Database). While there have been no direct studies of the GalOx CBM, it is believed to have affinity for galactosyls based on sequence similarity to a sialidase from *Micromonospora viridifaciens* (Gaskell, 1995; Newstead et al., 2005). Many enzymes that contain a CBM family 32 are involved in pathogenesis and interact with cell-surface exposed glycoconjugates containing galactosyls (Abbott et al., 2008). Since GalOx may play a similar role in nature (Halcrow et al., 2002), it is reasonable to expect that the GalOx CBM also binds galactosyls.

The CBM is an obvious target in efforts to improve GalOx affinity towards β(1→4) and β(1→6)-linked plant-derived polysaccharides. If the native GalOx CBM could be swapped for one with particular binding preference towards a given polysaccharide, then it follows that the
overall activity on that substrate might be improved. One mechanism by which CBMs increase overall enzyme activity is by increasing the local concentration of substrate (Bolam et al., 1998). Although GalOx is active on hemicelluloses such as galactoglucomannan, the relative activity is comparatively low and it seems likely that the native CBM is not a significant factor in this. In the present study, the native GalOx CBM was substituted with a member of CBM family 29 in an effort to improve GalOx activity on hemicelluloses that contain galactosyls. In addition, the wild-type enzyme and a variant GalOx with no CBM were also generated for comparative purposes. Like CBM family 32, CBMs from this family have a β-sandwich fold and belong to fold family 1 (Boraston et al., 2004). Also, CBM29 follows type B ligand-binding. The CBM29 used in the present study was a two-unit CBM from *Piromyces equi* that has previously been shown to have high affinity towards galactoglucomannan as well as other hemicelluloses (Freelove et al., 2001).

The methylotrophic yeast, *Pichia pastoris*, is a convenient and high-yielding expression host for recombinant protein production (Cereghino and Cregg, 2000). The pPICZα expression vector features a methanol-inducible AOX1 promoter, an upstream α-factor secretion signal sequence to direct extracellular secretion of recombinant protein and downstream myc and 6X-His tags to facilitate screening and affinity-based purification (Invitrogen, Figure 15). In addition, the pPICZα vector contains a resistance gene for the Zeocin™ antibiotic to enable selection of positive transformants in *Escherichia coli* and *P. pastoris*. Due to its ease-of-use and high yield of eukaryotic recombinant proteins, the *Pichia* system is an ideal platform for the production of GalOx variants (Figure 16).
4.1.2 Design of GalOx Variants

In carbohydrate-active enzymes, CBMs can be N-terminal or C-terminal to the catalytic domain (CD) (Shoseyov et al., 2006). Since the relative location of the CBM may be of importance for protein folding or stability, the order of CBM→CD was retained. The linker region, which connects the CBM to the CD can also be extremely important (Jeoh et al., 2008). It may provide needed flexibility, it may be glycosylated and it may align the CBM and CD so that polymeric substrates can bind and simultaneously undergo catalysis. For these reasons, the native GalOx linker was retained and the order of CBM→linker→CD was preserved. Three different expression cassettes were produced (Figure 17). GalOx variant #1 contained the CD only while variant #2 included the native CBM and the CD and variant #3 featured a CBM29 in place of the native CBM.
4.1.3 Construction of GalOx Variants

The GalOx gene was generously provided by Harry Brumer (University of British Columbia). This sequence was a codon-optimised version that was synthesized by GenScript (Piscataway, NJ, USA) for improved expression in *P. pastoris* (Spadiut et al., 2010). Similarly the CBM29 gene was also a codon-optimised sequenced produced by GenScript for this project.

All molecular cloning was completed using standard protocols with *E. coli* DH5α (Sambrook and Russell, 2001). All polymerase-chain reactions (PCR) were performed using *Pfu* Ultra II Fusion HS DNA polymerase (Agilent Technologies, Santa Clara, CA, USA). Agarose gel electrophoresis was performed using 1% agarose gels in a standard Tris-acetate-EDTA buffer system. Plasmid minipreps and DNA purifications were performed using the QIAquick Spin kits from Qiagen (Venlo, The Netherlands) and all dNTPs, restriction enzymes and ligases were obtained from New England Biolabs (Ipswich, MA, USA). Sequencing reactions were performed by The Centre for Applied Genomics (The Hospital for Sick Children, Toronto, ON). Primer sequences and reaction conditions are described in Appendix 5. Similarly, recipes for media and other reagents are included in Appendix 6.

To enable swapping of different CBM sequences, a restriction site was introduced between the linker and the CBM. This restriction site was designed in such a manner to avoid changing any amino acids in the linker region. The first two codons of the linker sequence were changed.
from GCT-TCA (Ala-Ser) to GCT-AGC (Ala-Ser) resulting in the introduction of a new NheI restriction site and no change in the amino acid sequence.

Variant #1 was produced first by PCR amplification of the GalOx CD with N-terminal EcoRI and NheI restriction sites and a C-terminal XhoI site. These restriction sites were added by appending the recognition sites to the primers used in PCR amplification. The PCR product was purified from the plasmid DNA template by excising and purifying the 1496 bp band obtained by agarose gel electrophoresis. After double restriction digests of the PCR product and the pPICZαA plasmid (3593 bp), a ligation reaction produced the intact plasmid that was transformed into *E. coli* DH5α. Thus, the expression construct was contained between the EcoRI and XhoI sites so as to minimise superfluous plasmid DNA between the signal sequence cleavage site and the stop codon. Variant #2 was produced when the native GalOx CBM (441 bp) was amplified with an N-terminal EcoRI site and a C-terminal NheI site and subsequently inserted into the plasmid carrying variant #1 by restriction digest and ligation. Finally, variant #3 was also produced from variant #1 by inserting the CBM29 sequence (915 bp) which was amplified using similar primers as with variant #2. All three variants were cloned into *E. coli* DH5α.

The three expression constructs were each verified by agarose gel electrophoresis and confirmed by DNA sequencing (Appendix 7). Figure 18 shows a triple restriction digest of the completed plasmids alongside empty pPICZαA plasmid.

**Figure 18.** Triple digest of GalOx expression constructs for CBM swapping project. Using EcoRI, NheI and Xbal. Lane 1: DNA ladder; Lane 2: pPICZαA; Lanes 3-5: GalOx variants 1-3.
4.1.4 Transformation and Expression in *P. pastoris*

The three expression constructs plus a vector-only control were transformed into the *P. pastoris* KM71H expression host according to the supplier instructions and the standard methodology (Weidner et al., 2010; Invitrogen, 2010). Plasmid DNA produced by miniprep was first linearised using *Pme*I restriction enzyme. Then, 10 μl containing 5 μg of plasmid DNA was used to transform 80 μL of electrocompetent cells. After electroshocking using 2 kV in a 2 mm electroporation cuvette, cells were recovered for two hours at 30ºC in 1 mL 1 M sorbitol. Transformed cells (~1.2 mL) were plated onto YPDS medium with Zeocin and transformants appeared as single colonies after three days incubation at 30ºC. The presence of the gene in several transformants of each variant was confirmed by colony-PCR.

4.1.5 Screening Transformants for GalOx activity

Transformants were transferred to BMM minimal media containing methanol to induce expression of GalOx variants. In addition to methanol, this media contained D-galactose (2 g L⁻¹), horseradish peroxidase (1500 U L⁻¹, Sigma-Aldrich) and ABTS (0.1 g L⁻¹, Sigma-Aldrich). Active secreted protein catalysed the oxidation of D-galactose to produce H₂O₂ (Baron et al., 1994). Horseradish peroxidase reduces H₂O₂ which transforms ABTS to a radical cation that displays a green colouration in the media (as depicted in Figure 16). Despite previous reports of this assay in the literature, peptone and yeast extract were found to be inhibitors of this reaction, even when buffered, and so were excluded from the minimal medium recipe (Figure 19). Instead of a lasting green colouration, a slow-forming purple colour was observed on plates and in reactions that contained those components.

The results of the plate-based assay are presented in Figure 20. For each variant, 75 to 100 transformants were screened. While transformants of variant #2 (wild-type gene) produced active protein, transformants of variants #1 and 3 did not appear significantly different from the vector-only controls.

---

**Figure 19. Troubleshooting the plate-based assay for GalOx activity. The 10 tubes contained the reaction components for the GalOx assay with additional media components as indicated.**

1. Control
2. Biotin
3. Phosphate buffer
4. Yeast Nitrogen Base
5. Methanol
6. Agar
7. Yeast extract
8. Peptone
9. Yeast extract + buffer
10. Peptone + buffer
4.1.6 The Native CBM is Necessary for GalOx Production

Results from the plate-based activity screens suggested that no active protein was produced from those variants which lacked the native CBM. Notably, certain transformants of variants #1 and #3, as well as the vector only control, revealed some low background activity as exemplified by slight green colouration. Alcohol oxidase naturally secreted from *P. pastoris* could account for the apparent background activity (Huang et al., 2011). To confirm that green colouration in variant #1 and variants #3 was not the result of low levels of GalOx activity, three transformants...
from each variant were cultivated in liquid inducing medium, and secreted proteins were analysed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Single colonies with the most apparent activity for each variant were each used to inoculate 200 mL of BMGY media in a 1000 mL baffled flask. After 16 hours growth at 30°C with shaking, the cells were pelleted by centrifugation at 3000 g and resuspended in 20 mL of BMMY media in a 250 mL baffled flask. This culture was incubated at 30°C with shaking and methanol was replaced to 0.5% every 24 hours. After three days, 20 μL supernatant aliquots were harvested and pelleted by centrifugation at 13,000 g. SDS-PAGE was performed according to the standard methodology using a 12% gel (Figure 21; Sambrook and Russell, 2001).

The expected size of the GalOx CD was approximately 52 kDa whereas the CBM was anticipated to be 23 kDa. Thus, the 69 kDa protein produced by variant #2 was most likely GalOx. This also agrees with results reported in the literature (McPherson et al., 1992). On the other hand, transformants of variant #1 and #3 did not produce GalOx proteins. In addition, protein secreted from colonies of variant #2 transformants contained the myc epitope as identified by an anti-myc colony blot (Western) of colonies grown on BMMY inducing media (Figure 22).

To confirm that the GalOx produced from variant #2 was catalytically active, a horseradish peroxidase-linked reaction was followed. A 5 μL aliquot of culture supernatant was added to 200 μL containing

![Figure 21. SDS-PAGE gel of culture supernatants of selected GalOx transformants. Lane 1: Marker; Lane 2: variant #2, Lane 3: variant #1; Lane 4: variant #3.](image)

![Figure 22. Colony blot of variant #2 transformants. Anti-myc antibody was used to confirm the secretion of myc-labelled protein.](image)
0.6 M d-galactose, 1 mM ABTS and 1.5 U horse-radish peroxidase in 100 mM sodium phosphate buffer. While the negative control reaction remained colourless, the reaction containing culture supernatant quickly turned green confirming the presence of active GalOx.

In summary, it was not possible to produce recombinant GalOx protein using the expression constructs that lacked the native N-terminal CBM. Since the CBM is N-terminal to the CD, it is conceivable that protein misfolding is the cause. Another possible explanation is that translation is problematic without the N-terminal sequence of the native CBM. Techniques to study mRNA production such as Northern blotting and quantitative PCR could be used to rule out the later hypothesis but were not in the present study. A previous study subjected GalOx to a mild protease treatment to fragment the protein (Mendonca and Zancan, 1988). Since this was before the structure was solved that showed the presence of binding and catalytic domains, it was surprising to see that a smaller 46 kDa fragment retained activity. In retrospect, it seems clear that this was likely the CD. Further analysis then revealed that this GalOx peptide was less stable, especially when deglycosylated. If the CBM is required for protein stability or proper folding, it is not surprising that variant #1 did not produce protein. But it is noteworthy that a replacement CBM was not adequate. This is especially interesting since both CBM29 and the native GalOx CBM belong to fold family 1.

### 4.2 Production of GalOx

Recombinant expression of GalOx without the native N-terminal CBM is apparently not possible. In order to explore the role of the CBM further, it will be necessary to isolate the CBM and CD and so I produced recombinant GalOx wild-type protein to enable future work.

#### 4.2.1 Introduction

CBM family 32, to which the GalOx CBM belongs, is a highly diverse family including bacterial, archeal and fungal members (Abbott et al., 2008). Binding affinities in this family range from galactosyls and mannosyls in cell-surface exposed glycoproteins to galactosaminyls in chitin (Ficko-Blean and Boraston, 2009). Indeed, most members of this family are involved in pathogenesis. Since CBM binding specificity normally corresponds well with the substrate
preference of the CD (Boraston et al., 2004), it is reasonable to predict that the GalOx CBM would bind galactosyls. Interestingly, the role of the GalOx CBM has not been examined despite the multitude of published enzymology studies.

While the catalytic mechanism has been the subject of considerable study, little is known about protein folding dynamics of the GalOx CD. Copper heteroatom insertion is thought to occur concurrently with pro-sequence cleavage (Rogers et al., 2000). But since heterologous production can occur successfully without the pro-sequence, cofactor assembly must be a self-processing event (Firbank et al., 2001). Also, the GalOx CD shares sequence homology with glyoxyl oxidases, which typically do not contain CBMs (Kerstein et al., 2005). Clearly, further study will be needed to fully understand the role of the CBM in GalOx.

To facilitate such future work, it was necessary to isolate the CD and CBM of GalOx. And since recombinant production of the CD was not possible, another approach was needed. In cellulases and other multi-domain proteins, limited proteolysis has been used to cleave the linker region and thereby separate CBM and CD (Fontana et al., 2004; Jeoh et al., 2008). Gel-filtration chromatography is subsequently used to isolate the resulting fragments. Recombinant GalOx was produced to enable such future studies.

### 4.2.2 Recombinant Production of Wild-type GalOx

A single colony of *P. pastoris* transformed with variant #2 was used to inoculate 10 mL of BMGY medium in a 250 mL baffled flask. After 16 hours growth with shaking at 30°C, this starter culture was used to inoculate 1000 mL of BMGY in a 4 L baffled flask. After an additional 24 hours incubation at 30°C with vigorous shaking, the cell density reached an optical density of over 2. The culture was then pelleted by centrifugation at 3000 g, the supernatant was decanted and the cells were

![Figure 23. SDS-PAGE of Pichia supernatants.](image-url)
resuspended in 200 mL of BMKY inducing media in a 1 L baffled flask. This culture was incubated at 30°C with shaking and methanol was replaced to 0.5% every 24 hours. To determine the optimum time for harvest, samples were taken each day for SDS-PAGE analysis (Figure 23).

The results from this timecourse suggested that endogenous proteases may have begun to digest GalOx protein after three days of growth. Therefore, a fresh culture was grown and harvested after three days growth. The culture was pelleted by centrifugation at 3000 g and the supernatant (200 ml) was harvested and filtered to remove all cells. The volume of the supernatant was reduced to 20 mL using a Jumbosep™ centrifugal concentrator unit (Pall Corporation, Port Washington, NY, USA) with a swinging-bucket rotor centrifuge. The media was then desalted using an FPLC unit with a HiTrap™ desalting column (GE Healthcare, Little Chalfont, UK) and an eluent of binding buffer containing 20 mM imidazole (see Appendix 6). Purified GalOx eluted first as a large peak and was followed by small peptides, salts and other small molecules (Figure 24). Fractions containing GalOx were pooled and loaded onto a GE HisTrap™ column which contained a Sepharose™ resin charged with Ni²⁺ that selectively binds 6X-HIS-tagged proteins. After washing the column with 10 column volumes of binding buffer, the GalOx protein was eluted by competitive binding using five column volumes of elution buffer containing 500 mM imidazole (Figure 25).
The fraction containing GalOx protein was buffer-exchanged with a 1000X dilution in sodium phosphate buffer using a Macrosep™ centrifugal concentrator unit (Pall Corporation). The final protein preparation was approximately 3 mL of 6 mg/mL protein as measured using the standard Bradford assay. Also, the protein prep was deemed to be quite pure by SDS-PAGE.

To confirm that the protein obtained was indeed recombinant GalOx, a Western blot was performed using anti-\textit{myc} antibody according to the standard methodology (Figure 26). Briefly, protein from an SDS-PAGE gel was transferred to a nitrocellulose membrane using a semi-dry transfer apparatus. The membrane was then incubated with shaking in blocking buffer, primary antibody (anti-\textit{myc}) and then secondary antibody (anti-mouse, alkaline phosphatase conjugated) for 1 hour each with three five-minute washing steps in between. Antibodies were diluted according to the manufacturer instructions in TBST buffer. Finally, the membrane was incubated in BCIP/NBT solution (Sigma-Aldrich) until the colour appeared.
Recombinant GalOx protein was produced for subsequent studies of the CBM and CD in isolation. Although mild protease has frequently been useful in the isolation of multi-domain proteins, each case is unique. Treatment of purified GalOx with the cysteine protease papain may require considerable trouble-shooting. Moreover, the linker region could be inaccessible to proteases due to the close association of CBM and CD. The purified GalOx CD will also facilitate studies on the role of the CBM in substrate binding.

4.3 Attempts to Engineer a Cellulose-6-oxidase

Although the native GalOx enzyme possesses no activity on D-glucose, a triple mutant that does has been reported (Sun et al., 2002). In an effort to create a novel polysaccharide oxidase that targets cellulosic substrates, I generated several GalOx variants using CBM swapping and site-directed mutagenesis. This project was done in parallel with the work described in Chapter 4.1.

4.3.1 Introduction

GalOx has been the subject of a great deal of research and there are over 1300 galactose-oxidase-related publications indexed in the ISI Web of Science database. While a large body of this work has been to study the peculiar thioether bridge in the active site, a number of projects have sought to engineer GalOx by mutagenesis (Deacon et al., 2004). One particular study generated over 10,000 variants by random mutagenesis and screened them using a variety of novel substrates (Sun et al., 2001). Through this work, a variant of GalOx was identified which possessed modest activity on the C6 primary hydroxyl of glucose (Sun et al., 2002). This so-called RQW variant contained three mutations: W290F, R330K and Q406T. All three residues are located adjacent to the active site and are thought to interact with the substrate. While these
mutations were not obvious, this study demonstrated that the GalOx active site is amenable to moderate changes (Moon et al., 2011).

Site-directed mutagenesis is a molecular technique that involves changing specific amino acids by purposefully altering the nucleotide sequence. For a gene contained in a plasmid, this can readily be achieved by performing a whole-plasmid PCR with mutagenic primers. Since the template plasmid DNA is methylated (having been isolated from Dam+ *E. coli*), it can be removed from the unmethylated DNA produced in the PCR reaction using a restriction digest with an enzyme that requires methylated sequences, such as *DpnI*. The result is a pool of plasmids carrying the desired mutation (Figure 27).

Although the RQW mutant has been tested with a range of monosaccharides and other small molecules containing primary alcohols, it has apparently never been examined with glucose-containing oligosaccharides or polysaccharides (Arnold et al., 2007). A cellulose-6-oxidase would be extremely useful since there are no existing enzymatic approaches for the derivatisation of cellulose. And so, the RQW mutant was produced through site-directed mutagenesis to test whether it could be useful in the oxidation of cellulose. Besides mutagenesis, CBM-swapping was also used in the production of five GalOx variants. In place of the native GalOx CBM, cellulose-binding CBMs were inserted. At this point, it is important to emphasize that the work described in this sub-chapter was performed in parallel with the
project described in Chapter 4.1 (i.e.: before it became clear that the native CBM is imperative for recombinant GalOx production in *P. pastoris*).

### 4.3.2 Design and Production of GalOx Variants

The pPICZ\(\alpha\) plasmid carrying the wild-type GalOx CD without the native CBM (i.e.: variant #1) was used as a template to mutagenise GalOx and produce variant #4 depicted in Figure 28. Site-directed mutagenesis was performed using the QuickChange™ II mutagenesis kit according to the manufacturer recommendations (Agilent Technologies, 2010). The three mutations were introduced in series using the mutagenic PCR primers and conditions described in Appendix 5. For example, the Trp\(^{290}\) codon was changed from TGG to TTC to produce the W290F mutation. The new codons were chosen in order to minimise the number of changes and in light of typical codon usage in *P. pastoris* to prevent expression problems due to rare codon usage by consulting the Kazusa database (Yasukazu Nakamura, Kazusa DNA Research Institute, Chiba, Japan). Following mutagenesis, the reaction mixtures were digested with *Dpn*I to remove the non-mutated parent template. DNA sequencing was used to confirm that each round of mutagenesis was successful (Appendix 7).

![Figure 28. GalOx expression constructs for cellulose-6-oxidase project.](image-url)
The remaining GalOx variants shown in Figure 28 were produced by inserting the various CBM sequences as was described in Chapter 4.1. In addition, one variant that included the native CBM was produced to serve as a control. Briefly, each CBM was amplified from the template plasmids using PCR primers which introduced an EcoRI restriction site at the 5’ end and an NheI restriction site at the 3’ end. Restriction digest, gel purification and subsequent ligation was used to produce variants #5-8.

The sequence encoding CBM11 (498 bp) was obtained from NZYTech (Lisbon, Portugal). This CBM is known to have affinity for cello-oligosaccharides. It is a type B CBM, has a β-sandwich fold structure and belongs to fold family 1 (Hashimoto, 2006). The two CBM family 1 cellulose-binding CBMs were cloned from *Phanerochaete carnosa* as part of the undergraduate thesis work of Pamuditha Silva. The CBM1 used in variant #7 (108 bp) was a CBM from a putative glycoside hydrolase from 7 (a family mostly composed of cellulases) and the CBM1 used in variant #8 (108 bp) was part of a glycoside hydrolase from family 61 (a family containing oxidative carbohydrate-active enzymes). CBM family 1 features type A ligand-binding and a cysteine-knot fold structure (fold family 3).

The expression constructs were each verified by agarose gel electrophoresis and confirmed by DNA sequencing (Appendix 7). Figure 29 shows a triple restriction digest of the completed plasmids alongside an empty pPICZα plasmid. *P. pastoris* KM71H was transformed using each plasmid as was described in Chapter 4.1.4. Single-colony transformants appeared after
three days growth on YPDS media.

4.3.3 Screening Transformants and Understanding Activity Results

A plate-based assay of GalOx transformants was performed as described in Chapter 4.1 (Figure 30). And following from the results in Chapter 4.1, the only mutant GalOx variant that showed any significant activity was the one which contained the native CBM (variant #5). The apparent activity was considerably lower than that observed with variant #2. This was anticipated since the RQW mutant is known to have 1000 times less activity on D-galactose than the wild-type enzyme (Sun et al., 2002). None of the other variants produced significant oxidation activity above the background levels. As in Chapter 4.1, several

Figure 30. Plate-based activity screen of GalOx transformants for variants #4-8. The colonies denoted with circles were vector-only control transformants.
transformants of each variant were assessed by anti-\textit{myc} colony blot and SDS-PAGE. Only variant \#5 produced GalOx protein.

\section*{4.4 Random Mutagenesis of GOOX}

Whereas random mutagenesis has been successfully used to generate carbohydrate oxidases with novel activity, I produced a library of mutant genes encoding glucooligosaccharide oxidase (GOOX). This project was done in collaboration with Pamuditha Silva.

\subsection*{4.3.1 Introduction}

Site-directed mutagenesis has often been used to study protein structure and function. While this approach is very useful in addressing specific questions in regards to structure-function relationships such as the identification of active-site residues, it is limited in its usefulness to identify which modifications could affect subtle enzyme function. In particular, choosing which residues to mutate to alter substrate tolerance may be a far more difficult question. Typically there are a number of candidate residues that may be in play. Also, changes in active site residues may not always provide the desired outcome and the residues behind the surface-exposed active-site amino acids might sometimes be better targets (Toscano et al., 2007).

GOOX has already been the subject of rational protein engineering efforts. For example, the residues involved in catalysis and cofactor binding have been identified by site-directed mutagenesis (Huang et al., 2008). Similarly, modest activity on non-natural substrates such as D-xylose and D-galactose has been introduced (Foumani et al., 2011). The GOOX active site is amenable to minor modifications which broaden the substrate tolerance. While site-directed mutagenesis has proven useful in studying and engineering GOOX in the past, more drastic change may be necessary to generate novel variants with useful biotechnological purpose. For example, a GOOX mutant with an elevated xylose oxidase activity could be very useful in the oxidation of hemicelluloses that could be used in a variety of advanced bioproducts. In addition, random mutagenesis of GOOX could produce new insights into the structure-function relationships of this enzyme and oligosaccharide oxidases more generally.
While there are many techniques for random mutagenesis, error-prone PCR is a versatile approach which can be used to introduce a high frequency of mutations. Although mutant libraries are often produced in *E. coli*, post-translational glycosylations may be important to the structure and function of GOOX. Accordingly, it was decided to recombinantly produce the mutant library in *P. pastoris*, a eukaryotic host. To that end, an error-prone library of GOOX mutants has been produced in *P. pastoris* to facilitate the genetic improvement of GOOX.

### 4.3.2 Error-prone Mutagenesis Strategy

Error-prone PCR involves the introduction of point mutations during the course of a PCR amplification reaction (Figure 31). This technique exploits the inherent error rate of polymerases by using a *Taq* polymerase that lacks error-checking. By altering the PCR reaction conditions, it is possible to modulate the frequency of mutation (Arnold and Georgiou, 2003). This is normally done by changing the concentration of Mg$^{2+}$ ions or the relative proportion of dNTPs. Such approaches typically result in a bias for adding one dNTP over another. On the other hand, commercial kits are available that balance the mutation bias and allow one to tweak the mutation frequency.

The GeneMorph™ II kit (Agilent Technologies) allows for the frequency of mutation to be altered by adjusting the concentration of template DNA. Typical mutation frequencies achieved...
range from 1-3 mutations/kb with high amounts of template DNA and up to 16 mutations/kb with low amounts of template DNA. Primers are designed around the sequence which is to be mutated and the error-prone PCR is performed using an error-prone Taq polymerase. The product of the error-prone reaction can then be used as a template for whole-plasmid PCR to regenerate plasmid DNA. Finally, the original template DNA, which is methylated from *E. coli*, is removed by *DpnI* digest.

### 4.3.3 Production of GOOX Mutant Library

The GOOX gene from *Acremonium strictum* was previously cloned and inserted into the pPICZα plasmid for recombinant expression in *P. pastoris* (Foumani et al., 2011). The GOOX sequence used in the present study was produced from cDNA prepared from *Acremonium strain CBS 346.70*. The 1422 bp gene contains no introns, but did contain 15 polymorphisms compared to the GOOX gene which has previously been reported (Lin et al., 2005). As in Chapter 4.1, the expression vector included the α-factor signal sequence that directs peptide export from the cell as well C-terminal 6X-HIS and myc tags which facilitate protein detection and affinity-based purification. The plasmid was isolated from a miniprep of a glycerol stock and its sequence was confirmed by sequencing (Appendix 7).

Error prone PCR was performed using the GeneMorph™ II kit (Agilent Technologies) according to the manufacturer instructions. The primers used in the mutagenic reaction spanned the entire gene so that the whole coding sequence was mutated. A template concentration of 12 ng/μL was used to achieve an expected average error rate of 4-8 mutations per 1000 bp. The resulting PCR product was confirmed by agarose gel electrophoresis and DNA sequencing confirmed that random mutagenesis occurred with no obvious bias. Whole plasmid PCR was performed using the product of the error-prone PCR reaction and the wild-type plasmid. Following PCR, a six-hour *DpnI* digest was used to remove the template plasmid leaving only mutant plasmids.

The plasmid pool was subsequently transformed in library efficiency *E. coli* cells, NEB Turbo high efficiency competent cells, according to the manufacturer’s instructions (New
England Biolabs). Transformed cells were aseptically spread onto five 150 mm Petri dishes of LSLB-agar with Zeocin. After 16 hours of incubation at 37°C, uniform and well-isolated single colonies were obtained. A sterile cell-scaper was used with 10 mL of liquid LB medium to collect the cells on each plate. Pooled cells (50 mL total) were subjected to plasmid DNA extraction using 25 separate miniprep extractions performed according to the manufacturer instructions. A total of 500 μL of deionised water was used to elute the plasmid DNA from the spin columns and the purity and concentration was verified by agarose gel and UV absorbance. A total of 200 mg of mutant library plasmid DNA was obtained and was stored at -20°C.

**4.3.4 Pichia Transformation for Future Screening**

This mutant library could be used to transform *P. pastoris* KM71H according to the standard methodology as described in Chapter 4.1. After three days on YPDS-Zeocin plates, transformants should be expected to appear as single colonies. The transformants of this GOOX library should then be used screening projects. For example, a plate-based assay could be used to screen for GOOX variants with improved activity on D-xylose. Colonies would be replicated onto transfer YPD plates and BMMY media which contain methanol to induce expression. After an additional three days of growth at 30°C, a solution of 0.3 % molten agarose containing 1 M D-xylose, 2 mM phenol, 0.4 mM 4-aminoantipyrine and 15 U/mL in 40 mM Tris-HCl pH 8.0 would be overlaid onto the BMMY plate containing colonies. Positive activity would be assessed by the formation of a pink colour near a given colony.
CHAPTER 5: SUMMARY AND RECOMMENDATIONS

This chapter includes brief summaries of the results from Chapter 3 and 4 and provides recommendations for future work.

5.1 Low-lignin Feedstocks

Through this project, xylem from eastern leatherwood was found to contain remarkably low levels of lignin in comparison to two reference species. The acid-insoluble lignin content was found to be 13.8 \% by mass whereas the acid-soluble fraction was 1.7 \%. Through FTIR and ToF-SIMS analysis, leatherwood lignin was found to be relatively rich in syringyl units. It is tempting to speculate that the greater hydrolysability of leatherwood lignin may be due to a lower degree of cross-linking. The middle lamella was not found to be rich in lignin although such a scenario cannot be ruled out. Finally, the crystallinity index of leatherwood cellulose was comparatively low and the xylan appeared to be highly acetylated.

While its small stature and slow growth make leatherwood impractical for use in bioenergy applications, low-lignin plants such as leatherwood are ideal model species for plant engineers trying to develop viable low-lignin genotypes (Neutelings, 2011). Besides inspiring efforts to engineer lignification directly, leatherwood features many secondary traits that have co-evolved with hypolignification and a thorough understanding of this balance will be indispensable for future commercial low-lignin trees. The following three recommendations provide a framework for future studies that will further explore the basis for hypolignification in leatherwood:

1. This study found that while leatherwood is deficient in lignin. In addition, the lignin is more readily hydrolysable and contains a greater proportion of syringyl units. NMR studies could confirm the proportion of syringyl units and could be useful in assessing the extent of cross-linking. Björkman lignin is more representative of actual lignin in the cell wall and may be a better candidate for such work than Klason (acid-insoluble) lignin.

2. The ToF-SIMS spectra suggest that lignin accumulation may not occur in the middle lamellae of leatherwood xylem. If true, the implication may be that whatever monolignols
are exported from the cell undergo deposition on to the cell wall rather than contributing
to the pool of lignin in the middle lamella. While a lack of lignin in the middle lamellae
could also simply be a function of lower total levels everywhere, the distribution of lignin
across the cell wall and middle lamella should be further examined by bromination and
electron microscopy.

3. Since the ratio of syringyl to guaiacyl lignin is an important factor in biomass
recalcitrance, the genetic basis of lignification in leatherwood should be studied. A
careful analysis of the transcriptome of the vascular cambium could reveal which genes
in the monolignol biosynthetic pathway control the production of S and G monolignols in
developing xylem. Special attention should also be given to glucosyltransferases,
peroxidases and laccases which could be involved in monolignol export and
polymerisation.

Efforts are underway to develop viable hypolignified commercial genotypes and further
study of leatherwood may guide those efforts. Several theories abound on the regulation of
lignification. While regulation at the level of monolignol biosynthesis is the most popular theory,
it is conceivable that monolignol export and deposition could also be regulated. Low-lignin
model plants such as leatherwood could help unlock these secrets. Moreover, compensatory traits
on display in leatherwood will also inspire future plant engineers in efforts to produce viable
hypolignified trees.

5.2 Carbohydrate Oxidases

Carbohydrate oxidases such as galactose oxidase (GalOx) and glucooligosaccharide oxidase
(GOOX) hold tremendous promise for the oxidation of carbohydrates in the production of high-
value bioproducts. GalOx is particularly attractive in light of its open active site which accepts
bulky substrates. On the other hand, GOOX permits a wide range of oligosaccharide substrates
and produces carboxylic acid derivatives.
The native carbohydrate-binding module (CBM) of GalOx is indispensable in recombinant production of GalOx in *P. pastoris*. The CBM and catalytic domains (CD) are joined by a short linker and are also associated by non-covalent interactions. The stability of the GalOx CD may be greatly reduced upon the removal of the CBM (Mendonca and Zancan, 1988). In the present study, GalOx variants that lacking the native CBM could not be produced in *P. pastoris*. In eukaryotes, unstable proteins that are poorly structured are targeted for proteolysis. Thus, it is conceivable that GalOx peptides were produced by *P. pastoris* transformants but were destroyed before maturation and export. In order to study the role of the GalOx CBM, a recombinant GalOx was produced from *P. pastoris* and purified by affinity-based protein chromatography. The following recommendations follow from the work with GalOx:

1. Subsequent partial digestion by the cysteine protease papain should be used to sever the linker region and facilitate the isolation of the CBM and CD by gel-filtration chromatography. The CD could be useful in comparative activity assays to indirectly evaluate the role of the CBM in activity. The CBM could be used in isothermal calorimetry studies to monitor the energy of substrate binding.

2. *Use in silico* modelling with the GalOx structure and site-directed mutagenesis to probe the significance of amino acids suspected to participate in CBM-CD interactions. While no residues are apparent that would be involved in hydrophobic interactions, there are a number of residues that may participate in hydrogen bonding between the CBM and CD.

3. Determine the effect of linker length and sequence on CBM-CD interaction and GalOx activity. Given the apparent role of the CBM in enzyme stability, GalOx could be an ideal enzyme for assessing the role of the linker sequence and length in carbohydrate-active enzymes that contain CBMs.

4. The eight GalOx expression cassettes could be used for future work in GalOx engineering. In those constructs which lack the native CBM, it could be inserted by using the *NheI* restriction site which remains in front of the CD coding sequence. A previous report in GalOx engineering added an additional CBM at the C-terminal (Lis and
Kuramitsu, 1997) and it remains to be seen whether the addition of an N-terminal CBM would produce functional protein.

As with GalOx, GOOX could be useful in a variety of biotechnological purposes. While the wild-type enzyme is capable of oxidising a broad range of substrates, activity on D-xylose, D-galactose and L-arabinose is limited (Lin et al., 1991). Protein engineering efforts have already been successful in expanding GOOX substrate range (Foumani et al., 2011). The new mutant library of GOOX variants may reveal mutations that produce even greater activity with non-natural substrates. The following recommendations follow from the GOOX work:

1. The GOOX mutant library should be screened using the plate-based assay with non-natural substrates such as D-galactose, D-xylose and L-arabinose. Mutants with apparent activity should then be screened using hemicellulose-derived oligosaccharides that contain these glycosyls.

2. *In silico* modelling using the published GOOX structure and the information gleaned from DNA sequencing of the promising GOOX mutants may inspire future protein engineering work.

Carbohydrate-active enzymes will be the lynchpins in future biorefineries. Aside from cellulases and xylanases, carbohydrate oxidases may play a pivotal role in economising such biomass utilisation processes. Through oxidation of plant cell wall-derived carbohydrates, these enzymes could enable further derivatisation and the production of a wide array of high-value bioproducts. Therefore, continued protein engineering efforts with GalOx and GOOX will be highly valuable.
REFERENCES


Bigelow, J. (1818) American Medical Botany. Cummings and Hilliard, Boston, MA. USA.


Fenerty C, 1844. A letter from Fenerty to the Acadian Recorder newspaper dated October 26, 1844.


TAPPI (2002a) Test method T222: Acid-insoluble lignin in wood and pulp. In *Tappi Test Methods*. Atlanta, GA.

TAPPI (2002b) Test method T250: Acid-soluble lignin in wood and pulp. In *Tappi Test Methods*. Atlanta, GA.


Valliant, S. 1708. *Histoire des plantes de Canada*. Excerpts of letters from Michel Sarrazin in which he describes the plants sent to France from 1698 to 1707. A copy is available in the archives of Bernard Boivin at Laval University, Quebec.


### APPENDIX 1: PEAK LIST FOR FTIR

<table>
<thead>
<tr>
<th>Wavenumber (cm⁻¹)</th>
<th>Description</th>
<th>Reference</th>
<th>Additional Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>3600-3300</td>
<td>H-bonded OH stretching</td>
<td>Marchessault, 1962</td>
<td></td>
</tr>
<tr>
<td>3570-3350</td>
<td>H-bonded and O-H group stretching</td>
<td>Fengel and Ludwig, 1991</td>
<td>cellulose and polysaccharides</td>
</tr>
<tr>
<td>3460-3412</td>
<td>O-H stretch</td>
<td>Fengel and Ludwig, 1991</td>
<td>cellulose and polysaccharides</td>
</tr>
<tr>
<td>3400-3390</td>
<td>OH stretching (H-bonded)</td>
<td>Hergert, 1971</td>
<td>conifer (Björkman lignin)</td>
</tr>
<tr>
<td>3400-3380</td>
<td>H-bond oriented OH stretching</td>
<td>Fengel and Ludwig, 1991</td>
<td>cellulose and polysaccharides</td>
</tr>
<tr>
<td>3419</td>
<td>O-H stretch (hydrogen bond)</td>
<td>Pandey, 1999</td>
<td>conifer (Klason lignin)</td>
</tr>
<tr>
<td>3414</td>
<td>O-H stretch (hydrogen bond)</td>
<td>Pandey, 1999</td>
<td>broadband (Klason lignin)</td>
</tr>
<tr>
<td>3348</td>
<td>O-H stretch (hydrogen bond)</td>
<td>Pandey, 1999</td>
<td>from cellulose</td>
</tr>
<tr>
<td>2981-2939</td>
<td>Asymmetrical CH₂ stretching</td>
<td>Fengel and Ludwig, 1991</td>
<td>cellulose and polysaccharides</td>
</tr>
<tr>
<td>2900</td>
<td>CH₂ antisymmetric stretching</td>
<td>Marchessault, 1962</td>
<td></td>
</tr>
<tr>
<td>2942</td>
<td>OH-stretch in methyl and methylene groups</td>
<td>Hergert, 1971</td>
<td>broadband (Björkman lignin)</td>
</tr>
<tr>
<td>2935</td>
<td>C-H stretching</td>
<td>Pandey, 1999</td>
<td>conifer (Klason lignin)</td>
</tr>
<tr>
<td>2920</td>
<td>OH-stretch in methyl and methylene groups</td>
<td>Hergert, 1971</td>
<td>broadband (Björkman lignin)</td>
</tr>
<tr>
<td>2919</td>
<td>C-H stretching</td>
<td>Pandey, 1999</td>
<td>broadband (Klason lignin)</td>
</tr>
<tr>
<td>2914-2970</td>
<td>OH stretching</td>
<td>Marchessault, 1962</td>
<td></td>
</tr>
<tr>
<td>2904-2850</td>
<td>Symmetrical CH₂ stretching</td>
<td>Fengel and Ludwig, 1991</td>
<td>cellulose and polysaccharides</td>
</tr>
<tr>
<td>2902</td>
<td>C-H stretching</td>
<td>Pandey, 1999</td>
<td>from cellulose</td>
</tr>
<tr>
<td>2850</td>
<td>C-H stretching</td>
<td>Pandey, 1999</td>
<td>broadband (Klason lignin)</td>
</tr>
<tr>
<td>3000-2842</td>
<td>C-H stretch in methyl and methylene groups</td>
<td>Faix, 1991</td>
<td>multiple peaks within this range (3000-3002, 2937-2942, 2979-2980, 2940-2960) (Björkman lignin)</td>
</tr>
<tr>
<td>3402</td>
<td>C-H stretching</td>
<td>Faix, 1991</td>
<td>conifer (Klason lignin)</td>
</tr>
<tr>
<td>2875-2850</td>
<td>OH-stretch in methyl and methylene groups</td>
<td>Hergert, 1971</td>
<td>broadband (Björkman lignin)</td>
</tr>
<tr>
<td>2845-2835</td>
<td>OH-stretch in methyl and methylene groups</td>
<td>Hergert, 1971</td>
<td>broadband (Björkman lignin)</td>
</tr>
<tr>
<td>2830</td>
<td>OH-stretch in methyl and methylene groups</td>
<td>Hergert, 1971</td>
<td>conifer (Björkman lignin)</td>
</tr>
<tr>
<td>1730-1709</td>
<td>C=O stretch in unconjugated ketones, carboxyls and in ester groups (frequently of carbohydrate origin); conjugated aldehydes and carboxylic acid absorb around and below 1700 cm⁻¹</td>
<td>Faix, 1991</td>
<td>conifer=1722, broadband=1735, herbaceous=1709 (Björkman lignin)</td>
</tr>
<tr>
<td>1730-1725</td>
<td>CHO stretching of acetyl or carboxylic acid</td>
<td>Marchessault, 1962</td>
<td></td>
</tr>
<tr>
<td>1725</td>
<td>C=O stretching of acetyl or COOH group</td>
<td>Fengel and Ludwig, 1991</td>
<td>cellulose and polysaccharides</td>
</tr>
<tr>
<td>1714</td>
<td>C=O stretching (unconjugated)</td>
<td>Pandey, 1999</td>
<td>conifer (Björkman lignin)</td>
</tr>
<tr>
<td>1711</td>
<td>C=O stretching</td>
<td>Pandey, 1999</td>
<td>broadband (Klason lignin)</td>
</tr>
<tr>
<td>1715</td>
<td>C=O stretching – unconjugated ketone and carboxylic groups</td>
<td>Hergert, 1971</td>
<td>conifer (Björkman lignin)</td>
</tr>
<tr>
<td>1715-1710</td>
<td>C=O stretching - unconjugated ketone and carboxylic groups</td>
<td>Hergert, 1971</td>
<td>broadband (Björkman lignin)</td>
</tr>
<tr>
<td>1670</td>
<td>Lignin</td>
<td>Marchessault, 1962</td>
<td></td>
</tr>
<tr>
<td>1675-1660</td>
<td>C=O stretching - para substituted ary ketone</td>
<td>Hergert, 1971</td>
<td>conifer (Björkman lignin)</td>
</tr>
<tr>
<td>1660</td>
<td>C=O stretching - para substituted ary ketone</td>
<td>Hergert, 1971</td>
<td>broadband (Björkman lignin)</td>
</tr>
<tr>
<td>1675-1655</td>
<td>C=O stretch in conjugated p-subst. ary ketones; strong electronegative substituents lower the wavenumber</td>
<td>Faix, 1991</td>
<td>conifer=1693, broadband=1658 (Björkman lignin)</td>
</tr>
<tr>
<td>1640</td>
<td>Absorbed O-H; conjugated C=O</td>
<td>Pandey, 1999</td>
<td>from cellulose</td>
</tr>
<tr>
<td>1635</td>
<td>Absorbed water</td>
<td>Fengel and Ludwig, 1991</td>
<td>cellulose and polysaccharides</td>
</tr>
<tr>
<td>1635</td>
<td>Absorbed water</td>
<td>Marchessault, 1962</td>
<td></td>
</tr>
<tr>
<td>1510</td>
<td>asymmetric skeletal vibration + C=O stretching</td>
<td>Pandey, 1999</td>
<td>broadband (Klason lignin)</td>
</tr>
<tr>
<td>1510</td>
<td>asymmetric skeletal vibration + C=O stretching</td>
<td>Pandey, 1999</td>
<td>from cellulose</td>
</tr>
<tr>
<td>1400</td>
<td>COO⁻ ion</td>
<td>Marchessault, 1962</td>
<td></td>
</tr>
<tr>
<td>1460</td>
<td>C-H deformation</td>
<td>Pandey, 1999</td>
<td>conifer (Björkman lignin)</td>
</tr>
<tr>
<td>1460</td>
<td>Lignin and CH₂ symmetrical bending on pyran ring</td>
<td>Marchessault, 1962</td>
<td></td>
</tr>
<tr>
<td>1455-1400</td>
<td>OH-in-plane bending</td>
<td>Marchessault, 1962</td>
<td></td>
</tr>
<tr>
<td>1430</td>
<td>asymmetric skeletal vibrations</td>
<td>Hergert, 1971</td>
<td>conifer (Björkman lignin)</td>
</tr>
<tr>
<td>1430</td>
<td>CH₃ symmetrical bending modes of hydroxymethyl</td>
<td>Marchessault, 1962</td>
<td></td>
</tr>
<tr>
<td>1430</td>
<td>C-H deformation (symmetric)</td>
<td>Pandey, 1999</td>
<td>from cellulose</td>
</tr>
<tr>
<td>1340</td>
<td>CH₂ deformation</td>
<td>Pandey, 1999</td>
<td>conifer (Klason lignin)</td>
</tr>
<tr>
<td>1370</td>
<td>O-H in-plane deformation</td>
<td>Pandey, 1999</td>
<td>from cellulose</td>
</tr>
<tr>
<td>1330-1315</td>
<td>OH deformations</td>
<td>Fengel and Ludwig, 1991</td>
<td>cellulose and polysaccharides</td>
</tr>
<tr>
<td>1335-1315</td>
<td>CH₂ wagging</td>
<td>Marchessault, 1962</td>
<td></td>
</tr>
<tr>
<td>1315</td>
<td>C=O of pyrimidine ring</td>
<td>Pandey, 1999</td>
<td>broadband (Klason lignin)</td>
</tr>
<tr>
<td>1282-1277</td>
<td>CH deformations</td>
<td>Fengel and Ludwig, 1991</td>
<td>cellulose and polysaccharides</td>
</tr>
</tbody>
</table>
1035-1030 aromatic C-H in-plane deformation, G>S; plus C-O deformation in primary alcohols; plus C=O stretching Fengel and Ludwig, 1991 cellulose and polysaccharides

1060-1015 C-O stretching Fengel and Ludwig, 1991 cellulose and polysaccharides

1230-1221 C-C

1128-1125 aromatic C-H in-plane deformation; typical for S units; plus secondary alcohols plus C=O stretching Fengel and Ludwig, 1991 cellulose and polysaccharides

1110-1107 asymmetrical ring stretching Fengel and Ludwig, 1991 cellulose and polysaccharides

1205-1200 OH group deformations Fengel and Ludwig, 1991 cellulose and polysaccharides

1162-1125 asymmetrical streching of C-O-C group Fengel and Ludwig, 1991 cellulose and polysaccharides

1270-1266 G ring and side group vibrations C-C, C-OH, C-H Ka

990-966 -HC=CH- out-of-plane deformation (trans) Faix, 1991 ? (Björkman lignin)

925-915 C-H out-of-plane; aromatic Faix, 1991 conifer=919, broadleaf=925 (Björkman lignin)

930-925 pyranose ring stretching Fengel and Ludwig, 1991 cellulose and polysaccharides

895-892 stretching and formation of anomeric C-H group Pandey, 1999 from cellulose

835-834 C-H out-of-plane in position 2 and 6 of S and in all positions of H units Faix, 1991 broadleaf=835, herbaceous=834 (Björkman lignin)

858-853 C-H out-of-plane in position 2, 5 and 6 of G units Faix, 1991 broadleaf=858, herbaceous=853 (Björkman lignin)

832-817 C-H out-of-plane in position 2, 5 and 6 of G units Faix, 1991 broadleaf=832, herbaceous=817 (Björkman lignin)

770-750 aromatic C-H out-of-plane deformations Herrgert, 1971 broadleaf and conifer (Björkman lignin)

768 Mode due to arabinogalactan Pandey, 1999 from cellulose

700-650 OH out-of-plane twinning Marchessault, 1962
## APPENDIX 2: PEAK LIST FOR ToF-SIMS

<table>
<thead>
<tr>
<th>m/z value</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.023</td>
<td>CH₃⁺ from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>19.019</td>
<td>H3O⁺ from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>27.024</td>
<td>C₂H₃⁺ possibly from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>29.003</td>
<td>CHO⁺ possibly from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>29.040</td>
<td>C₂H₅⁺ possibly from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>30.037</td>
<td>CH₄N from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>31.019</td>
<td>CΗ₃O⁺ from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>39.022</td>
<td>C₃H₃⁺ possibly from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>41.039</td>
<td>C₃H₅⁺ possibly from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>42.009</td>
<td>C₂H₂O⁺ possibly from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>42.042</td>
<td>C₃H₆⁺ possibly from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>43.019</td>
<td>C₂H₃O⁺ possibly from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>43.055</td>
<td>C₃H₇⁺ possibly from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>44.023</td>
<td>C₄H₃O⁺ from polysaccharide</td>
<td>Dell, 1987</td>
</tr>
<tr>
<td>44.053</td>
<td>C₂CH⁺ or C₂H6⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>45.034</td>
<td>C₂H₆O⁺ from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>47.013</td>
<td>CH₃O⁺ from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>51.021</td>
<td>CH₃⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>53.038</td>
<td>C₄H₅⁺ possibly from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>55.019</td>
<td>C₃H₃O⁺ possibly from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>55.056</td>
<td>C₄H₇⁺ possibly from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>56.058</td>
<td>C₃H₈N⁺ or C₄H₄⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>57.035</td>
<td>C₃H₅O⁺ possibly from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>57.073</td>
<td>C₄H₆⁺ possibly from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>58.071</td>
<td>C₃H₈N⁺ or C₄H₁₀⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>59.015</td>
<td>C₂H₃O₂⁺ from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>60.021</td>
<td>C₂H₄O₂⁺ from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>63.022</td>
<td>C₅H₃⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>65.038</td>
<td>C₅H₅⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>67.006</td>
<td>C₅H₇⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>68.060</td>
<td>C₆H₈N⁺ or C₅H₈⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>69.995</td>
<td>C₃H₂O₂⁺ from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>69.036</td>
<td>C₄H₅⁺ from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>69.074</td>
<td>C₃H₅⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>70.074</td>
<td>C₄H₈N⁺ or C₅H₁₀⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>71.014</td>
<td>C₃H₃O₂⁺ from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>72.090</td>
<td>C₄H₁₀N⁺ or C₅H₁₂⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>73.032</td>
<td>C₃H₅O⁺ from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>77.037</td>
<td>C₆H₅⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>79.056</td>
<td>C₆H₇⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>81.036</td>
<td>C₅H₅O⁺ from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>81.075</td>
<td>C₆H₉⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>83.013</td>
<td>C₄H₂O₂⁺ from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>85.034</td>
<td>C₄H₂O₂⁺ from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>86.108</td>
<td>C₆H₁₁⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>87.051</td>
<td>C₄H₂O₂⁺ from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>90.014</td>
<td>C₇H₁₀⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>90.037</td>
<td>C₅H₇O⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>95.082</td>
<td>C₇H₉⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>97.032</td>
<td>C₅H₂O₂⁺ from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>99.049</td>
<td>C₅H₂O₂⁺ from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>101.029</td>
<td>C₆H₃O⁺ from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>103.045</td>
<td>C₄H₂O₃⁺ or C₈H₇⁺ possibly from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>105.071</td>
<td>C₇H₁₀⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>107.050</td>
<td>C₇H₁₀⁺ from lignin</td>
<td>Saito et al., 2006</td>
</tr>
<tr>
<td>107.044</td>
<td>C₇H₁₀⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>109.033</td>
<td>C₆H₂O₂⁺ from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>113.026</td>
<td>C₅H₃O⁺ from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>115.040</td>
<td>C₅H₂O⁺ from xylan</td>
<td>Dell, 1987</td>
</tr>
<tr>
<td>115.045</td>
<td>C₅H₃O⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>121.029</td>
<td>C₆H₂O₂⁺ from lignin</td>
<td>Saito et al., 2006</td>
</tr>
<tr>
<td>127.040</td>
<td>C₇H₁₀⁺ from cellulose</td>
<td>Dell, 1987</td>
</tr>
<tr>
<td>127.041</td>
<td>C₆H₁₀O⁺ from cellulose</td>
<td>Dell, 1987</td>
</tr>
<tr>
<td>128.051</td>
<td>C₆H₂O⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>131.049</td>
<td>C₇H₂O₂⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>133.050</td>
<td>C₅H₂O⁺ from xylan</td>
<td>Dell, 1987</td>
</tr>
<tr>
<td>137.060</td>
<td>C₅H₂O⁺ from lignin</td>
<td>Saito et al., 2005</td>
</tr>
<tr>
<td>137.063</td>
<td>C₅H₂O⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>145.050</td>
<td>C₆H₂O⁺ from cellulose</td>
<td>Dell, 1987</td>
</tr>
<tr>
<td>145.061</td>
<td>C₅H₂O⁺ from polysaccharide</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>147.068</td>
<td>C₇H₂O⁺ or C₆H₁₀O⁺ or C₁₀H₁₁O⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>151.040</td>
<td>C₇H₁₀O⁺ from lignin</td>
<td>Saito et al., 2005</td>
</tr>
<tr>
<td>151.049</td>
<td>C₇H₁₀O⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>152.050</td>
<td>C₅H₂O⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>153.049</td>
<td>C₅H₂O⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>165.059</td>
<td>C₅H₂O⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
<tr>
<td>167.071</td>
<td>C₇H₁₁O⁺ from lignin</td>
<td>Saito et al., 2005</td>
</tr>
<tr>
<td>181.050</td>
<td>C₇H₂O⁺ from lignin</td>
<td>Saito et al., 2005</td>
</tr>
<tr>
<td>189.059</td>
<td>C₁₁H₂O⁺ from lignin</td>
<td>Goacher et al., 2011</td>
</tr>
</tbody>
</table>
To complement the analysis of leatherwood holocellulose, I also employed high-performance liquid chromatography (HPLC). These data are presented here rather than in the main text since it was the result of collaborative work and since there were several problems encountered during data analysis.

**A1.1 Methodology**

Holocellulose composition was evaluated by HPLC in collaboration with Dr. D. Jeremic using a CarboPac™ PA1 anion-exchange column (2 x 250 mm) and an ICS-3000 system (Dionex Corporation, Sunnyvale, California, USA) equipped with an electrochemical detector and a guard column. Samples of acid-hydrolysis filtrate were diluted so that the glucose content was roughly 100 ppm. Good separation of the main holocellulose sugars was obtained using an injection volume of 3 μL and an eluent of 3 mM sodium hydroxide at a flow rate of 3 mL min\(^{-1}\). Between samples, a column clean-up was performed using 200 mM sodium hydroxide. Calibration standards of known concentration enabled quantification. In addition, sugar recovery standards were processed alongside powdered wood samples in order to correct for loss upon hydrolysis. L-fucose was added to all samples as an internal standard.

**A1.2 Results and Discussion**

HPLC was successfully used to resolve arabinose, galactose, glucose, xylose and mannose in the acid hydrolysate. Table A3 shows the ratios of the main wood sugars relative to glucose. The levels of arabinose and galactose were routinely below the lower limits of the standard curve. Similarly, mannose levels in leatherwood and aspen and xylose levels in spruce were outside the range of the standard curve. The peak areas from calibration standards were used to prepare standard curves for quantification. All results were first normalized using the L-fucose internal standard.
Table A3. Holocellulose composition.

<table>
<thead>
<tr>
<th>Species</th>
<th>Ara/Rha : Glu</th>
<th>Gal : Glu</th>
<th>Xyl : Glu</th>
<th>Man : Glu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern leatherwood</td>
<td>0.025 ± 0.006</td>
<td>0.019 ± 0.001</td>
<td>0.630 ± 0.006</td>
<td>0.004 ± 0.001</td>
</tr>
<tr>
<td>Trembling aspen</td>
<td>0.013 ± 0.012</td>
<td>0.008 ± 0.002</td>
<td>0.371 ± 0.019</td>
<td>0.022 ± 0.008</td>
</tr>
<tr>
<td>White spruce</td>
<td>0.033 ± 0.011</td>
<td>0.030 ± 0.002</td>
<td>0.135 ± 0.003</td>
<td>0.268 ± 0.012</td>
</tr>
</tbody>
</table>

Three replicates were analysed and standard deviations are shown alongside mean values. The values shown in italics are based on only two replicates since the third value was below the detection limit.

Problems with dilution levels and the standard curves mean that these data should be considered with caution. Nonetheless, it is clear that leatherwood contained considerable more xylose than spruce or even aspen. Similarly, leatherwood has very little mannose. The relative ratios of the different sugars for aspen and spruce correspond reasonably well with the literature values (Pettersen, 1984).
APPENDIX 4: CELLULOSE CRYSTALLINITY EVALUATED BY FTIR

In Chapter 3, a curve fitting procedure was performed on FTIR spectra of lignin to assess the S/G lignin ratio. A similar procedure was performed using the holocellulose spectra in order to assess crystallinity.

A2.1 Methodology

Peak deconvolution was performed as was done for the lignin spectra. Peaks were identified using the second derivative scanning mode of OPUS version 5.1 with 9 smoothing points and curve fitting was performed with these peaks modelled using a pseudo-Voigt shape. The Levenberg-Maquardt algorithm was performed with the peak positions fixed and peak intensity, width and shape unfixed. The crystallinity index was assessed using the ratio of deconvoluted peak heights at 1730cm⁻¹ and 895cm⁻¹ (Kataoka and Kondo, 1998).

A2.1 Results and Discussion

Figure A4 shows a representative spectrum of leatherwood holocellulose with the resulting deconvoluted peaks. Unlike the curve fitting results for the lignin spectra in Chapter 3, these results should be viewed with caution since the resulting peaks are quite broad and poorly defined. Indeed many of the peak position that were used as input in the fitting procedure lead to zero-intensity peaks after the peak fitting iterations. This is largely due to the poorly defined nature of the input spectrum. Compared the lignin spectra, the holocellulose spectra are clearly composed of broader and more overlapping peaks. This is due in part to residual lignin which remains in the holocellulose preparation in as much as 4% (Browning, 1967). But for the most part, holocellulose spectra are composed of broad overlapping peaks which do not lend themselves well to such a curve-fitting procedure.
Table A3 presents the crystallinity index values calculated using the FTIR spectra of holocellulose. These data are presented in an Appendix since it should be viewed with caution. The very large standard deviation values, especially for leatherwood, represent the inadequacy of the curve fitting procedure applied to these spectra. Nonetheless, these data suggest that spruce possesses a much higher crystallinity index than the other two species. Such a result is in accordance with the XRD measurements described in Chapter 3. Also, since the ratio of these peak intensities describes the lateral order cellulose microfibrils, a larger value could correspond to thicker microfibrils (Åkerholm et al., 2004).

Table A3. Crystallinity Index Calculated by FTIR.

<table>
<thead>
<tr>
<th>Species</th>
<th>Crystallinity Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern leatherwood</td>
<td>1.11 ± 1.08</td>
</tr>
<tr>
<td>Trembling aspen</td>
<td>0.55 ± 0.22</td>
</tr>
<tr>
<td>White spruce</td>
<td>6.81 ± 1.79</td>
</tr>
</tbody>
</table>

Three replicates were analysed and standard deviations are shown alongside mean values.
### APPENDIX 5: PRIMERS AND PCR CONDITIONS

<table>
<thead>
<tr>
<th>Reaction description</th>
<th>PCR primers / Reaction conditions</th>
</tr>
</thead>
</table>
| PCR amplification of GalOx linker + CD (with N-terminal EcoRI and NheI sites and C-terminal XbaI site) | Forward primer, YM-GALOXENG-PR1: 5'-CCCGAATTCTCAGCTTACACAGCAGTCCACAG-3  
Reverse primer, YM-GALOXENG-PR11: 5'-CCCTCTAGAGCTTGAGTAACTCTAATTGTTGAAG-3  
1 μL of plasmid template DNA, 1 μL each of primer (100 μM), 0.5 μL dNTP mix (25 mM each), 2.5 μL of 10X reaction buffer, 18.5 μL of dH2O, 0.5 μL Pfu DNA polymerase.  
95°C for 5 mins; Repeat 5 times: 95°C for 60 s, 50°C for 40 s, 72°C for 30 s; Repeat 25 times: 95°C for 60 s, 55°C for 40 s, 72°C for 30 s; 72°C for 5 mins. |
| PCR amplification of GalOx CBM (with N-terminal EcoRI site and C-terminal XbaI site) | Forward primer, YM-GALOXENG-PR12: 5'-CCCGAATTCGCCCTCTGCTCCAATTGGTTC-3  
Reverse primer, YM-GALOXENG-PR4: 5'-CCCGCTAGCTTGAAAAACGTTGATCTCAGC-3  
As described for PCR of GalOX linker + CD. |
| PCR amplification of CBM29 (with N-terminal EcoRI site and C-terminal XbaI site) | Forward primer, YM-GALOXENG-PR24: 5'-CCCGAATTCAAATTCAACTTTGAAGACGGCACG-3  
Reverse primer, YM-GALOXENG-PR25: 5'-CCCGCTAGCCCAGTCAATAGCATCTACATAGATAG-3  
1 μL of plasmid template DNA, 1 μL each of primer (100 μM), 0.5 μL dNTP mix (25 mM each), 2.5 μL of 10X reaction buffer, 18.5 μL of dH2O, 0.5 μL Pfu DNA polymerase.  
95°C for 5 mins; Repeat 5 times: 95°C for 60 s, 55°C for 40 s, 72°C for 30 s; Repeat 25 times: 95°C for 60 s, 60°C for 40 s, 72°C for 30 s; 72°C for 5 mins. |
| Mutagenic PCR to create W290F GalOx | Forward primer, YM-GALOXENG-PR20: 5'-GGTAGAGTTTTTACTATCGGTGGATCCTTCTCAGGTGGAGTTTTC-3  
Reverse primer, YM-GALOXENG-PR21: 5'-GAAAACTCCACCTGAGAAGGATCCACCGATAGTAAAAACTCTACC-3  
0.2 μL of plasmid template DNA (25 ng), 1 μL each of primer (125 ng each), 1 μL dNTP mix (25 mM each), 1 μL Agilent “Quick” solution, 2.5 μL of 10X reaction buffer, 17.3 μL of dH2O, 0.5 μL Pfu DNA polymerase.  
95°C for 2 mins; Repeat 18 times: 95°C for 60 s, 60°C for 50 s, 68°C for 5 mins; 68°C for 10 mins. |
| Mutagenic PCR to create R330K GalOx | Forward primer, YM-GALOXENG-PR13: 5'-GACTGCTGATAAGCAAGGTTTGTATAAATCTGACAATCACG-3  
Reverse primer, YM-GALOXENG-PR14: 5'-CTGAGATGAGTTTTTGTTCTAGA-3  
Same as described from W290F mutagenesis. |
| Mutagenic PCR to create Q406T GalOx | Forward primer, YM-GALOXENG-PR15: 5'-GGTGGATCTCCTGACTATTACGATTCCGACGCTACTACA-3  
Reverse primer, YM-GALOXENG-PR16: 5'-TGTAGTAGCGTCGGAATCGTAATAGTCAGGAGATCCACC-3  
Same as described from W290F mutagenesis. |
| PCR amplification of CBM11 (GH7) (with N-terminal EcoRI site and C-terminal XbaI site) | Forward primer, YM-GALOXENG-PR28: 5'-CCCGAATTCGCCCTCTGCTCCAATTGGTTC-3  
Reverse primer, YM-GALOXENG-PR29: 5'-CCCGCTAGCTTGAAAAACGTTGATCTCAGC-3  
As described from PCR of GalOX linker + CD. |
| PCR amplification of CBM1(GH61) (with N-terminal EcoRI site and C-terminal XbaI site) | Forward primer, YM-GALOXENG-PR30: 5'-CCCGAATTCGCCCTCTGCTCCAATTGGTTC-3  
Reverse primer, YM-GALOXENG-PR31: 5'-CCCGCTAGCTTGAAAAACGTTGATCTCAGC-3  
As described from PCR of GalOX linker + CD. |
| Mutagenic PCR to generate pool of mutant G0OX genes | Forward primer, YM-RANMUT-PR3: 5'-GAGAGGCTGAAGCTGAATTC-3  
Reverse primer, YM-RANMUT-PR4: 5'-CTGAGATGAGTTTTTGTTCTAGA-3  
As described for PCR of GalOX linker + CD.  
0.25 μL of plasmid template DNA (8 ng), 1 μL each of primer (125 ng each), 1 μL dNTP mix (10 mM each), 5 μL of 10X Agilent Mutazyme II reaction buffer, 40.75 μL of dH2O, 0.5 μL Pfu DNA polymerase.  
95°C for 2 mins; Repeat 29 times: 95°C for 30 s, 55°C for 30 s, 72°C for 90 s; 72°C for 10 mins. |
APPENDIX 6: RECIPES

1 M phosphate buffer, pH 6.0, for Pichia media
1. Combine the following:
   - 132 mL 1 M K₂HPO₄
   - 868 mL 1 M KH₂PO₄
2. Adjust the pH to 6.0 if necessary using H₃PO₄ or KOH
3. Autoclave at 121°C for 20 minutes
4. Store at 4°C

10X D, 20% glucose for Pichia media
1. Dissolve the following in 1000 mL deionised water:
   - 200 g D-glucose
2. Autoclave at 121°C for 20 minutes

10X M, 5% methanol for Pichia media
1. Add the following to 475 mL deionised water:
   - 25 mL 100% methanol
2. Filter-sterilise and store at 4°C

10X GY, 10% glycerol for Pichia media
1. Add the following to 900 mL deionised water:
   - 100 mL glycerol
2. Autoclave at 121°C for 20 minutes
3. Store at 4°C

10X YNB, 13.4% yeast nitrogen base for Pichia media
1. Dissolve the following in 1000 mL deionised water:
   - 134 g yeast nitrogen base with ammonium sulphate and without amino acids
2. Filter-sterilise and store at 4°C

500X B, 0.02% biotin for Pichia media
1. Add the following to 50 mL deionised water:
   - 10 mg biotin
2. Filter-sterilise and store at 4°C

Binding buffer, for purification of Ni²⁺ proteins
- 20 mM sodium phosphate buffer, pH 7.4
- 0.5 M NaCl
- 20 mM imidazole

Blocking buffer, for Western blot
- 5% skim milk powder in TBST buffer

BMM, buffered minimal media plates for plate-based GalOx assay of Pichia transformants
1. Dissolve the following in 61.8 mL deionised water:
   - 2 g agar
2. Autoclave at 121°C for 20 minutes
3. Cool to around 50°C and add the following:
   - 0.2 mL 500X B
   - 1 mL of 1 mg/mL horseradish peroxidase (at least 1500 U/mg)
   - 2 mL 50 mg/mL ABTS
   - 5 mL of 400 g/L D-galactose
   - 10 mL 1 M potassium phosphate buffer, pH 6.0
   - 10 mL 10X YNB
   - 10 mL 10X M

4. Stir and pour quickly before media solidifies
5. Store at 4°C

**BMGY, buffered minimal media for Pichia growth**

1. Dissolve the following in 700 mL deionised water:
   - 10 g yeast extract
   - 20 g peptone
   - 20 g agar (only for plates, exclude for liquid culture)
2. Autoclave at 121°C for 20 minutes
3. Cool to around 50°C and add the following:
   - 10 mL 1 M potassium phosphate buffer, pH 6.0
   - 10 mL 10X YNB
   - 10 mL 10X GY
   - 2 mL 500X B
4. Stir and pour quickly before media solidifies
5. Store at 4°C

**BMMY, buffered minimal media to induce Pichia expression**

1. Dissolve the following in 700 mL deionised water:
   - 10 g yeast extract
   - 20 g peptone
   - 20 g agar (only for plates, exclude for liquid culture)
2. Autoclave at 121°C for 20 minutes
3. Cool to around 50°C and add the following:
   - 10 mL 1 M potassium phosphate buffer, pH 6.0
   - 10 mL 10X YNB
   - 10 mL 10X M
   - 2 mL 500X B
4. Stir and pour quickly before media solidifies
5. Store at 4°C

**Elution buffer, for purification of Ni^{2+} proteins**

- 20 mM sodium phosphate buffer, pH 7.4
- 0.5 M NaCL
- 500 mM imidazole

**LSLB.-Zeocin low-salt Luria-Bertani media for E. coli growth**

1. Dissolve the following in 1000 mL deionised water:
   - 10 g tryptone
   - 5 g yeast extract
   - 5 g NaCL
2. Adjust the pH to 7.5 with 1N NaOH
3. Autoclave at 121°C for 20 minutes
4. Cool to around 50°C and add the following:
- 0.25 mL 100 mg/mL Zeocin
4. Stir and pour quickly before media solidifies
5. Store at 4°C

Stable amorphous cellulose, for use in X-ray diffraction studies
1. Bubble SO₂ gas through DMSO for 20 minutes. Titrate a 100X dilution of SO₂-DMSO with 0.5 NaOH using phenolphthalein indicator. Calculate the concentration of SO₂ in DMSO considering a 1:2 molar ratio of SO₂ to NaOH. Stable for at least one month.
2. Add 1 g of microcrystalline cellulose to 20 mL DMSO. Allow swelling to occur. This process will occur eventually at room temperature but can be accelerated by 10 minutes at ~60°C. Add the correct volume of SO₂-DMSO equivalent to 1.19 g SO₂. Then add 1.35 g diethylamine (1.9 mL) and shake vigourously.
3. Once completely solubilised, the above solution is slowly but continuously poured into an excess volume of deionised water (~250 mL dH₂O). The regenerated cellulose is immediately obvious as a gel in the form of strands. Wash by decanting and replacing the deionised water (four washes with 100 mL). Retain the solution as chemical waste.
4. After the final decanting, the cellulose gel pieces should be transferred to a mortar and placed in the freeze drier until dry. Gentle grinding with a pestle will yield stable amorphous cellulose powder.

TBST buffer, for Western blot
- 50 mM Tris-HCL, pH 7.5
- 137 mM NaCl
- 0.1% Tween-20

YPDS-Zeocin, rich media for recovering *Pichia* transformants
1. Dissolve the following in 900 mL deionised water:
   - 10 g yeast extract
   - 182.2 g sorbitol
   - 20 g peptone
   - 20 g agar
2. Autoclave at 121°C for 20 minutes
3. Cool to around 50°C and add the following:
   - 100 mL 10X D
   - 1 mL 100 mg/mL Zeocin
4. Stir and pour quickly before media solidifies
5. Store at 4°C
Appendix 7: Sequences

GalOx Variant #1:
- Start and stop codons are shown in purple
- Key restriction sites are shown in red
- The GalOx linker sequence is shown in orange
- The GalOx CD sequence is shown in green
- The uncoloured sequence is from the pPICZα plasmid

GalOx Variant #2:
- Start and stop codons are shown in purple
- Key restriction sites are shown in red
- The GalOx CBM sequence is shown in blue
- The GalOx linker sequence is shown in orange
- The GalOx CD sequence is shown in green
- The uncoloured sequence is from the pPICZα plasmid
GalOx Variant #3:
- Start and stop codons are shown in purple
- Key restriction sites are shown in red
- The CBM29 sequence is shown in blue
- The GalOx linker sequence is shown in orange
- The GalOx CD sequence is shown in green
- The uncoloured sequence is from the pPICZα plasmid

GalOx Variant #4:
- Start and stop codons are shown in purple
- Key restriction sites are shown in red
- The GalOx linker sequence is shown in orange
- The GalOx CD sequence is shown in green
- The three codons mutated for the RQW variant are underlined
- The uncoloured sequence is from the pPICZα plasmid
TCATTGTAGTACATTTCTCGAGCATCTCGTTAATGCAATGTACGTATGCAAGTTGCTAGAGGATATCAGTCATCTGCTACAATGTCTGACGGTAGAGTTTTTACTATCGGTGGATCCTTC

TCAGGTGGAGTTTTCGAAAAGAACGGAGAGGTTTACTCCCCATCCTCAAAGACATGGACTTCATTGCCAAATGC

TAAAGTTAACCATATGTAGCTGCAGTAAGACGGATTTGTTGTTGTTGAAAGAAAGATCAGTTTCCAGGCTGACCTTCTACAGCTATGAACTGGTACTATACTTCTGGTTCCGGAGATGTTAAGTCCGCTGGTAAAAGACAATCAAAATAGAGGAGTTGCTCCAGATGCTATGTGTGGTAACGCTGTTATGTACGACGCTGTTAAGGGAAAAATTGGTTGTTTGCTGCTGCCATCACTTGACGTAGTTTGCAGAACCATCTCCAAACTCTCAAGTG

GCTGGACCTTCTACAGCTATGAACTGGTACTATACTTCTGGTTCCGGAGATGTTAAGTCCGCTGGTAAAAGACAATCAAATAGAGGAGTTGCTCCAGATGCTATGTGTGGTAACGCTGTTATGTACGACGCTGTTAAGGGAAAAATTGGTTGTTTGCTGCTGCCATCACTTGACGTAGTTTGCAGAACCATCTCCAAACTCTCAAGTG

GalOx Variant #5:
- Start and stop codons are shown in purple
- Key restriction sites are shown in red
- The GalOx CBM sequence is shown in blue
- The GalOx linker sequence is shown in orange
- The GalOx CD sequence is shown in green
- The three codons mutated for the RQW variant are underlined
- The uncoloured sequence is from the pPICZα plasmid

GalOx Variant #6:
- Start and stop codons are shown in purple
- Key restriction sites are shown in red
- The CBM11 sequence is shown in blue
- The GalOx linker sequence is shown in orange
- The GalOx CD sequence is shown in green
- The three codons mutated for the RQW variant are underlined
- The uncoloured sequence is from the pPICZα plasmid
GalOx Variant #7:

- Start and stop codons are shown in purple
- The CBM1 (GH7) sequence is shown in red
- The GalOx linker sequence is shown in orange
- The GalOx CD sequence is shown in green
- The three codons mutated for the RQW variant are underlined
- The uncoloured sequence is from the pPICZα plasmid

GalOx Variant #8:

- Start and stop codons are shown in purple
- Key restriction sites are shown in red
- The CBM1 (GH61) sequence is shown in blue
- The GalOx linker sequence is shown in orange
- The GalOx CD sequence is shown in green
- The three codons mutated for the RQW variant are underlined
- The uncoloured sequence is from the pPICZα plasmid

ATGAGATTTTCTCCAATTTTTTACTGCTGTTTTTATTCGAGCATCCTCCGCCATAGCTGCTCGCAGTCACACTACAACAGAAGATGAAACGGCACAAAATTCGAGCAGCATCCTGCTGCAAGCTCAGTACTAAAAAGAGGGTTTGTAATTGTGATTAACATTATCTGCCTGCTAAATCTATTCGTTCTTCTGGTCTGTTTTGCCATTTTCCAACAGCACAAATAACGGGTTATTGTTTGGATGTGTTTGAAGTTGAGGAGGTGCTGACCTATATAGAAATGACGCTTTTTAGTGGAGATTCTCCTGGGATCCTTCAACTGGTATCGTTTCTGACAGAACTGTTACAGTTACTAAGCATGATATGTTCTGTCCAGGAATTTCTATGGACGGTAATGGACAAATCGTTGTTACAGGTGGAAACGATGCTAAGAAAACTTCATTGTACGATTCATCTTCCGACTCTTGGATTCCAGGTCCTGATATGCAAGTTGCTAGAGGATATCAGTCATCTGCTACAATGTCTGACGGTAGAGTTTTTACTATCGGTGGATCCTTC

GOOX gene contained in pPICZα plasmid:
- Start and stop codons are shown in purple
- Key restriction sites are shown in red
- The GOOX sequence is shown in green
- The uncoloured sequence is from the pPICZα plasmid

ATGAGATTTTCTCCAATTTTTTACTGCTGTTTTTATTCGAGCATCCTCCGCCATAGCTGCTCGCAGTCACACTACAACAGAAGATGAAACGGCACAAAATTCGAGCAGCATCCTGCTGCAAGCTCAGTACTAAAAAGAGGGTTTGTAATTGTGATTAACATTATCTGCCTGCTAAATCTATTCGTTCTTCTGGTCTGTTTTGCCATTTTCCAACAGCACAAATAACGGGTTATTGTTTGGATGTGTTTGAAGTTGAGGAGGTGCTGACCTATATAGAAATGACGCTTTTTAGTGGAGATTCTCCTGGGATCCTTCAACTGGTATCGTTTCTGACAGAACTGTTACAGTTACTAAGCATGATATGTTCTGTCCAGGAATTTCTATGGACGGTAATGGACAAATCGTTGTTACAGGTGGAAACGATGCTAAGAAAACTTCATTGTACGATTCATCTTCCGACTCTTGGATTCCAGGTCCTGATATGCAAGTTGCTAGAGGATATCAGTCATCTGCTACAATGTCTGACGGTAGAGTTTTTACTATCGGTGGATCCTTC

CATCATTGCTTCTCAATTTTTTACTGCTGTTTTTATTCGAGCATCCTCCGCCATAGCTGCTCGCAGTCACACTACAACAGAAGATGAAACGGCACAAAATTCGAGCAGCATCCTGCTGCAAGCTCAGTACTAAAAAGAGGGTTTGTAATTGTGATTAACATTATCTGCCTGCTAAATCTATTCGTTCTTCTGGTCTGTTTTGCCATTTTCCAACAGCACAAATAACGGGTTATTGTTTGGATGTGTTTGAAGTTGAGGAGGTGCTGACCTATATAGAAATGACGCTTTTTAGTGGAGATTCTCCTGGGATCCTTCAACTGGTATCGTTTCTGACAGAACTGTTACAGTTACTAAGCATGATATGTTCTGTCCAGGAATTTCTATGGACGGTAATGGACAAATCGTTGTTACAGGTGGAAACGATGCTAAGAAAACTTCATTGTACGATTCATCTTCCGACTCTTGGATTCCAGGTCCTGATATGCAAGTTGCTAGAGGATATCAGTCATCTGCTACAATGTCTGACGGTAGAGTTTTTACTATCGGTGGATCCTTC

CATCATTGCTTCTCAATTTTTTACTGCTGTTTTTATTCGAGCATCCTCCGCCATAGCTGCTCGCAGTCACACTACAACAGAAGATGAAACGGCACAAAATTCGAGCAGCATCCTGCTGCAAGCTCAGTACTAAAAAGAGGGTTTGTAATTGTGATTAACATTATCTGCCTGCTAAATCTATTCGTTCTTCTGGTCTGTTTTGCCATTTTCCAACAGCACAAATAACGGGTTATTGTTTGGATGTGTTTGAAGTTGAGGAGGTGCTGACCTATATAGAAATGACGCTTTTTAGTGGAGATTCTCCTGGGATCCTTCAACTGGTATCGTTTCTGACAGAACTGTTACAGTTACTAAGCATGATATGTTCTGTCCAGGAATTTCTATGGACGGTAATGGACAAATCGTTGTTACAGGTGGAAACGATGCTAAGAAAACTTCATTGTACGATTCATCTTCCGACTCTTGGATTCCAGGTCCTGATATGCAAGTTGCTAGAGGATATCAGTCATCTGCTACAATGTCTGACGGTAGAGTTTTTACTATCGGTGGATCCTTC

CATCATTGCTTCTCAATTTTTTACTGCTGTTTTTATTCGAGCATCCTCCGCCATAGCTGCTCGCAGTCACACTACAACAGAAGATGAAACGGCACAAAATTCGAGCAGCATCCTGCTGCAAGCTCAGTACTAAAAAGAGGGTTTGTAATTGTGATTAACATTATCTGCCTGCTAAATCTATTCGTTCTTCTGGTCTGTTTTGCCATTTTCCAACAGCACAAATAACGGGTTATTGTTTGGATGTGTTTGAAGTTGAGGAGGTGCTGACCTATATAGAAATGACGCTTTTTAGTGGAGATTCTCCTGGGATCCTTCAACTGGTATCGTTTCTGACAGAACTGTTACAGTTACTAAGCATGATATGTTCTGTCCAGGAATTTCTATGGACGGTAATGGACAAATCGTTGTTACAGGTGGAAACGATGCTAAGAAAACTTCATTGTACGATTCATCTTCCGACTCTTGGATTCCAGGTCCTGATATGCAAGTTGCTAGAGGATATCAGTCATCTGCTACAATGTCTGACGGTAGAGTTTTTACTATCGGTGGATCCTTC

CATCATTGCTTCTCAATTTTTTACTGCTGTTTTTATTCGAGCATCCTCCGCCATAGCTGCTCGCAGTCACACTACAACAGAAGATGAAACGGCACAAAATTCGAGCAGCATCCTGCTGCAAGCTCAGTACTAAAAAGAGGGTTTGTAATTGTGATTAACATTATCTGCCTGCTAAATCTATTCGTTCTTCTGGTCTGTTTTGCCATTTTCCAACAGCACAAATAACGGGTTATTGTTTGGATGTGTTTGAAGTTGAGGAGGTGCTGACCTATATAGAAATGACGCTTTTTAGTGGAGATTCTCCTGGGATCCTTCAACTGGTATCGTTTCTGACAGAACTGTTACAGTTACTAAGCATGATATGTTCTGTCCAGGAATTTCTATGGACGGTAATGGACAAATCGTTGTTACAGGTGGAAACGATGCTAAGAAAACTTCATTGTACGATTCATCTTCCGACTCTTGGATTCCAGGTCCTGATATGCAAGTTGCTAGAGGATATCAGTCATCTGCTACAATGTCTGACGGTAGAGTTTTTACTATCGGTGGATCCTTC