MULTIVARIATE CHARACTERIZATION OF LIGNOCELLULOSIC BIOMASS AND GRAFT MODIFICATION OF NATURAL POLYMERS

by

Daniel J. Krasznai (B.Sc. Eng.)

A thesis submitted to the Department of Chemical Engineering

In conformity with the requirements for

the degree of Master of Applied Science

Queen's University

Kingston, Ontario, Canada

February 2012

Abstract

The cell walls of plants contain significant quantities of renewable polymers in the form of cellulose, hemicellulose, and lignin. These three renewable polymers have the potential to complement or replace synthetic polymers in a variety of applications. Rapidly determining the relative quantities of these polysaccharides in lignocellulosic biomass is important yet difficult since plant biomass is recalcitrant and highly variable in composition.

Part of this contribution outlines a novel compositional analysis protocol using infrared spectroscopy and multivariate regression techniques that is rapid and inexpensive. Multivariate regression models based on calibration mixtures can be used to discern between populations of lignocellulosic biomass or to predict cellulose, hemicellulose, and lignin quantities. Thus, the compositional analysis step can be expedited so that other processes, like fractionation of the lignocellulose polymers, can be tuned accordingly to maximize the value of the final product.

Hybrid materials were also generated using a variety of polymerization techniques and post-polymerization modifications. A novel controlled/living radical polymerization initiator was synthesized (2-bromo-2-methylpropane hydrazide) containing a hydrazide functionality that was covalently linked to the reducing-end of dextran. Despite the rapid coupling of the hydrazide-based initiator to the reducing-end of dextran, the instability of the alkyl bromide bond resulted in several unsuccessful attempts at Cu(0)-mediated controlled/living radical polymerization. A number of recommendations were given to improve the stability of this compound; however, an alternative approach to synthesizing hybrid copolymers was also investigated in parallel.

Hyperbranched polymers were synthesized using commercially available vinyl and divinyl monomers in the presence of a cobalt(II) complex that enabled control over the size, architecture, and mol% of pendant vinyl groups amenable to post-polymerization modification.

Modifying the ratio of divinyl monomer to cobalt(II) complex provided a series of hyperbranched polymers with variable morphology and mol% pendant vinyl groups. The pendant vinyl bonds were subsequently converted to amines *via* thiol additions with cysteamine. These amine functionalized hyperbranched polymers were then used in a subsequent reductive amination reaction with the reducing-end of dextran to produce the amphiphilic core-shell copolymer poly(methyl methacrylate-*co*-ethylene glycol dimethacrylate)-*b*-dextran. These amphiphilic copolymers mimicked the colloidal behaviour of conventional block copolymer micelles without requiring difficult syntheses or tedious self-assembly steps.

Co-Authorship

Chapter 1 was written by Daniel Krasznai and reviewed by Pascale Champagne and Michael Cunningham.

Chapter 2 was written by Daniel Krasznai and coauthored by Pascale Champagne and Michael Cunningham. Chapter 2 was submitted for publication in *Biomass and Bioenergy* on December 27, 2011 (see reference 1 below).

Chapter 3 was written by Daniel Krasznai and coauthored by Pascale Champagne and Michael Cunningham. Chapter 3 was accepted for publication (with revisions) in *Bioresource Technology* on December 4, 2011 and resubmitted January 13, 2012 (see reference 2 below).

Chapter 4 was written by Daniel Krasznai and reviewed by Pascale Champagne and Michael Cunningham. Chapter 4 was not submitted for publication.

The first draft of Chapter 5 was written by Daniel Krasznai and reviewed by Niels Smeets, Pascale Champagne, Michael Cunningham, and Timothy McKenna. Niels Smeets made revisions prior to submitting significant portions of this chapter for publication in *Polymer Chemistry* on December 19, 2011 (see reference 3 below).

The corresponding references for the chapters submitted for publication are as follows:

(1) **Krasznai, D.J.**, Champagne, P., and Cunningham, M.F. *A Historical Review: The Compositional Characterization of Lignocellulosic Biomass*. Biomass and Bioenergy (Submitted December 27, 2011; Ref.: JBB-D-11-01243).

(2) **Krasznai, D.J.**, Champagne, P., and Cunningham, M.F. *Quantitative Characterization of Lignocellulosic Biomass Using Surrogate Mixtures and Multivariate Techniques*. Bioresource Technology (Submitted September 18, 2011 ; Accepted January 17, 2011; DOI: 10.1016/j.biortech.2012.01.089).

(3) **Krasznai, D.J.**, McKenna, T.F.L., Cunningham, M.F., Champagne, P., and Smeets, N.M.B. *Polysaccharide-stabilized Core Cross-linked Polymer Micelle Analogues*. Polymer Chemistry (Submitted December 19, 2011; Accepted January 28, 2011; DOI: 10.1039/C2PY00601D).

Acknowledgements

I would like to thank my thesis project advisors, Dr. Pascale Champagne and Dr. Michael Cunningham, for their patience, support, and guidance. I would also like to express my gratitude towards Dr. Ralph Whitney and Dr. Ronald Neufeld for their advice regarding organic and carbohydrate chemistry.

For their expertise, advice, and collaboration, I thank Mr. Nicky Chan, Dr. Niels Smeets, Dr. Jeffery Wood, and Dr. Julien Pinaud. To all of my lab and office mates, thank you for your help and collegiality. Thanks to the Department of Chemical Engineering at Queen's University for providing me with the courses, equipment, personnel, and expertise necessary to make this thesis possible.

And last, but not least, I would like to express my deepest gratitude towards my parents, because they truly fostered my curiosity and interest in science.

"Nature uses only the longest threads to weave her patterns,
so that each small piece of her fabric reveals the organization of the entire tapestry."

– Richard Feynman, *The Character of Physical Law*

Table of Contents

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of Tables

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of Abbreviations

AACC	American Association for Clinical Chemistry
ADF	Acid Detergent Fibre
ADL	Acid Detergent Lignin
aNDF	α-Amylase Neutral Detergent Fibre
AOAC	Association of Official Agricultural Chemists
ARGET	Activators Regenerated by Electron Transfer
ASTM	American Society for Testing and Materials
ATR	Attenuated Total Reflectance
ATRP	Atom Transfer Radical Polymerization
BA	2-Bromoacetamide
BOC	*tert*-Butyloxycarbonyl
BOC-hydrazine	*tert*-Butyl carbazate
BriB	α-Bromoisobutyryl bromide
CA	Cysteamine hydrochloride
CCTP	Catalytic Chain Transfer Polymerization
CF	Crude Fibre
COBF	Bis(boron difluorodimethylglyoximate)
CP/MAS	Cross Polarization Magic Angle Spinning
CRP	Controlled Radical Polymerization
CuAAc	Copper(I)-catalyzed azide-alkyne cycloaddition
CV	Cross Validation
DDW	Distilled Deionized Water
DEX	Dextran
DF	Dietary Fibre
DFRC	Derivatization Followed by Reductive Cleavage
DLS	Dynamic Light Scattering
DMAc	Dimethyl acetamide
DMAP	4-(Dimethylamino) pyridine
DMF	Dimethyl formamide
DMSO	Dimethyl sulfoxide
DP	Degree of Polymerization
DSC	Differential Scanning Calorimetry
EBiB	Ethyl α-bromoisobutyrate
EGDMA	Ethylene glycol dimethacrylate
EMSC	Extended Multiplicative Scatter/Signal Correction
EtOAc	Ethyl acetate
EtOH	Ethanol
FPL	Forest Products Laboratory
FTIR	Fourier Transform Infrared

GC	Gas Chromatography
GPC	Gel Permeation Chromatography
HPLC	High Performance Liquid Chromatography
HSQC	Heteronuclear Single-Quantum Correlation
iPrOH	Isopropanol
IV	Intrinsic Viscosity
KL	Klason Lignin
LALS/RALS	Left/Right Angle Light Scattering
LAP	Laboratory Analytical Procedure
LRP	Living Radical Polymerization
LSEC	Linear Size Exclusion Chromatography
MA	Methyl acrylate
MBiP	Methyl 2-bromopropionate
Me_6TREN	Tris[2-(dimethylamino)ethyl]amine
MEA	Monoethanolamine
MeOH	Methanol
MLR	Multiple Linear Regression
MMA	Methyl methacrylate
MSC	Multiplicative Scatter/Signal Correction
MWCO	Molecular Weight Cutoff
MWD	Molecular Weight Distribution
NDF	Neutral Detergent Fibre
NIPALS	Nonlinear Iterative Partial Least Squares
NIR	Near Infrared
NMR	Nuclear Magnetic Resonance
NREL	National Renewable Energy Laboratory
OSC	Orthogonal Scatter/Signal Correction
OSET	Outer Shell Electron Transfer
PAT	Process Analytical Technology
PCA	Principle Component Analysis
PCR	Principle Component Regression
PLS	Projection of Latent Structures or Partial Least Squares
pMA/PMA	Poly(methyl acrylate)
pMMA/PMMA	Poly(methyl methacrylate)
PSD	Particle Size Distribution
PTFE	Poly(tetrafluoroethylene)
pyMBMS	Pyrolysis Molecular Beam Mass Spectrometry
RMSEC/RMSECV	Root Mean Square Error of Calibration/Cross Validation
RMSEE	Root Mean Square Error of Estimation
RMSEP	Root Mean Square Error of Prediction
SCVP	Self-condensing Vinyl Polymerization

SEC	Size Exclusion Chromatography
SET	Single Electron Transfer
SNV	Standard Normal Variate
SSF	Simultaneous Saccharification and Fermentation
TAPPI	Technical Association of the Pulp and Paper Industry
TBAF	Tetrabutylammonium fluoride
TDSEC	Triple Detector Size Exclusion Chromatography
TEA	Triethylamine
TEAm-Br	Triethylammonium bromide
TEM	Transmission Electron Microscopy
TFA	Trifluoroacetic acid
TGA	Thermogravimetric Analysis
THF	Tetrahydrofuran
TMS	Trimethylsilyl
TPMA	Tris(2-pyridylmethyl)amine
UK	United Kingdom
USDA	United States Department of Agriculture
UV/VIS	Ultraviolet/Visible
vazo67	2,2'-Azobis(2-methylbutyronitrile)

List of Symbols and Nomenclature

$[\eta]$	Intrinsic viscosity [$mL \cdot g^{-1}$]
$[I]$	Initiator concentration [$mol \cdot L^{-1}$]
$[M]$	Monomer concentration [$mol \cdot L^{-1}$]
$[M]_0$	Initial monomer concentration [$mol \cdot L^{-1}$]
a	Mark-Houwink-Kuhn-Sakurada parameter [-]
C_T	Intrinsic activity [-]
D	Dispersity [-]
$D_{p(N)}$	Number-average particle size [nm]
$D_{p(V)}$	Volume-average particle size [nm]
E	Error matrix [-]
F	Error matrix [-]
f_{tr}	Chain transfer frequency [s^{-1}]
g'	Zimm branching factor [-]
I_{eff}	Initiator efficiency [%]
K	Mark-Houwink-Kuhn-Sakurada parameter [$dL \cdot g^{-1}$]
k	Number of variables in **Y** [-]
k_{act}	Activation rate coefficient [$mol \cdot g^{-1} \cdot s^{-1}$]
k_{tr}	Chain transfer rate coefficient [$L \cdot mol^{-1} \cdot s^{-1}$]
M_n	Number-average molecular weight [$g \cdot mol^{-1}$]
M_v	Viscosity-average molecular weight [$g \cdot mol^{-1}$]
M_w	Weight-average molecular weight [$g \cdot mol^{-1}$]
n	Number of samples [-]
n_v	Number of samples left out for cross-validation [-]
P	Loadings matrix [-]
Q	Loadings matrix [-]
R_H	Hydrodynamic radius [nm]
T	Scores matrix [-]
T_g	Glass transition temperature [^oC]
U	Scores matrix [-]
X	Matrix consisting of FT-IR absorbance values [-]
$\mathbf{x}_i$	Vector of spectral absorbance values for a single sample [-]
$x_{i,k}$	Absorbance value of *i*th sample at *k*th wavenumber [arbitrary units]
Y	Matrix consisting of reference weight fractions [-]
Z-avg.	Z-average particle size [nm]
α	Alpha anomer [-]
β	Beta anomer [-]
δ	Chemical shift [ppm]
ζ	Zeta potential [mV]
φ	Ratio of f_{tr} to [EGDMA] [$s \cdot mol\%^{-1}$]

Chapter 1

Introduction

1.1 Background and Motivation

Lignocellulose is a complex polymer network containing polysaccharides (cellulose and hemicellulose) and phenolic polymer (lignin) that is found within the cell walls of plants. Cellulose, a β(1→4) linked polymer of D-glucopyranose, is the most abundant carbohydrate on Earth (and arguably the most abundant polymer on Earth) and is produced at a rate of about 100 billion tons per year.[1] In nature, the chemical composition of lignocellulose is inherently variable due to differences in plant species, cultivating environment, and growing seasons. Furthermore, the recalcitrant polymer network that is formed by cellulose, hemicellulose, and lignin presents a hindrance to the chemical processing and analysis of lignocellulose.[2] Numerous chemical techniques have been developed over the past 150 years to determine the composition of lignocellulosic biomass (see Chapter 2). These chemical techniques focus on selectively fractionating (i.e., extracting) the three main constituents of lignocellulose and obtaining the relative composition by a mass balance. In the mid-1970's, the statistician Herman Wold developed the idea of "soft modeling" where multivariate statistics are used to recognize and predict patterns in large multivariate datasets.[3] As a result of many years of statistical and computational development, multivariate regression techniques, such as Partial Least Squares Regression (PLSR), have been used to predict the composition of lignocellulosic biomass using Fourier Transform Infrared (FT-IR) and Near Infrared (NIR) spectroscopic data.[4,5] In addition, FT-IR-based multivariate regression models have been used to assess the sugar composition in hydrolyzed biomass liquors.[6] Compared to standard chromatographic methods (e.g., HPLC) FT-IR is non-destructive and allows for high throughput screening of biomass and its hydrolysate as

long as it is used in conjunction with a well-calibrated multivariate regression model.[5] As a result the chemical composition of lignocellulosic biomass can be easily predicted using spectroscopic techniques, thus reducing the need for more tedious and expensive wet-chemical analyses (see Chapter 3).[5] Aside from the compositional analysis of lignocellulose, infrared-based multivariate models have also been used in synthetic polymer applications to rapidly determine conversions[7], compositional information[8] or morphological characteristics[9]. Because of its versatility, expedience, and low cost, the application of infrared spectroscopy and multivariate techniques to the characterization of polymers and lignocellulosic biomass provides an expedient and cost-effective Process Analytical Technology (PAT).[10]

Differences in the chemical functionality, architecture, and molecular weight of the polysaccharide ultimately dictate the types of synthetic polymer chemistries that can be employed. Thus, the aforementioned compositional analysis techniques constitute the first step in converting lignocellulosic biomass into value-added polymeric materials. For example, cellulose is a renewable, inexpensive, and versatile polymer having an electrophilic reducing-end moiety and abundant hydroxyl functionalities on its backbone.[11] Economic and environmental drivers, coupled with low margin between cellulosic ethanol and feedstock prices[12], have led to numerous well-defined cellulose derivatives through modification of the backbone hydroxyls or reducing-end hemiacetal functionality. Cellulose, like many polysaccharides, is hydrophilic but is insoluble in most conventional solvents due to its crystalline, hydrogen-bonded structure.[11] Thus, water- or DMSO-soluble model compounds, such as cellobiose or dextran, are often used instead of insoluble polysaccharides in order to study novel systems.[13] Likewise, protecting the backbone groups of polysaccharides often facilitates the polymer's dissolution in organic solvents, making the material amenable to the homogenous functionalization and polymerization in conventional media.[14] Alternatively, ionic liquids can be used to dissolve recalcitrant polysaccharides, like

cellulose, which facilitates the homogenous modification or polymerization of these materials without prior deriviatization.[15,16] Recent advances in controlled/living radical polymerization techniques have produced several examples of amphiphilic macromolecules composed of carbohydrates (e.g., polysaccharides) and synthetic polymers.[17,18] Under certain conditions, enthalpic and entropic effects tend to cause the aggregation of these amphiphilic molecules giving colloidal structures with numerous potential applications in the pharmaceutical and chemical industries.[19] A multitude of polymer chemistries can be used to produce well-defined graft[20] and block[21] copolymers containing polysaccharide (e.g., cellulose or dextran) and synthetic units, including Atom Transfer Radical Polymerization (ATRP)[22] and Single Electron Transfer Living Radical Polymerization (SET-LRP)[23]. These approaches generally rely on coupling synthetic polymers with the polysaccharide (i.e., grafting-to)[24]; growing a synthetic polymer from an initiating site introduced on the polysaccharide backbone or reducing-end (i.e., grafting-from)[24,21]; or using carbohydrate-based monomers in conjunction with conventional synthetic monomers[25]. In this manner, controlled/living radical polymerization techniques can be used to produce potentially high-value amphiphilic copolymers with tuneable properties by manipulating the reaction conditions or polymerization recipes. Reactions and polymerizations with polysaccharides are generally conducted in either homogenous solvents (e.g., water, ionic liquids, dimethyl sulfoxide, or dimethyl acetamide with LiBr) or in conventional organic solvents by protecting the polysaccharide backbone groups prior to modification. However, the poor solubility of polysaccharides in conventional solvents is still one of the primary hindrances to the synthesis of these amphiphilic copolymers. Furthermore, the amination of the polysaccharide at the reducing-end is generally slow and inefficient, often taking days to reach quantitative conversions.[26] Alternative polymer chemistries that utilize commercially available initiators and low quantities of inexpensive catalyst systems, such as Catalytic Chain Transfer Polymerization

(CCTP)[27], could also provide an inexpensive route towards amphiphilic polysaccharide-containing copolymers. Techniques that facilitate the homogenous, inexpensive, expedient, and efficient polymerization or modification of polysaccharides could lead to a wide array of high-value amphiphilic, biocompatible, and biodegradable copolymers.

1.2 Scope and Organization of Thesis

The first objective of this thesis was to expedite the compositional analysis of lignocellulosic biomass using infrared spectroscopy and multivariate analysis techniques. A comprehensive literature review was written on traditional and contemporary methods for the compositional analysis of lignocellulosic biomass to provide the basis for subsequent work involving novel multivariate analysis techniques. Subsequently, multivariate analysis of infrared spectra derived from powdered samples of lignocellulosic biomass was used to determine polysaccharide (cellulose and hemicellulose) and lignin quantities in an attempt to mitigate the tedious wet-chemical compositional analysis techniques that have been traditionally invoked. Furthermore, amphiphilic hybrid materials containing covalently linked polysaccharide (e.g., cellulose) and synthetic blocks were considered to be of interest as value-added advanced renewable materials. Consequently, the synthesis of amphiphilic polysaccharide-containing block copolymers was the second major objective of this thesis. A variety of polymerization and post polymerization chemistries were explored with the goal of generating a number of materials containing both polysaccharide and well-defined synthetic parts.

Chapter 2 presents a comprehensive literature review of traditional and novel methods for the compositional analysis of lignocellulosic biomass. After this review, it became apparent that recently developed compositional analysis techniques involving multivariate regression models were both expedient and accurate. This observation was the motivation for Chapter 3, where mixtures of cellulose, hemicellulose, and lignin were used to generate a calibration sample set for

a completely FT-IR based multivariate regression model. Various data preprocessing regimes were studied and the final PLS regression model that was built could predict, with reasonable certainty, the amount of polysaccharides and lignin within a given sample of lignocellulosic biomass.

Chapter 4 describes the synthesis of a hydrazide-containing controlled/living radical polymerization initiator that would covalently attach to the reducing-end of a polysaccharide. The goal was to expedite the coupling reaction between the initiating species and the reducing-end by utilizing a highly reactive hydrazide moiety with an alkyl halide initiator on the terminal end. The next intention was to polymerize synthetic monomers *via* Cu(0)-mediated LRP from the reducing-end initiating site on the polysaccharide, resulting in an amphiphilic block copolymer. Unfortunately, the high reactivity of the hydrazide, coupled with the labile nature of the alkyl bromide moiety resulted in the loss of the bromide functionality while coupling to the reducing-end of dextran. Possible sources of this instability were discussed as well as plausible solutions for synthesizing similar polysaccharide containing block copolymers.

Chapter 5 provides a facile approach towards polysaccharide containing micelle analogues by grafting the reducing-end of dextran to hyperbranched synthetic copolymers containing pendant amine groups. First, CCTP was used to produce poly(MMA-*co*-EGDMA) by polymerizing MMA and EGDMA in the presence of catalytic quantities of a cobalt(II) complex. After isolating the resulting hyperbranched poly(MMA-*co*-EGDMA), thiol additions were performed on the pendant vinyl groups with cysteamine hydrochloride in the presence of organic base to give an amine terminated hyperbranched polymer, poly(MMA-*co*-EGMDA)-NH_2. After isolating the polymers following thiol additions, dextran was attached at its reducing end to the pendant amine functionalities of the hyperbranched polymer *via* reductive amination. The resulting core-shell polymers, poly(MMA-*co*-EGDMA)-*b*-dextran, showed amphiphilic

behaviour in water that could be tuned based on the polymerization recipe of the hyperbranched core.

Chapter 6 contains a summary of the thesis and recommendations to improve or complement the endeavours discussed herein.

1.3 References

(1) Kunz, M. *Biocatal Biotransformation* **2011**, *26*, 128–132.
(2) da Costa Sousa, L.; Chundawat, S. P. S.; Balan, V.; Dale, B. E. *Current Opinion in Biotechnology* **2009**, 339–347.
(3) Wold, H. *Mathematical Statistics* **1980**, *6*, 333–346.
(4) Chen, H.; Ferrari, C.; Angiuli, M.; Yao, J.; Raspi, C.; Bramanti, E. *Carbohydrate Polymers* **2010**, *82*, 772–778.
(5) Hames, B.; Thomas, S.; Sluiter, A.; Roth, C.; Templeton, D. *Appl Biochem Biotechnol* **2003**, *105*, 5.
(6) Tucker, M. P.; Nguyen, Q. A.; Eddy, F. P.; Kadam, K. L.; Gedvilas, L. M.; Webb, J. D. *Appl Biochem Biotechnol* **2001**, *91-93*, 51–62.
(7) Cherfi, A.; Févotte, G. *Macromol. Chem. Phys.* **2002**, *203*, 1188–1193.
(8) Dube, M.; Li, L. *Polymer-Plastics Tech. & Eng.* **2010**, *49*, 648–656.
(9) de Faria, J. M., Jr.; Machado, F.; Lima, E. L.; Pinto, J. C. *Macromol. React. Eng.* **2010**, *4*, 11–24.
(10) Workman, J.; Koch, M.; Lavine, B.; Chrisman, R. *Anal. Chem.* **2009**, *81*, 4623–4643.
(11) Klemm, D.; Heublein, B.; Fink, H.-P.; Bohn, A. *Angew. Chem. Int. Ed.* **2005**, *44*, 3358–3393.
(12) Zhang, Y. H. P. *J Ind Microbiol Biotechnol* **2008**, *35*, 367–375.
(13) Zhang, J.; Zhang, H.; Wu, J.; Zhang, J.; He, J.; Xiang, J. *Phys. Chem. Chem. Phys.* **2010**, *12*, 1941–1947.
(14) Houga, C. M.; Meins, J.-F. O. L.; Borsali, R.; Taton, D.; Gnanou, Y. *Chemical Communications* **2007**, 3063.
(15) Seoud, El, O. A.; Koschella, A.; Fidale, L. C.; Dorn, S.; Heinze, T. *Biomacromolecules* **2007**, *8*, 2629–2647.
(16) Mäki-Arvela, P.; Anugwom, I.; Virtanen, P.; Sjöholm, R.; Mikkola, J. P. *Industrial Crops & Products* **2010**, *32*, 175–201.
(17) Spain, S. G.; Gibson, M. I.; Cameron, N. R. *J. Polym. Sci. A Polym. Chem.* **2007**, *45*, 2059–2072.
(18) Tizzotti, M.; Charlot, A.; Fleury, E.; Stenzel, M.; Bernard, J. *Macromol. Rapid Commun.* **2010**, *31*, 1751–1772.
(19) Grzybowski, B. A.; Wilmer, C. E.; Kim, J.; Browne, K. P.; Bishop, K. J. M. *Soft Matter* **2009**, *5*, 1110.
(20) Malmström, E.; Carlmark, A. *Polym. Chem.* **2012**, DOI: 10.1039/c1py00445j.
(21) Schatz, C.; Lecommandoux, S. *Macromol. Rapid Commun.* **2010**, *31*, 1664–1684.
(22) Wang, J.; Matyjaszewski, K. *Journal of the American Chemical Society* **1995**, *117*, 5614–5615.
(23) Percec, V.; Guliashvili, T.; Ladislaw, J. S.; Wistrand, A.; Stjerndahl, A.; Sienkowska, M.

J.; Monteiro, M. J.; Sahoo, S. *Journal of the American Chemical Society* **2006**, *128*, 14156–14165.
(24) Roy, D.; Semsarilar, M.; Guthrie, J. T.; Perrier, S. *Chem. Soc. Rev.* **2009**, *38*, 2046–2064.
(25) Voit, B.; Appelhans, D. *Macromol. Chem. Phys.* **2010**, *211*, 727–735.
(26) Verma, M. S.; Gu, F. X. *Carbohydrate Polymers,* **2012**, DOI: 10.1016/j.carbpol.2011.11.025.
(27) Heuts, J. P. A.; Smeets, N. M. B. *Polym. Chem.* **2011**, *2*, 2407–2423.

Chapter 2

Literature Review

Compositional Analysis of Lignocellulosic Biomass: Conventional Methodologies and Future Outlook

Significant portions of this chapter were submitted for publication in *Biomass and Bioenergy* on December 27, 2011 (Ref.: JBB-D-11-01243).

Abstract

As mentioned in Chapter 1, the composition and structural properties of lignocellulosic biomass have significant effects on its conversion to fuels, biomaterials, and building-block chemicals. Specifically, the recalcitrance to modification and compositional variability of lignocellulose make it challenging to optimize and control the conditions under which the conversion takes place. Various characterization protocols have been developed over the past 150 years to elucidate the structural properties and compositional patterns that affect the processing of lignocellulose. The early characterization techniques were developed to estimate the relative digestibility and nutritional value of plant material after ingestion by ruminants and humans alike (e.g., dietary fibre). Over the years, these empirical techniques have evolved into statistical approaches that give a broader and more informative analysis of lignocellulose for conversion processes, to the point where an entire compositional and structural analysis of lignocellulosic biomass can be completed in minutes, rather than weeks. The use of modern spectroscopy and chemometric techniques has shown promise as a rapid and cost effective alternative to traditional empirical techniques. This review serves as an overview of the compositional analysis techniques that have been developed for lignocellulosic biomass in an

effort to highlight the motivation and migration towards rapid, accurate, and cost-effective data-driven chemometric methods. These rapid analysis techniques can potentially be used to optimize future biorefinery unit operations where large quantities of lignocellulose are continually processed into products of high value.

2.1 Introduction

Lignocellulose is derived from the cell walls of plants and is mainly composed of cellulose (30-50%), hemicellulose (15-35%), and lignin (10-30%).[1] Cellulose is a linear polysaccharide comprising β(1→4) linked glucose units with a high degree of hydrogen bonding that makes a significant contribution to the structural integrity and recalcitrance to modification of the lignocellulose composite. Cellulose can be found in various macromolecular conformations that can be generally classified as crystalline or amorphous, where crystalline cellulose poses a greater energy penalty to hydrolyze than amorphous cellulose. Hemicellulose is a branched heteropolysaccharide of both pentose and hexose units that interact with cellulose chains and can sometimes be found as a crosslinking material between cellulose and lignin, mainly imparting enhanced structural stability to the cell wall.[2] Finally, lignin is a complex aromatic heteropolymer that can be found as a binding agent along the outside of the lignocellulose microfibril, providing structural rigidity and protection to the energy rich cellulose fibres.[3,4] Lignin comprises β-aryl ether, biaryl ether, phenylcoumarane, biphenyl, pinoresinol, or diaryl propane linked *p*-coumaryl, coniferyl, and sinapyl alcohol units.[4] Because of its highly branched and irregular structure formed *via* random oxidative radicalization of phenols[3,5], lignin poses a formidable barrier to digestion *via* enzymatic routes. While enzymes capable of catalyzing cellulose hydrolysis (e.g. cellulase) are produced by a wide variety of prokaryotes, eukaryotes, insects, and even crustaceans, the layers of lignin surrounding the cellulose fibres serves to protect the sugar-rich constituents from rapid degradation. Despite this protective role

of lignin, a number of microorganisms are capable of lignin degradation; for example, white-rot fungi produce lignin peroxidase and manganese peroxidase enzymes with lignolytic activity.[4] Interestingly, a number of these microorganisms with lignolytic activity seem to function in a symbiotic relationship with the host plant.[4] Human beings, on the other hand, have discovered that lignin is the primary hindrance to the facile production of fuels and chemicals from lignocellulosic materials, including bioethanol.[3] Figure 2.1 illustrates the source of lignocellulose, the geometric arrangement of its components within the microscopic fibrillar structure, and the composition of each of the three major biopolymers found therein.

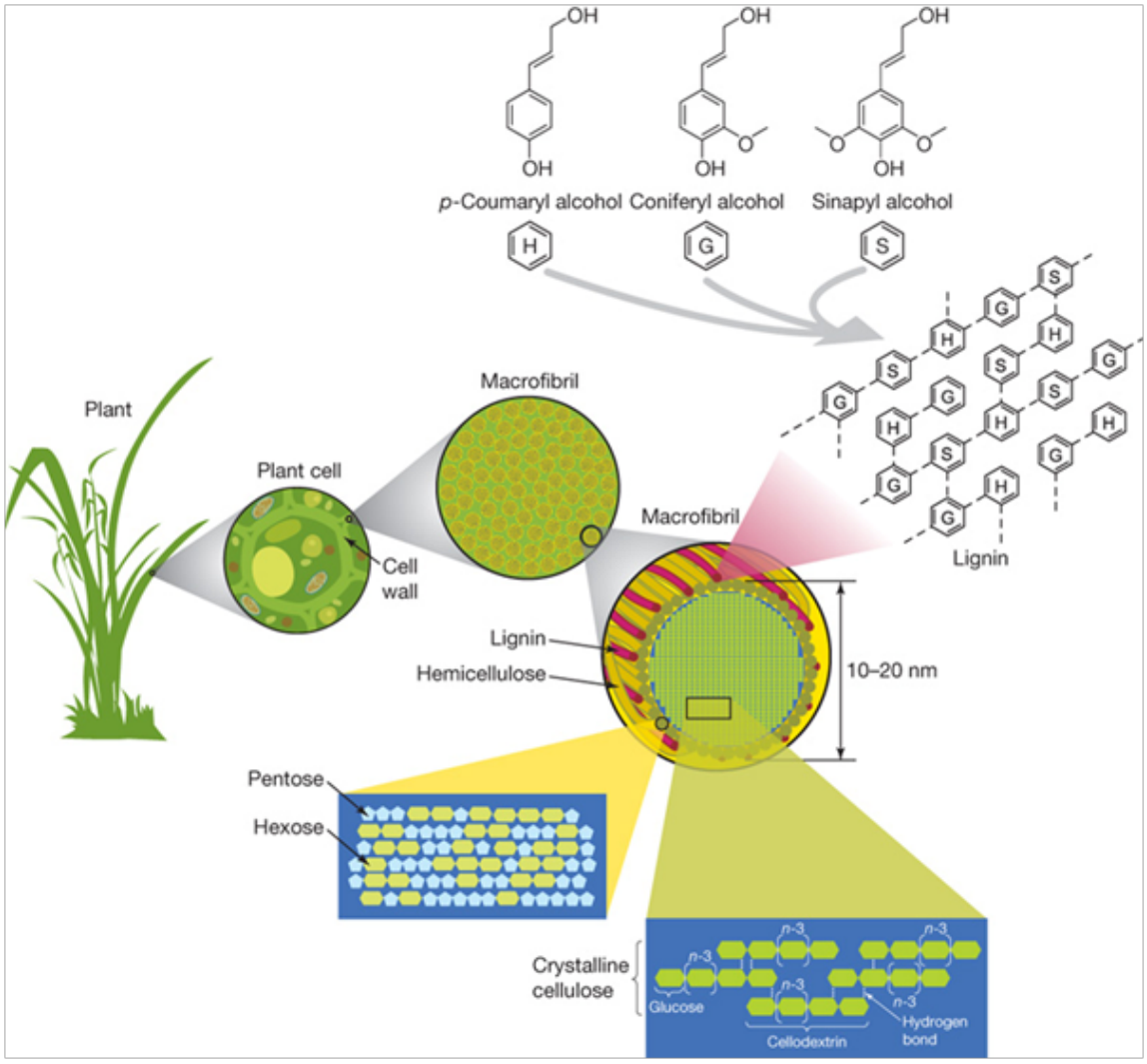


Figure 2.1. The main components of lignocellulose (cellulose, hemicellulose, and lignin) are found within the plant cell wall. Cellulose consists of hydrogen-bonded homopolymer chains of glucose that are attached to one another *via* glycosidic bonds. Hemicellulose is a branched and irregular heteropolymer of both pentose and hexose sugar units. Lignin is a highly complex and aromatic heteropolymer of *p*-coumaryl, coniferyl, and sinapyl alcohol units (Reproduced with permission from Rubin[6]).

The characterization of raw materials, whether for composition, structure, or some other salient property, is a fundamental first step in most experiments and industrial conversion processes. As reviewed by FitzPatrick et al., a promising area of research and industrial interest

is biorefineries utilizing lignocellulosic biomass for the production of value-added chemicals and fuels from non-food crops.[7] These lignocellulosic biorefineries seek to capitalize on the benefits of an inexpensive and renewable feedstock as opposed to more traditional processes that utilize finite petroleum sources. However, a thorough characterization protocol is required prior to the conversion of lignocellulosic biomass because of its highly variable and recalcitrant composition. These properties ultimately limit the range of products that can be derived from a particular lignocellulosic feedstock in a cost efficient manner. Furthermore, the composition and structure of the lignocellulosic feedstock serve a critical role in closing the mass balance of the process and also have a large influence on the efficacy of pretreatment and subsequent processing steps.[8] Because the most expensive aspects of lignocellulose conversion depend on the nature of the raw material, accurate and rapid front-end biomass characterization – either in the field where the biomass is grown or on the processing line where the conversion takes place – is absolutely necessary for an economically feasible biorefinery.

There are many options to consider with respect to characterizing lignocellulose, but very few published papers discussing their relative advantages and disadvantages with respect to industrial applications. For example, Giger-Reverdin[1] reviewed the various protocols available at the time for estimating cell wall composition and digestibility of lignocellulosic materials, but presented the review from a ruminant nutrition perspective. However, one exception is a thorough review paper recently published by Sluiter et al.[9]. Sluiter et al. reviewed the use of sulfuric acid to affect various stages of cell wall dissolution during the summative compositional analysis of lignocellulosic feedstocks.[9] The authors also mentioned the limitations of these methods as well as some suggested improvements, most of which will also be discussed in the paragraphs below.[9] In addition, Sluiter et al. mention that multivariate calibration methods involving NIR can expedite the compositional analysis process.[9] While some univariate

analytical techniques are well suited for the laboratory environment, other multivariate approaches are more robust and can be implemented in a real-time process control scenario over a wide array of feedstock compositions and product formulations.

The following review will serve as a comprehensive examination of the analytical methods available for the characterization of lignocellulosic biomass, its derivatives, and their suitability for application in biorefinery processes. Furthermore, it will be shown that an increase in research focused on biorefinery-relevant Process Analytical Technologies (PAT) and data processing techniques could present rapid and inexpensive alternatives for experimental purposes and quality assurance in both research and industrial setting.

2.2 Historical Perspective on the Compositional Analysis of Lignocellulose

Contemporary methods for the compositional analysis of lignocellulosic materials have been ultimately derived from those developed to assess the digestibility of plant-based animal feed. A parallel interest in the assessment of the chemical composition of plant-based lignocellulosic materials has arisen from the practical needs related to the production of renewable fuels, chemicals, and materials. Both of the aforementioned metrics, nutrient requirements and chemical composition, have served as the basis for the empirical study of lignocellulosic biomass for over 150 years. The following review tracks the progress and evolution of lignocellulosic characterization methods from laborious, wet-chemical methods towards rapid, high-throughput techniques that take advantage of a data-driven approach congruent with the latest developments in the field of chemometrics.

2.2.1 Weende Method

The seminal method for quantifying crude fibre (CF) was the proximate or Weende analysis proposed by Henneberg and Stohmann[10], where samples of animal feed are first dried at

100 °C, extracted using ether to remove lipids, then refluxed using a weak base (1.25% sodium hydroxide) followed by a strong acid (1.25% sulfuric acid) to digest cell wall material leaving behind an indigestible residue, providing a gravimetric estimate of crude fibre (CF).[11] Despite being standardized by Henneberg and Stohmann, the roots of the Weende analysis system can be traced back to the alkaline extractions by Gorham[12] and similar acid-alkaline extractions by Horsford[1,13]. Although it was modified and accepted as an official AOAC method for the analysis of crude fibre in food AOAC 978.10[14], the Weende system can yield inaccurate results and may underestimate fibre content by more than 50% under the conditions prescribed by the method.[15] The limitations associated with the crude fibre method can be attributed to complexities in the fractionation and analysis of the substrate: the highly variable recovery of cellulose, hemicellulose, and lignin in the crude fibre fraction depends on the ratios of the three components in the starting material. Materials with high hemicellulose and moderate lignin content like grasses tend to provide the lowest recovery of crude fibre, while vegetables with low lignin and high cellulose content, tend to provide the highest cell wall recovery as crude fibre.[11] The analysis is further complicated by the heterogeneous nature of the constituent macromolecules of lignocellulosic biomass, lignin being one of the most highly variable structures.[16] Despite its shortcomings, the Weende system is still routinely used in the food and agricultural industries for the determination of CF – a well-known and understood parameter for estimating fibre digestibility. For example, the Weende analysis was used by Rzedzicki et al.[17] to estimate CF in breakfast cereals. To date, the Weende system has not been used extensively in a lignocellulosic biorefinery context.

2.2.2 Klason Lignin Method

In 1893, Klason published two papers on the analysis of black liquor from the Kraft pulping process and developed his techniques over the next 40 years.[18] The pioneering

techniques published by Klason[19] are still commonly used for the quantification of lignin in biomass. Klason lignin (KL) is defined as the mass of the residue left behind after a sample is treated with 64-72% sulfuric acid thus removing all polysaccharides.[20] While the Klason technique has been accepted as a TAPPI standard (TAPPI Test Method T222 om-88[21]) it has some limitations: lignocellulosic materials with a high fraction of acid-soluble lignin will result in underestimated lignin content, while materials with highly recalcitrant lignin that are resistant to acid treatment will result in overestimated lignin values.[22] Despite these limitations, the KL method has been modified and adapted for the quantification of both acid insoluble (TAPPI Test Method T222 om-06[23]) and soluble (TAPPI Test Method T250[24]) lignin in lignocellulosic biomass as reviewed by Giger[25] and Van Soest[11]. Furthermore, the KL method remains in rigorous use today in agricultural, food and biochemical research.[26-28]

2.2.3 USDA FPL Methods

Over the next 20 years following the development of the KL technique, the United States Department of Agriculture (USDA) Forest Products Laboratory (FPL) applied the KL methods for the determination of lignin in wood samples.[29] Furthermore, scientists from the USDA FPL adopted the methods of Saeman et al.[30] for the quantitative saccharification of wood and cellulose in their research. Briefly, the hydrolysis of cellulose to reducing sugars was achieved by treating ground (30 mesh) and dried woody material with 72% sulfuric acid for 45 minutes at 30 °C followed by a rapid secondary hydrolysis in an autoclave at 15 psi of steam pressure for 1 hour or by boiling the solution for 4.5 hours at atmospheric pressure. The hydrolysate was then assessed for reducing sugar content using the Shaffer and Somogyi[31] method. The use of a secondary autoclave step where numerous samples could be hydrolyzed at once was introduced to decrease the time required for reducing sugar analysis while ensuring reproducible results. Saeman et al. did not discuss lignin quantification, however, in a follow-up publication, they elaborated on this

technique and acknowledged that their use of 72% sulfuric acid for carbohydrate removal was similar to the KL method and that the acid concentration selected was satisfactory since it was both expedient and accurate.[30,32] The attempt to improve the reproducibility and throughput of these methods by Saeman et al.[30,32] is notable, as these metrics defined a future approach in process analytics for lignocellulosic biomass. The hydrolysis methods developed by Saeman et al.[32] have recently been used by Tsubaki et al.[33] to assess the monosaccharides present in *Prunus mume* stones – the heavily lignified seed within the fruit that is often discarded.

2.2.4 Dietary Fibre

As stated in the introduction, agricultural sciences have greatly influenced progress in the compositional analysis of lignocellulosic biomass. In 1953, Hipsley first used the term "dietary fibre" as a description for plant cell wall materials comprising cellulose, hemicellulose, and lignin that are indigestible by the human digestive tract.[34,35] Figure 2.2 provides a description of the relationship between dietary fibre fractions and chemical constituents.

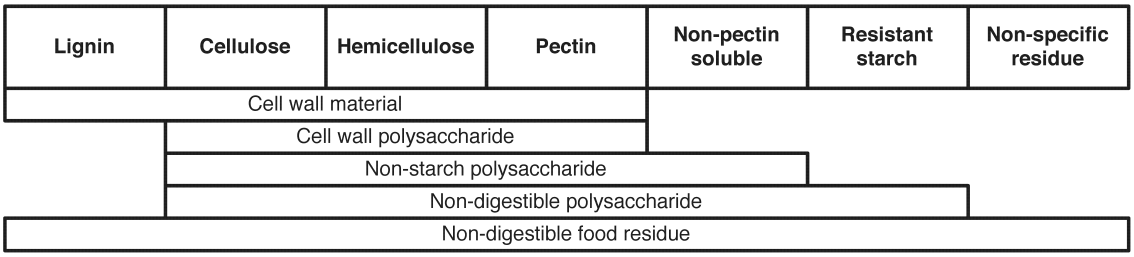


Figure 2.2. Definitions of dietary fibre fractions (Reproduced with permission from Monro and Burlingame[36]).

This definition of dietary fibre became the basis for future work in the field of agricultural and food sciences, where an understanding of the nutritional value of feedstuffs is integral. The underlying concepts behind the dietary fibre and USDA FPL methods for the determination of

cellulose, hemicellulose and lignin content were subsequently used almost interchangeably for the next 50 years.

2.2.5 Saeman Pentosan Method

Approximately 10 years after their method development for wood hydrolysis and reducing sugar determination, Saeman et al. optimized the chromatographic techniques that were adopted as a TAPPI standard for pentosan (i.e. five carbon sugar polymer) determination.[37] The authors showed that although satisfactory agreement with the TAPPI standard was observed for materials with high pentosan content, at low pentosan content, the standard showed high errors due to residual pentosan bound to unhydrolyzed cellulose. However, the errors caused by residual pentosan were partially eliminated by introducing a two step hydrolysis whereby the pulp was first treated in strong mineral acid followed by a secondary hydrolysis with dilute acid.[37] Despite improvements to pulp fractionation (i.e., introducing a two step hydrolysis), Saeman et al.[37] questioned the adequacy of using paper chromatography for the quantification of released pentosans in the hydrolysate. At this time, the difficulties associated with the selection of hydrolysis conditions were becoming apparent. Conditions that were too mild resulted in incomplete hydrolysis, and conditions that were too harsh yielded a hydrolysate with mono-sugars that were partially decomposed. The selection of appropriate hydrolysis conditions was further confounded with inter- and intra-species variation in lignocellulosic biomass, where methods applied for softwood species might not be effective for hardwoods or herbaceous species like switchgrass. The methods of Saeman et al.[32] have recently been used in a biorefinery context by Houghton et al.[38] for quantitative carbohydrate and lignin determination of wheat straw and corn stover subjected to drying temperatures between 45 °C and 100 °C. Houghton et al. were interested in determining the optimum drying temperature for lignocellulose samples because drying is often used as the first step in conversion processes and the quality of the biomass can be

negatively impacted at higher temperatures.[38] Thus, careful selection of drying temperature can afford energy savings, while minimizing sample deterioration.

2.2.6 Monoethanolamine Method

Nelson and Leming optimized the monoethanolamine (MEA) method for the gravimetric quantification cellulose after lignin and other carbohydrates have been removed.[39] Originally developed approximately 18 years earlier by Wise et al.[40] for woody biomass, Nelson and Leming determined the optimal conditions for the MEA method with respect to conditions of reflux time and temperature, MEA-solids ratio, and the strength of the bleach.[39] The usefulness of the MEA method on three representative types of agricultural residues, fibrous (i.e. wheat straw), hard fibrous (i.e. abaca plant), and non-fibrous (i.e. peach pits) was evaluated in the study.[39] A report published by Reid et al.[41] determined that the original MEA method[40] "compared favourably" with similar methods of the time, namely the Cross and Bevan (as cited by Reid et al.[41]) and Norman-Jenkins[42] methodologies for cellulose determination. Furthermore, the harsh conditions employed by the Kürschner-Hoffer[43] method resulted in cellulose fractions that were considerably lower than the Nelson and Leming MEA method[39].

The MEA method was extensively used by Cunningham et al.[44-47] to quantify cellulose in wheat straw and sorghum. Foyle et al. also used the MEA method for the analysis of sugar in waste paper and straw.[8] In their study, Foyle et al.[8] reported that the MEA method was time consuming and questionable since Claus et al.[48] raised concerns as to whether MEA alone could remove sufficient quantities of hemicellulose from beech and spruce wood. As a result, Claus et al. suggested that an additional sulfuric acid hydrolysis stage might need to be introduced to remove hemicelluloses.[48] More recently, Banerjee et al.[49] used the MEA method described by Foyle et al.[8] to determine total cellulose content in rice hull following wet air oxidation

pretreatment. Aside from the aforementioned studies, MEA is scarcely used in lignocellulosic biorefinery research.

2.2.7 Permanganate Method

Tasman and Berzins[50] developed the permanganate method for lignin, which served as the basis for measuring the Kappa number of wood in the pulp and paper industry. A linear relationship was found between the Kappa number and Klason lignin for both Kraft and sulphite cooking reagents up to approximately 22% and 15% lignin content for softwoods and hardwoods, respectively.[51] Further relationships are available that relate Kappa, permanganate, chlorine, and hypochlorite numbers to Klason lignin values.[52] A significant limitation of the permanganate method is the non-specific oxidation and retention of phenolic and unsaturated substances like pigments or proteins that will overestimate lignin content.[53] Because of these limitations, the permanganate method as developed by Tasman and Berzins[50] is rarely used in the analysis of lignin content.

2.2.8 Detergent Fibre Methods

The detergent fibre procedure, as elaborated by Van Soest[54,55], Van Soest and Wine[56] and described in procedural form by Goering and Van Soest[57], has been widely used for the fractionation and compositional analysis of plant-based animal feed. Although generally used in agricultural sciences to gauge the digestibility of animal feed, this forage fibre technique can also be adapted for the analysis of lignocellulose in a biorefinery context. The neutral detergent fibre (NDF) and acid detergent fibre (ADF) analyses are based on the selective chemical extraction of the entire cell wall material using neutral detergent solution and then acid detergent solution under reflux, followed by filtration and recording the mass of the remaining residue. The solid residue remaining after treatment with neutral detergent solution, that is NDF, contains attached

protein, hemicellulose, cellulose, lignin, lignified nitrogen compounds, heat-damaged proteins, keratin, and silica. The acid detergent solution removes hemicellulose (leaving ADF residue) and a final step is involved where the sample is treated with 72% sulfuric acid to completely solubilize highly recalcitrant materials like lignin, heat-damaged proteins, or keratin, leaving acid detergent lignin (ADL) residue.[58] Table 2.1 lists the groups of plant materials that can be roughly quantified using the NDF and ADF procedures.

Table 2.1. Various forage organic matter that can be determined using detergents (adapted from Van Soest[59]).

Fraction	**Components**
Cell contents (soluble in neutral detergent)	Lipids Sugars, organic acids, water soluble-material Starch Non-protein nitrogen Soluble protein Pectin
Cell-wall constituents (fibre insoluble in neutral detergent)	Attached protein
Soluble in acid detergent	Hemicellulose
Insoluble in acid detergent (acid detergent fibre)	Cellulose Lignin Lignified nitrogen compounds Heat-damaged proteins Keratin Silica
Soluble in 72% sulfuric acid	Cellulose

Thus, ADF minus ADL can be used as a rough estimate of cellulose content, however these values are often overestimated due to the presence of hemicelluloses (e.g., xylans) in ADF and underestimated when heat-damaged proteins contaminate ADL.[15] Likewise, NDF minus ADF can be used to estimate hemicellulose content but these values can be overestimated by non-extracted protein in NDF and underestimated by non-extracted hemicelluloses in ADF.[15] Total lignin content is often underestimated by ADL and better predicted by the KL method, although

ADL tends to be more consistently correlated with digestibility than KL.[60] These limitations were supported by a study involving several herbaceous biomass energy crops at various stages of maturity where the detergent fibre method consistently overestimated cellulose and hemicellulose content while lignin content was substantially underestimated.[61] Despite its shortcomings, the NDF procedure has seen widespread use as a pretreatment for the AOAC standard ADF method because of its popularity in the agricultural and food sciences and the vast amount of data that has been accumulated using the technique over the years.[15] In actuality, the cell wall material is often underestimated in NDF since pectic substances can be solubilized in the process.[11,15] Moreover, the neutral detergent resistant residue can often become contaminated by starch, animal keratin, and soil minerals.[1,11]

Various modifications of the forage fibre analysis technique have been developed in the last 40 years. For example, Cherney et al. showed that overestimates of fibre content caused by high levels of starch in the forage fibre samples could be mitigated using the highly specific and environmentally benign enzyme α-amylase instead of decahydronaphthalene (decalin).[62] Moreover, Cherney et al.[62] showed that 2-ethoxyethanol, a harmful and environmentally hazardous compound, was no longer required for the forage fibre analysis. An inter-disciplinary study by Mertens[63] confirmed that this new α-amylase neutral detergent fibre (aNDF) procedure was generally applicable to a variety of forage feeds, providing reproducible results from a number of laboratories. As a result, Mertens suggested that the aNDF procedure be accepted as a standard protocol.[63] In addition to the requirement of large amounts of harmful reagents, the forage fibre technique also suffers from high equipment costs due to its low throughput. Each large, representative sample must be refluxed and filtered separately using adequate amounts of detergent solution and solvent. Efforts have been made to reduce sample size, processing time, and subsequent cost of the forage fibre technique through semi-automation, but a fully-stocked

laboratory with experienced staff is still a prerequisite for reproducible results.[15,64] Other official methods for determining fibre content in foodstuffs have been reviewed in DeVries et al.[35]. Despite its limitations, the detergent fibre method has recently been used by Chen et al. as a reference technique for advanced multivariate analysis of wood samples using infrared spectroscopy.[65]

2.2.9 Moore and Johnson (Modified USDA FPL) Method

Moore and Johnson[66,67] developed a method for the analysis of pulp and wood sugars including pentosans at the USDA FPL. In the Moore and Johnson method, hydrochloric acid is used to determine pentosan content by hydrolyzing and converting pentoses to furfural.[8] Furfural is then quantified using the orcinol technique. Foyle et al.[8] found that the Moore and Johnson method[66,67] underestimated pentose content compared to a 64% sulfuric acid digestion technique and attributed this discrepancy to incomplete hydrolysis of pentoses or incomplete conversion to furfural.

2.2.10 Trifluoroacetic Acid Method

Fengel and Wegner[68] developed the dilute trifluoroacetic acid (TFA) method for the determination of cellulose and hemicellulose by proxy of their hydrolysate sugars. Two methods were described by Fengel and Wegner for materials with high and low lignin content, respectively.[68] A description of the two methods, along with a modified method for waste paper was published by Foyle et al., although all TFA methods sometimes gave poor results due to incomplete hydrolysis, low recovery rates, and low reproducibility.[8] Furthermore, hydrolysis of lignocellulosic materials like oat straw using TFA under mild temperatures (e.g. 37 ^{o}C) can take more than 8 days for significant quantities of reducing sugars to be released.[69] Notwithstanding the aforementioned limitations, Windeisen et al.[70] used the TFA method to monitor the

compositional changes of beech wood under different temperature loads. Aside from these few studies, the TFA method has been rarely applied to lignocellulosic biorefinery research.

2.2.11 Grohmann Method

Grohmann et al.[71] modified the methods of Moore and Johnson[66,67] and developed a milder 64% sulfuric acid hydrolysis technique for the determination of monomeric sugars *via* ion-moderated partition chromatography using a Bio-Rad HPX-87P column. The sulfuric acid concentration was chosen to minimize acid consumption and sulfonation of cellulose while ensuring reasonable rates of hemicellulose removal.[71] With the exception of Niu et al.[72] who employed these methods to determine cellulose, hemicellulose and lignin content in rice straw prior to a novel photocatalytic pretreatment, the methods developed by Grohmann et al.[71] have rarely been used in recent years.

2.2.12 Prosky Dietary Fibre Method

Prosky et al.[73] developed an enzymatic and chemical dietary fibre method that allowed for the entire cell wall to be retained, including starch proteinacious material, rendering this analytical technique unsuitable for the quantification of individual cell wall components: cellulose, hemicellulose and lignin. Following modifications to the concentration of alcohol and buffers, time of incubation, sample preparation, as well as an inter-laboratory study[74], the Prosky dietary fibre (DF) technique was designated as AOAC Official Method 985.29, Total Dietary Fibre in Foods—Enzymatic-Gravimetric Method and as AACC Approved Method 32-05[35]. These methods retain all components of the cell wall in their analyses and are, therefore, poorly suited for the compositional analysis of lignocellulosic biomass.

2.2.13 Uppsala Method

Soon after Prosky's work, the Uppsala method was developed by Theander and

Westerlund[75] to estimate the composition of cell wall components using a chemical approach. The Uppsala method allows for the determination of uronic acid residues, extractives, water-soluble carbohydrates, acid-soluble lignin, and Klason lignin. The Uppsala method begins by repeated sonication of the sample in 80% (v/v) ethanol followed by centrifugation to pellet form and fresh ethanol extraction.[52] An additional extraction step using petroleum ether can be introduced to remove fatty substances if present in significant quantities (i.e. over 5%).[76] Following extraction, starch is removed by placing the sample in acetate buffer using thermostable α-amylase and amyloglucosidase. The starch-free, ethanol-soluble fibre is then precipitated using fresh 80% (v/v) ethanol and subjected to sulfuric acid hydrolysis. Released neutral sugars are quantified using colorimetry or chromatography (i.e. HPLC or GC) as alditol acetates.[15,77] Acidic cell wall sugars like galacturonic and glucuronic acids, are measured colorimetrically with 3,5-dimethylphenol or as carbon dioxide released as a result of decarboxylation.[15,76,77] Lignin is estimated gravimetrically as Klason lignin: ash-free acid-insoluble residue.[15,77] Total dietary fibre is reported as amylase-resistant polysaccharides plus Klason lignin.[78]

Following successful inter-laboratory studies by Milne et al.[79] and Theander et al.[78], the Uppsala method was adopted as AOAC Official Method 994.13 for the quantification of soluble and insoluble polysaccharides, including resistant starch and lignin. The UK (United Kingdom) method is quite similar to the Uppsala method with the exception that it only includes plant cell wall non-starch polysaccharides in the analysis rather than dietary fibre, Klason lignin and enzyme-resistant starch as elucidated by the Uppsala method.[76] The Uppsala and UK methods have been contrasted in greater detail in Theander et al.[76]. It should be noted that, in addition to being complex and laborious, the Uppsala method is susceptible to significant protein contamination, leading to overestimated KL values.[52]

The dietary fibre methods, while reasonable predictors of lignocellulose content in biomass, are generally used in validation or screening studies for human and animal nutrition. These standardized fibre methods using wet-chemical procedures were not developed with a continuous biorefinery process in mind. This is an important factor to consider when choosing a characterization protocol: time and material costs associated with wet-chemical techniques will increase correspondingly if integrated in an industrial-scale process. This underscores the need for rapid, non-destructive compositional analysis techniques for future lignocellulosic biorefineries.

2.2.14 Acetyl Bromide (Derivatization Followed by Reductive Cleavage) Method

Lignin remains one of the most difficult constituents of lignocellulosic biomass to isolate and characterize with no single method being considered a standard for all feedstocks.[52] Of the numerous wet-chemical methods available for quantifying lignin and its structure, the acetyl bromide or Derivatization Followed by Reductive Cleavage (DFRC) method stands out as one of the most consistent methods for determining its digestibility.[16,53] The DFRC method overcomes some of the limitations introduced by the ADL, KL or permanganate techniques; for example, syringyl-rich lignins can become solubilized and lost during ADL determination, KL values of legumes can appear inflated because of protein contamination, and applying the permanganate method to legumes can inflate lignin values because of uronosyl oxidation.[53] Figure 2.3 shows the general reaction scheme of the DFRC method for guaiacyl β-ether units, a lignin model compound.

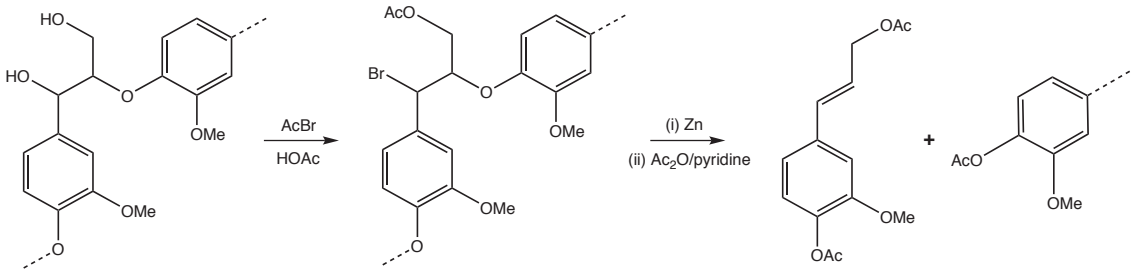


Figure 2.3. General reaction scheme of DFRC method on guaiacyl β-ether units. Acetyl bromide solubilizes the cell wall. The acetylated and brominated lignin units are then reductively cleaved using zinc dust and acetylated using acetic anhydride and pyridine to produce acetylated lignin monomers that are quantifiable using gas chromatography (adapted from Ralph[16]).

In the DFRC method, lignin is first dissolved and brominated in acetyl bromide and acetic acid, followed by reductive cleavage of the aryl ether bonds using zinc dust and acetylation using acetic anhydride and pyridine.[80] The products are acetylated lignin monomers that are generally quantified using mass spectroscopy and gas chromatography. Methods for lignin isolation and characterization have been extensively contrasted and reviewed by Fukushima and Hatfield[53] and Hatfield and Fukushima[52]. While the DFRC method is a convenient method for lignin analysis, xylan degradation is considered a major interference that could result in overestimated lignin content, thus introducing a significant limitation on the method.[81] In order to minimize xylan degradation, the hydrolysis step can be undertaken at a reduced temperature (i.e. 50 °C) for an extended duration (i.e. 2-4 h).[81] A modified DFRC method involving dissolution at 50 °C for 2 h in propionyl bromide and propionic acid instead of acetyl reagents was utilized by Del Río et al. to study native acetylated lignin derived from vascular plants.[80]

2.2.15 NREL Methods (Laboratory Analytical Procedures)

Most methods for the compositional analysis of lignocellulose are bifurcated between fibre analyses, for the food and agricultural sciences, and more rigorous chemical and physical

methods for the chemical and engineering sciences. In recent years, the methods commonly used for the characterization of lignocellulosic biomass for engineering applications have been the standards developed by the National Renewable Energy Laboratory (NREL). Laboratory analytical procedures (LAPs) have been developed for the determination of structural carbohydrates and lignin in biomass[82]; extractives in biomass; total solids in biomass and total dissolved solids in liquid process samples; ash in biomass; sugars, byproducts, and degradation products in liquid fraction process samples; starch in biomass samples using HPLC; protein content in biomass; insoluble solids in pretreated biomass; as well as lignocellulosic biomass hydrolysis; and, fermentation under simultaneous saccharification and fermentation (SSF) protocols.[83] The LAP for the determination of structural carbohydrates and lignin in biomass was evolved from the Uppsala method and ASTM E1758-01[84] with the addition of a section on the determination of lignin using UV/VIS spectroscopy.[85] Briefly, the biomass sample is fractionated using 72% sulfuric acid and filtered through fritted glass crucibles, similar to the NDF procedure. Carbohydrate content in the hydrolysate (filtrate) is then analyzed using high performance liquid chromatography (HPLC) using known hexose and pentose standards. Likewise, acid soluble lignin is determined from the same hydrolysate using UV/VIS spectroscopy. The retentate is ashed and used to determine acid insoluble lignin (and acid insoluble ash and protein, if necessary) using gravimetric techniques. Furthermore, this LAP can be modified for samples containing high quantities of extractives or protein.[82] The main advantage of the NREL LAPs, like most standard procedures, is the ability to compare results from different research projects. However, as with most standard procedures, slight variations or substitutions in feedstock or fractionation procedure can lead to comparability issues between research groups; in other words, a simplified compositional analysis technique might be warranted given the laborious and costly nature of complex procedures requiring fractionation. Currently, there is a publicly available

database containing compositional data of various plant species using the aforementioned NREL LAPs and ASTM methods.[86] The expansion of such databases to include various species and growing conditions will undoubtedly aid the standardized characterization of feedstocks in future biorefineries.

NREL LAP TP-510-42618 has been used by a number of researchers for the compositional analysis of lignocellulosic biomass. [38,87-93] In addition, a number of studies have described techniques other than those developed by NREL for the compositional analysis of lignocellulose using similar fractionation and measurement methods, specifically concentrated sulfuric acid treatment and HPLC analysis.[94] A study conducted by Foyle et al. outlined some major limitations of techniques involving concentrated sulfuric acid and HPLC, in which they stated that the NREL LAP could under-estimate cellulose content in samples analyzed using 72% sulfuric acid because of the sulfonation of cellulose and the inability to accurately measure these compounds using the prescribed BioRad Aminex HPX-87P HPLC column.[8] Furthermore, concentrated acid solutions can degrade hydrolysate sugars to the point where they become undetectable using HPLC.[8] Thus, Foyle et al. opted for a more conservative method developed by Grohmann et al.[71] where a 64% sulfuric acid was used instead of the 72% sulfuric acid technique proposed by Moore and Johnson[66,67] and later adopted by Sluiter et al.[82] for the NREL LAP. However, Foyle et al.[8] noted that because of the relatively mild conditions employed by the Grohmann et al.[71] method, low concentrations of sugars close to the detection limit of the HPLC may not be accounted for, thus contributing to experimental error. Recently, Scarlata and Hyman[95] made further improvement to NREL's HPLC technique by reducing the time required for organic acid, alcohol and furan analysis from about 55 minutes to 10 minutes through the use of a proton form cation-exchange resin stationary phase (Phenomenex Rezex RFQ column) capable of higher temperatures and flow rates relative to the conventional Aminex columns.

Further work by Templeton et al.[96] showed that the uncertainties associated with NREL LAPs, in terms of relative standard deviation (i.e., standard deviation divided by mean), were more heavily influenced by the type of analytical procedure invoked rather than the type of lignocellulosic biomass analyzed. Moreover, the authors suggest that a well-trained analyst should conduct the wet-chemical compositional analysis from beginning to end in order to minimize the uncertainties associated with what was described as a "manually-intensive" process.[96]

The intense research focus on improving lignocellulosic biomass characterization techniques reflects the desire to quickly commercialize novel conversion technologies. Table 2.2 provides a brief summary of the wet-chemical methods discussed in this paper along with their limitations.

Table 2.2. Summary of wet-chemical methods for the compositional analysis of lignocellulosic biomass.

Method of Analysis	Fraction Measured	Limitations of Method	Reference(s)
Proximate, Weende, or Crude fibre	Portion of plant cell wall, complete cellulose recovery	Most non-cellulosic polysaccharides and lignin removed May underestimate fibre content by more than 50%	15, 97, 98
Neutral detergent fibre (NDF)	Incompletely digestible feed fraction, almost complete recovery of grass cell walls	Pectin almost completely removed, protein and starch removal can be problematic	15
Acid detergent fibre (ADF)	Portion of plant cell wall, complete cellulose recovery	A significant portion of lignin is solubilized, especially for grasses (50% or more dissolved)	15, 52
ADF minus ADL	Cellulose	Suffers from limitations of ADF and ADL methods	15
NDF minus ADF	Hemicellulose	Suffers from limitations of NDF and ADF methods	15
Acid detergent lignin (ADL)	Lignin	Lignin Solubilization at ADF step, especially grasses	15
Dietary fibre	Complete recovery	Protein and starch removal	15

	of cell wall polymers	can be problematic	
Uppsala dietary fibre	Total cell wall recovery and composition of the cell wall	Complex and laborious A significant portion of the cell wall protein can remain, potentially inflating Klason lignin values	15, 52
Crampton and Maynard	Cellulose	Small amount of xylan contamination on cellulose Significant cellulose degradation	15, 39, 99
Klason lignin	Lignin	Possible protein and carbohydrate contamination Might underestimate/overestimate lignin content based on ratio of acid soluble/insoluble lignin (especially in legumes)	15, 22, 53
Monoethanolamine (MEA)	Cellulose after lignin and other carbohydrates removed	Time consuming and hemicellulose components might not be adequately removed	8, 39, 41, 48
USDA FPL	Reducing sugars and lignin in woody samples	Limitations of reducing sugar assays Choice of secondary hydrolysis conditions are not always clear	30, 100
UK	Plant cell wall non-starch polysaccharides	Non-specific towards polysaccharides Lignin not quantified	76
Trifluoroacetic acid (TFA)	Carbohydrate polymers: cellulose and hemicellulose	Incomplete hydrolysis is likely to occur (can take days for sugars to be released)	8, 69
DFRC	Lignin	Xylan degradation and contamination can be an issue at elevated temperatures (> 50 °C)	81
NREL LAP TP-510-42618	Structural carbohydrates and lignin in biomass	Underestimated cellulose content due to sulfonation of cellulose at high (72%) sulfuric acid concentration	8

Figure 2.4 illustrates, in chronological format, the 150-year evolution of wet-chemical compositional analysis protocols that have been mentioned in this paper.

Traditional methods for the compositional analysis of lignocellulose appear laborious and expensive, as evidenced by the numerous wet-chemical techniques discussed in the sections above. Consequently, new techniques have been developed utilizing spectroscopic instrumentation and computational algorithms in an effort to replace the traditional wet-chemical methods with an inexpensive and rapid alternative capable of real-time or on-line analysis.

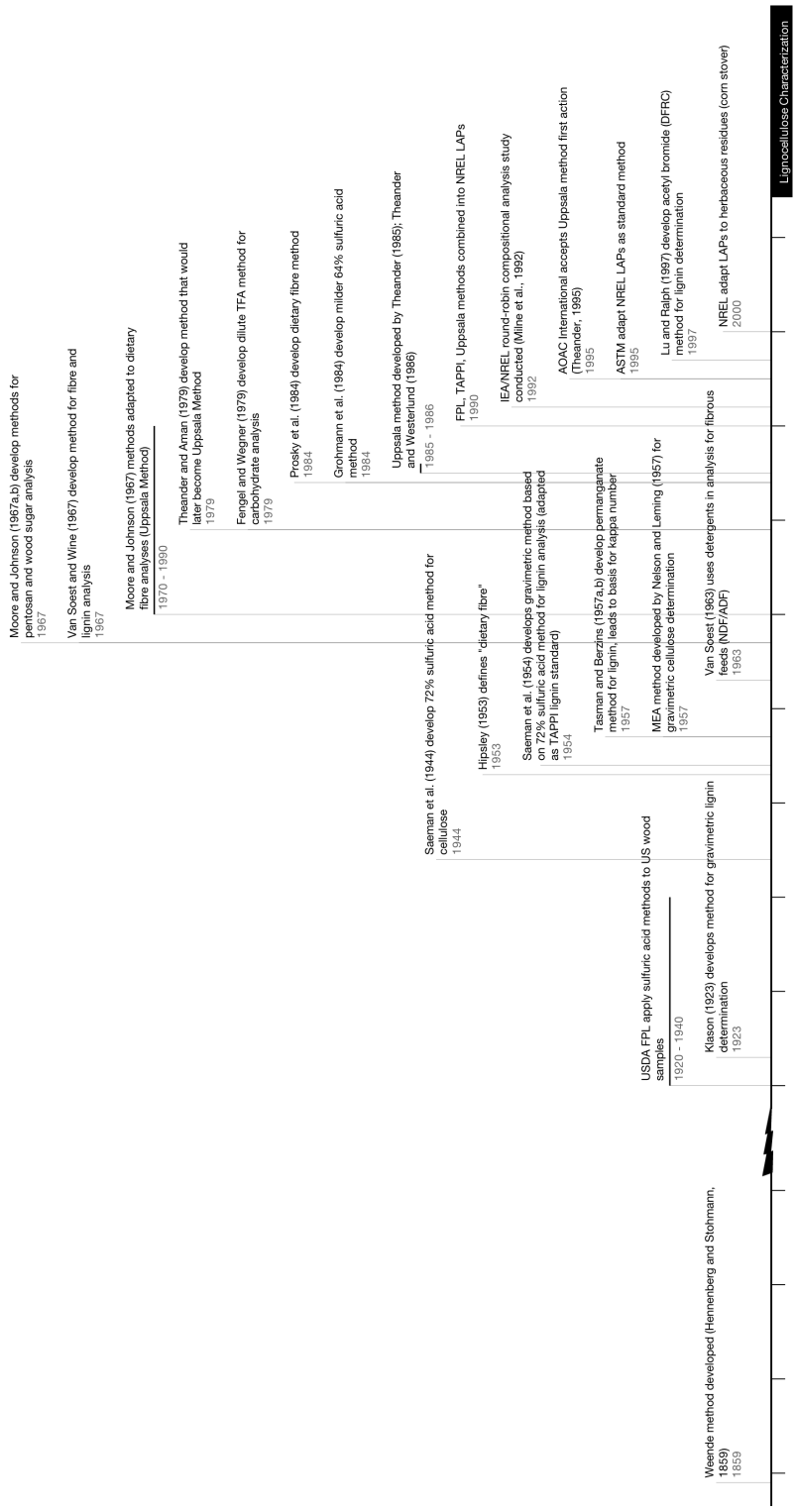


Figure 2.4. Chronology of conventional methods for the compositional analysis of lignocellulosic biomass.[15,29,35,101]

2.3 New Opportunities: Chemometrics and Spectroscopy for Lignocellulose Characterization

2.3.1 Introduction

For a complex biocomposite like lignocellulose, the primary hindrance to an expedient and accurate compositional analysis ultimately depends on the expedience and efficiency of the following criteria:

- The lignocellulose fractionation or separation process and
- The analytical methods used to analyze the components following fractionation.

Near-infrared (NIR), mid-infrared (MIR) absorption and Raman scattering spectroscopy have been used to characterize a diverse range of materials since the required instrumentation is inexpensive, expedient and gives a broad spectrum of compositional data for macromolecules. Under most circumstances, scientists are more interested in comparing lignocellulosic materials rather than determining absolutes based on thorough compositional analyses.[15] As a result, the proliferation of calibration techniques involving NIR, MIR, Raman, and even NMR spectroscopy has enabled large quantities of compositional and structural data to be acquired in real-time. The rapid accumulation of compositional information from spectroscopic techniques can be further supplemented through the use of databases and web groups that are congruent to the latest collaborative efforts in the field of chemometrics.[102] In fact, NIR is a relatively mature technology and has been used for *in situ* measurement of samples since the late 1970s.[103] The use of Fourier Transform Infrared (FTIR) and Fourier Transform Raman (FT-Raman) spectroscopy has also been applied in biomedical research to identify disease patterns in human tissue.[104,105] Furthermore, NIR, FTIR, and FT-Raman can be used as a non-destructive tool in the field of archaeology to obtain compositional information of artifacts, which is required if the item is to be preserved without causing undo damage or modification.[106,107] In fact, infrared instrumentation

coupled with multivariate statistical techniques, which will be discussed in the paragraphs below, has been widely utilized for characterization and quality control purposes in pulp and paper applications.[108-112] Reviews have also been published on the use of infrared and/or Raman spectroscopy for the analysis of lignocellulosic materials such as plant materials[113], wood[114] and carbohydrates[115]. Furthermore, Adapa et al.[116] published a review on the use of infrared and Raman spectroscopy and microspectroscopy for the compositional analysis of agricultural biomass. Likewise, Gierlinger and Schwanninger[117] reviewed the use of near infrared FT-Raman, resonance Raman, and surface-enhanced Raman spectroscopy for the chemical and structural analysis of lignocellulosic materials. As reviewed by Gierlinger and Schwanninger, Raman imaging can also provide valuable information on the spatial distribution of various compounds that comprise the cell-walls of lignocellulosic biomass down to the sub-micron level.[117] Thus, the quantity and distribution of cellulose, hemicellulose, and lignin can be elucidated using infrared absorption or Raman scattering microscopy and imaging techniques.

The use of quantitative ^{31}P-NMR in wood chemistry was reviewed by Argyropoulos[118]; however, the aim of most NMR approaches is to elucidate minute structural differences in the lignocellulose polymers. Thus, it is unlikely that NMR approaches will see widespread adoption as PAT in modern lignocellulosic biorefineries due to their expense and complexity of interpretation.

It is clear that infrared, Raman, and to some extent NMR spectroscopy have shown promise as rapid and accurate methods for the analysis of numerous materials; however, the main problem associated with these techniques can be attributed to the convoluted nature of spectra when analyzing mixtures where difficulties arise in accurately choosing peaks associated with chemicals of interest. However, over the past 20 to 30 years in published literature, the use of chemometric techniques have increased in prevalence. Chemometrics is a term used to describe

the science of applying mathematical algorithms – often implemented using computer software – to extract useful information from large or irregular chemical datasets. For example, the phenolic hydroxyl groups in milled wood lignins were determined using FTIR and chemometrics (i.e., Partial Least Squares (PLS) and Principal Components Regression (PCR)) that were calibrated using aminolysis as the reference technique.[119] Likewise, Tsuchikawa[120] reviewed the application of near infrared (NIR) spectroscopy to wood and paper, which included a discussion on the use of chemometrics to extract information from infrared spectra. In essence, Tsuchikawa [120] demonstrated that spectroscopy and chemometrics have applications in the pulp and paper industry where rapid monitoring and analysis can improve process economics.

The application of multivariate statistics in industry has improved process control, product diversity and process economics.[121-123] The proliferation of spectroscopic instruments in industry has been aided by the application of multivariate statistical approaches to large spectral datasets. For example, Principal Component Analysis (PCA) can be used to segregate populations of samples based solely on spectral data, allowing for high throughput compositional screening of various biomass populations.[124,125] Likewise, PCR and PLS regression can be used to rapidly and accurately determine the composition of lignocellulosic biomass using a well-calibrated mass, absorption, or vibrational spectroscopy models.[65,126,127] PLS is often referred to as Partial Least Squares (PLS) regression, although this definition does not fully capture the essential motive behind PLS, that is, to model the *hidden structure* of two datasets *via projections* onto a reduced or *latent* variable space.[128] A full mathematical development of the PLS-regression algorithm is available elsewhere[128].

PLS regression produces a predictive model while PCA simply highlights the differences within a given dataset. As a result, PCA is often used to simply discriminate between groups of samples based on spectral information; however, it can be adapted to produce a regression model.

To improve predictability, multivariate calibration models (e.g., PLS) often involve a data-preprocessing step whereby the data is transformed *via* one or more methods. Some of the more common preprocessing techniques will be discussed below.

2.3.2 Data Pre-processing

2.3.2.1 Mean Centering and Unit Variance Scaling

Mean centering and unit variance scaling are ubiquitous with multivariate calibration methods and are often referred to as auto-scaling. Mean centering a sample's IR, NMR, or mass spectrum can be accomplished by subtracting the mean intensity value from the intensity at each independent variable. This particular sample-based mean centering procedure is outlined mathematically in Equation 2.1 where *k* represents the total number of infrared spectral variables used in the model.

$$x_{i,k}^{MC,row} = x_{i,k} - \frac{1}{K}\left(\sum_{k=1}^{K} x_{i,k}\right)$$
$$\text{where } x_{i,k} \equiv \text{absorbance value of } i\text{th sample and } k\text{th spectral variable (wavenumber)} \tag{2.1}$$

Likewise, the spectra of multiple samples can be mean centred based on their individual variables (i.e. wavenumbers) as described in Equation 2.2.

$$x_{i,k}^{MC,column} = x_{i,k} - \frac{1}{i}\left(\sum_{i=1}^{i} x_{i,k}\right)$$
$$\text{where } x_{i,k} \equiv \text{absorbance value of } i\text{th sample and } k\text{th spectral variable (wavenumber)} \tag{2.2}$$

In summary, Equation 2.1 operates on **X**'s rows while Equation 2.2 operates on **X**'s columns.

Graphically, mean centering the dataset will center the data about the origin, thus eliminating the intercept term in the regression model.[129] In this way, the mean centered model becomes easier to interpret since all of the samples in the dataset are centered at the origin. Extra caution must be exercised for regression models where the mean will drift over time, thereby introducing bias in the dataset, a topic covered in greater detail elsewhere[130]. Unit variance scaling can also be applied to the rows (samples) or columns (variables) of a dataset. The technique is analogous to mean centering: each data point is divided by either the standard deviation of the sample's spectrum or the standard deviation of a single variable for all samples. Unit variance scaling ultimately eliminates scaling effects; for example, by reducing the scale of very large variables and increasing the scale of very small variables. Unit variance scaling is important since multivariate techniques like PCA and PLS are very sensitive to scaling.

2.3.2.2 Standard Normal Variate (SNV)

Row-operating unit variance scaling is often employed in conjunction with mean scaling in what is referred to as standard normal variate (SNV) scaling. Applying SNV scaling to spectra reduces particle size (multiplicative) effects that are especially troublesome for solid mixtures.[131] During SNV scaling, each sample spectrum is first row centred about an average intensity as shown in Equation 2.1. The standard deviation of the sample spectrum is then calculated and the mean centered data is then divided by this value. Equation 2.3 describes the SNV procedure in mathematical terms.

$$x_{i,k}^{SNV} = \frac{x_{i,k} - \frac{1}{K}\sum_{k=1}^{K} x_{i,k}}{s}$$

where $x_{i,k} \equiv i$th sample comprising k spectral variables (wavenumbers)

$s \equiv$ calculated standard deviation of all k variables of the ith spectrum (2.3)

Spectra treated using SNV have a mean of zero and a variance equal to unity. Detrending is sometimes used in conjunction with SNV-treated data since the latter are still susceptible to baseline curvature issues, which can be corrected by the former.

2.3.2.3 Smoothing

Savitzky-Golay[132] smoothing is an averaging algorithm that fits a piecewise polynomial to the dataset and then predicts the point of interest (e.g., spectral values) from the resulting polynomial equation. Savitzky-Golay smoothing is often applied in conjunction with first and second derivatives.[132]

2.3.2.4 First Derivative

Taking the first derivative of a spectrum gives the slope of the curve at each data point. Taking the first derivative of a spectrum can reduce baseline offset and linear baseline issues; however, it has been noted to cause a shift in the characteristic peaks, thus making the resulting spectrum difficult to interpret using peak-picking techniques. Nkansah et al. preprocessed their raw NIR spectra using Savitzky-Golay first derivatives (i.e., 15 smoothing points, second degree polynomial fit, followed by differentiation) in order to improve the quality, interpretation, and predictive ability of the PLS model.[124]

2.3.2.5 Second Derivative

Taking the second derivative of a spectrum gives the change in the slope of the curve at each data point. Similar to first derivative, second derivative spectra remove baseline offset and linear baseline issues. In addition, second derivative spectra retain the position of peaks so interpretation is generally easier than first derivative spectra. It should be noted that both first and second derivatives likely intensify noise in the model. In addition to first derivative preprocessing, Nkansah et al. also preprocessed their raw NIR spectra using second derivatives although no significant improvement to the predictive ability of the PLS regression model was observed.[124] Krongtaew et al. applied Savitzky-Golay second derivatives (i.e., 17 smoothing points, second degree polynomial fit, followed by differentiation) to their raw NIR spectra, thus improving the predictive ability of their PLS regression models.[133]

2.3.2.6 Multiplicative Signal Correction

Multiplicative Signal/Scatter Correction (MSC) is a method that can be used to reduce multiplicative and additive scatter effects in a dataset, such as those caused by path length variations, offset shifts, interferences, and particle size effects.[131] Mathematically, MSC is a transformation that first involves regressing the sample spectrum against an "ideal" spectrum, often estimated as the set-mean-spectra, $\bar{\mathbf{x}}(k)$, to fit parameters that describe additive and multiplicative effects.[134] The additive and multiplicative effects are estimated from the ordinary least squares solution to Equation 2.4.

$$\mathbf{x}_i = a_i\mathbf{1} + b_i\bar{\mathbf{x}} + e_i$$

where $\mathbf{x}_i \equiv$ vector of spectral absorbances of the *i*th sample

$a_i \equiv$ regression coefficient describing additive effects

$\mathbf{1} \equiv$ row vector of 1's introduced for formality

$b_i \equiv$ regression slope describing multiplicative effects

$\bar{\mathbf{x}} \equiv$ set-mean-spectra for all i samples, $\frac{1}{i}\sum_{i=1}^{i} x_{i,k}$

$e_i \equiv$ regression model error

(2.4)

Once the intercept (a_i) and slope (b_i) parameters have been determined, each sample spectrum is corrected as shown in Equation 2.5.

$$\mathbf{x}_i^{MSC} = \frac{\left(\mathbf{x}_i - \hat{a}_i\right)}{\hat{b}_i}$$

where $\mathbf{x}_i^{MSC} \equiv$ MSC-corrected vector of spectral absorbances of the *i*th sample

$\hat{a}_i \equiv$ estimated regression coefficient describing additive effects

$\hat{b}_i \equiv$ estimated regression slope describing multiplicative effects

(2.5)

The similarities between MSC and SNV can be noted by considering Equations 2.3 and 2.5 as well as the discussion presented by Dhanoa et al.[131]. A more detailed mathematical description of the MSC and Extended MSC (EMSC) techniques are described elsewhere[134-136]. Wallbäcks et al. found that MSC preprocessing of their raw NIR spectra slightly improved the predictability of their PLS regression models.[137] However, Nkansah et al. also preprocessed their raw NIR spectra using MSC although no significant improvement to the predictive ability of the PLS regression

model was observed.[124] Liu et al. applied EMSC to their raw NIR spectra in order to remove physical interferences and to subsequently improve the predictive ability of their PLS regression model.[138]

2.3.2.7 Orthogonal Signal Correction

Orthogonal Signal Correction (OSC) is a data preprocessing technique that removes information from **X** that is unrelated (orthogonal) to **Y**. OSC is analogous to the PLS algorithm (NIPALS) except that OSC attempts to minimize instead of maximize the covariance between **X** and **Y**.[139] Thus, each OSC component will remove information from **X** that shares minimal covariance (maximal orthogonality) with **Y**. Caution should be exercised when choosing the number of OSC components followed by PLS: if too many OSC components are used, the result will approach the multiple linear regression (MLR) solution and overfitting may take place.[139] A full mathematical description of the OSC process is outlined elsewhere[139].

2.3.2.8 Baseline Correction

Baseline correction seeks to minimize linear offset in the spectral data by subtracting either a minimum or user-defined value from the dataset. Similarly, constructing a line between two user-defined values can impose a linear baseline. For spectral data, a linear baseline is generally within a fingerprint region and each subsequent wavenumber in the spectrum is normalized accordingly. For example, the general form of baseline correction is to subtract a baseline absorbance value, which is a function of the wavenumber, from the corresponding absorbance value at the same wavenumber for the spectrum to be corrected, as described by Equation 2.6.

$$x^*_{i,k} = x_{i,k} - f(k)$$

where $x^*_{i,k} \equiv$ baseline corrected absorbance at wavenumber k of the ith sample

$x_{i,k} \equiv$ absorbance at wavenumber k of the ith sample

$f(k) \equiv$ absorbance value corresponding to some function of the wavenumber, k, for sample i (e.g. min(k) for traditional baseline offset correction or a linear function of the form mk+ b) (2.6)

2.3.2.9 Detrending

Detrending is a spectral correction technique, similar to and often used in conjunction with SNV, that attempts to reduce nonlinear trends by fitting a polynomial equation to each spectra that is later used for baseline correction. The detrended spectra are calculated as the difference between the original spectrum and the polynomial equation describing the new baseline. Other data pretreatment or preprocessing techniques are also available, such as various forms of normalization (e.g., minimum/maximum spectral normalization[133]), but for the sake of brevity they will not be discussed here. Following the acquisition, preprocessing and model building exercises, a technique known as Cross-Validation (CV) is implemented to optimize the predictive capability of the multivariate model.

2.3.3 Cross Validation

CV is often used to determine the optimum number of principal components to include in a multivariate model. There is currently no consensus on when to apply a particular CV procedure. The default CV procedure in ProSensus MultiVariate is to split the calibration data into 7 segments or groups with k samples per segment (i.e., the data are divided such that approximately 14.28% of the data is in each group). The CV procedure then leaves out one of the

7 segments (*k* samples) for validation of the regression model, which is built using the remaining 6 segments (i.e., 85.72% of the data calibrates the model). Other software such as CAMO Unscrambler® X allows for the specification of the number of segments and samples. Hames et al.[140] used a full CV procedure, a leave-one-out technique (i.e., each sample is its own segment or group), which is not recommended as this procedure tends to perform well for calibration but poorly for prediction.[141,142] As the number of samples increases, the calibration model appears to perform better under the leave-one-out CV procedure because there are increasingly more samples used for calibration but the same number of samples (i.e., one) used for validation. Thus, the leave-one-out method often results in overfitting of the data and an underestimation of true predictive error.[142] Alternatively, a larger number of samples to be left out for CV, n_v, could be selected such that $n_v / n \to 1$ as $n \to \infty$.[141] However, the leave-n_v-out procedure can be computationally taxing in its search for an optimal number of components with such a limited number of samples from which to build the calibration model. An alternative method called Monte Carlo CV was developed and was described by Xu and Liang[142]; however, this particular CV technique has not been extensively used for studies involving lignocellulose. In fact, a recent study by Krongtaew et al.[133] used FT-NIR (wet-chemistry as reference), PLS and leave-one-out CV to predict and monitor the chemical pretreatment and delignification of wheat and oat straw. The leave-one-out technique was selected due to the small number of samples used for the calibration set. The authors were able to extract quantitative information regarding total residual lignin content, enzymatically released reducing sugars, total solids, volatile solids, and biogas yield using the FT-NIR PLS regression model.[133]

As a rule of thumb, it is generally advisable to leave approximately 10% of the data out of the predictive model for CV purposes. Clearly, there is an inherent variability in multivariate model building since so many data-preprocessing techniques and CV protocols currently exist,

each with its own strengths and weaknesses. Thus, it is imperative that future studies involving lignocellulose and chemometrics thoroughly describe and support the methods used in the development of the model in addition to reporting the Root Mean Square Error of Estimation/Calibration (RMSEE) and Root Mean Square Error of Prediction (RMSEP) – the latter being the best indicator of predictive performance. Figure 2.5 provides a summary of the approach for producing a PCA or PLS compositional analysis using spectroscopy.

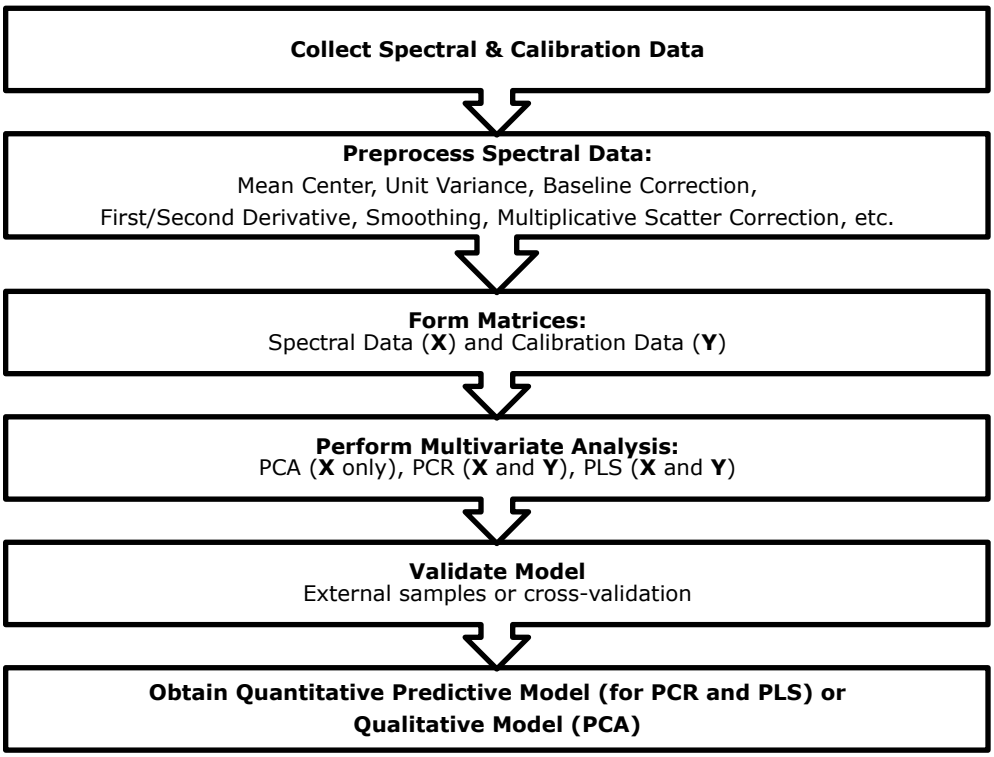


Figure 2.5. General framework for the development of a multivariate model.

A valuable mini-review was presented by Kalivas[143] on the development of multivariate calibration models and the types of problems one might encounter.

2.3.4 Calibration Model Development

Similar to univariate techniques, the performance of multivariate regression models is largely determined by the quality of the calibration samples. Thus, multivariate calibration should be implemented using ASTM E1655-05[144], as well as ASTM E2056-04[145] if surrogate mixtures are used to build the calibration set. If the multivariate model involves three variables or less in **Y**, the response matrix, then there should be a minimum of 24 sample spectra in **X** used for calibration purposes (following outlier removal).[144] Alternatively, if the model is developed using more than three variables ($k > 3$) in **Y,** then **X** should contain at least $6k$ spectra following outlier removal and $6(k + 1)$ spectra if the dataset is mean centered.[144] Surrogate mixtures can be used to model a small subset of the chemical constituents in a material, for example, cellulose, hemicellulose, and lignin for lignocellulosic biomass. The surrogate mixture method vastly simplifies experimental designs by reducing the number of samples required for model calibration. The use of surrogate mixtures for calibrating a PLS model based on FT-NIR spectra of pharmaceutical tablets was studied by Cournoyer et al.[146] and showed that an accurate predictive model could be constructed for tablets containing various concentrations of acetylsalicylic acid, caffeine, microcrystalline cellulose, sodium lauryl sulfate and starch, thus affording a new avenue for PAT in the production process. Despite the success of surrogate mixtures for calibrating chemometric models, the use of these techniques for lignocellulose biomass applications has been limited to date. Thus, NREL wet chemical methods have been used for calibration, whereby spectral data is regressed against wet chemical data to form a predictive regression model.[140] Alternatively, the detergent fibre methods of Van Soest[55] and Van Soest and Wine[56] have been recently used by Chen et al.[65] to calibrate a PLS model of FTIR spectra on wood samples. Although the authors were able to develop an adequate calibration model for the prediction of cellulose, hemicellulose and lignin, the incorporation of nine principal

components in the final model could be considered to be overfitting the model. It is generally advisable that a minimum number of principal components be selected to maximize the modeled trend (and minimize modeled error).

2.3.5 Recent Applications of Chemometrics in Lignocellulosic Biomass Research

Recently, Liu et al.[138] developed a broad FT-NIR model for the compositional analysis of three different lignocellulosic feedstocks: corn stover, switchgrass and wheat straw. Glucan, xylan, galactan, arabinan, mannan, lignin, and ash were selected as the analytes of interest and a calibration model was developed using PLS regression with EMSC data pretreatment to remove physical interferences in the spectra such as particle size effects. The experimental procedure employed by Liu et al.[138] is illustrated in Figure 2.6.

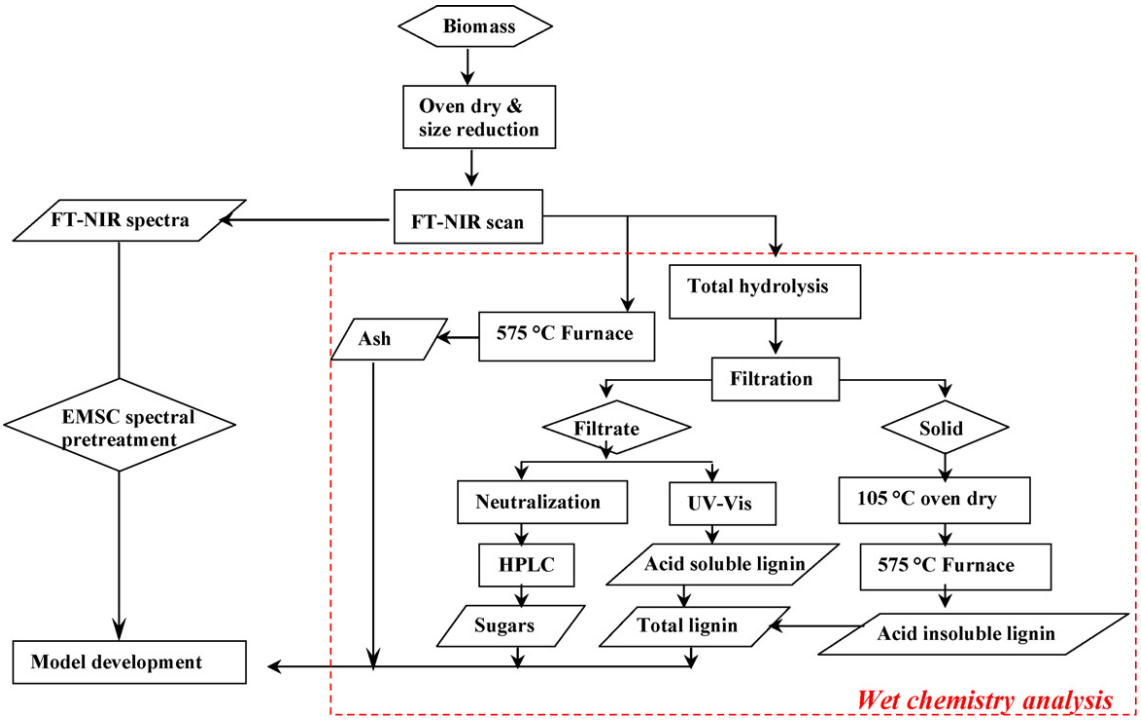


Figure 2.6. Experimental procedures used for the PLS model developed using FT-NIR spectra of corn stover, switchgrass, and wheat straw (Reproduced with permission from Liu et al.[138]).

Samples of corn stover and switchgrass were subjected to conventional wet chemical fractionation and analysis as well as FT-NIR spectroscopy. The calibration set comprised the wet chemical data while the prediction set comprised the FT-NIR spectra. Wheat straw was used to verify the robustness of the model that was developed using two dissimilar plant species: corn stover and switchgrass. Independent samples of corn stover and switchgrass that were not used in the calibration model were also employed as validation standards. In addition, leave-one-out cross validation was implemented. The model could adequately predict the three major components in each of the samples, glucan, xylan, and lignin with relative errors of 1.99%, 2.37%, and 3.62% respectively. Ash and the other minor sugars exhibited much higher relative predictive errors, which was attributed to the inability of inorganic compounds in ash to absorb in the NIR range, as well as the relatively low concentrations of the minor sugars present within the samples. Despite this, the authors demonstrated that a robust PLS model could be generated from FT-NIR data that could apply to samples with dissimilar compositions to those within the calibration set.

In an early study by Wallbäcks et al.[137], the changes in the chemical composition of pulp samples were analyzed over the duration of the birch Kraft pulping process using ^{13}C Cross Polarization Magic Angle Spinning (CP/MAS) NMR, FTIR, and NIR coupled with PLS regression. In addition to producing an accurate regression model capable of predicting glucose, xylose, and Klason lignin content, the authors showed that the combination of spectroscopic datasets from different instruments can sometimes yield models of greater predictive capability compared to a model built with only one dataset. The authors also suggested that NIR, which was demonstrated to have the best predictive capability of the three techniques, should be implemented for on-line measurement of samples using robust fibre optic probes.

More recently, Brink et al.[147] developed an on-line apparatus for the prediction of aspen and birch bark contents in unbleached hardwood pulp using NIR spectroscopy and multivariate techniques (PCA and PLS). The authors produced a semi-quantitative model that could be used to monitor the effect of hardwood pulp composition on final product quality at a paperboard mill. Temperature and moisture sensitivities were cited as the two most significant hindrances to the development of the model. As such, a meticulous preparation of the reference system was suggested for future work.

Pyrolysis molecular beam mass spectrometry (pyMBMS) was studied Tuskan et al.[126] as a rapid method for the compositional analysis of lignocellulosic biomass. The authors showed that PLS regression could be applied to pyMBMS data to obtain accurate and rapid estimates of specific gravity, extractives, galactan, mannan, xylan, alpha cellulose, and lignin contents in loblolly pine. Arabinan concentrations could not be quantified due to the low concentrations present in loblolly pine as well as difficulties in estimating arabinan using standard wet-chemical practices.[126] Full CV was applied to the pyMBMS PLS regression model to determine the optimal number of components. Data preprocessing was not mentioned.

The conversion of lignocellulosic biomass into high value or commodity products can be considered a multivariate process. For example, the enzymatic hydrolysis and fermentation of lignocellulosic biomass depends on variables such as the chemical composition of the substrate before pretreatment, degree cellulose crystallinity, macromolecular structure characteristics, pretreatment conditions, hydrolysis conditions, production of inhibitors, enzyme activities, and fermentation conditions.[148] Further, variation in the type or maturity of the lignocellulosic substrate can affect optimal pretreatment conditions.[61] Because the use of chemometrics and infrared spectroscopy, namely FT-NIR with fibre optic probes or FTIR with Attenuated Total Reflectance (ATR) accessories, have been shown to afford accurate, *in situ,* inexpensive, and

real-time analysis of polymer systems[149,150], it is easy to see the promise in applying these techniques in future lignocellulosic biorefineries. For more robust PAT applications, NIR spectroscopy might be considered more advantageous than FTIR, Raman, or NMR due to the high availability and industrial use of inexpensive NIR hardware. In addition, the near infrared region is generally less sensitive to water and thus more robust for samples with high water content. However, the added robustness of NIR devices comes at a cost to ability to resolve complex mixtures, in which case FTIR, Raman, and especially NMR will provide much more detailed chemical spectra.

2.4 Summary and Recommendations

Table 2.3 summarizes the use of various chemometric techniques for the compositional analysis of lignocellulosic biomass.

Table 2.3. Analytical method, sample type, and parameters used in studies involving spectroscopy and multivariate techniques.

Method or Technique	Sample Type	Description/Parameters	Reference
FT-NIR and PCA	Wood	First, second derivative, MSC	124
pyMBMS and PLS regression	Hybrid poplar and loblolly pine	PLS applied to pyMBMS data; full CV applied to pyMBMS PLS model	126
FTIR and PCR/PLS regression	Milled wood lignins	Phenolic hydroxyl group content	119
FT-NIR/FTIR and PLS regression	Corn stover	Full CV	140
FT-NIR and PLS regression	Swedish pine, Swedish spruce, Polish pine	MSC applied to NIR spectra	109
FTIR ATR and PCA	Wood, fibres, and particles of Beech and Grand Fir	First derivative applied to FTIR ATR spectra	111
FTIR comparison study	30 wood species	Two spectral zones analyzed: 1800–1100 cm^{-1} and 900–800 cm^{-1}	151

FTIR ATR, NIR and PLS regression	Various wood species	Baseline corrected, second derivative infrared spectra	152
CP/MAS NMR, FTIR DRIFT, FT-NIR with PLS regression	Birch Kraft pulp	60-115 ppm region used for NMR , 2000-715 cm^{-1} region used for FTIR DRIFT, MSC applied to both FTIR and NIR spectra	137
On-line FT-NIR and PCA and PLS regression	Aspen and birch bark	First derivative preprocessing used on NIR spectra; CV with 25% of data left out	147
FT-NIR and PLS regression	Corn stover, switchgrass, and wheat straw	EMSC applied to FT-NIR spectra; full CV used in addition to external validation standards	138
FT-NIR and PLS regression	Wheat and oat straw	Second derivative, second order polynomial fit, or minimum/maximum absorbance normalization applied to FT-NIR spectra; full CV	133
VIS spectra with PCR and PLS regression	Unbleached Kraft pulps of Scots pine, Norway spruce, and Silver birch	VIS data was mean centered prior to multivariate analysis; CV applied to model as per Wold, 1978	127

Applying chemometrics to spectroscopy can be troublesome, despite the many advantages over traditional chemical routes. As a result, future work should address and improve the following:

1. Careful selection and reporting of data preprocessing, modeling, and validation parameters, as these can greatly influence the efficacy and utility of the multivariate technique
2. The accuracy of the chemometric method, which is only as good as its reference data, is often determined using conventional, laborious wet-chemical routes, therefore surrogate mixtures should be investigated for simplified model building exercises
3. Large datasets that span a wide concentration or property range are required to produce an accurate, predictive multivariate model; perhaps further development of broad-based databases could aid in experimental design and verification. A data-driven modeling approach, such as

the Wiki-based management of chemometric projects described by Alsberg and Clare[102], appears to offer both an avenue towards collaboration and standardization in the compositional analysis of lignocellulose.

2.5 Conclusions and Future Outlook

This historical perspective on the compositional analysis of lignocellulosic biomass was presented in approximate chronological order with respect to the development of each method or technique. The compositional analysis of lignocellulosic biomass has evolved over the past 150 years, ranging from laborious wet-chemical fractionation techniques originally aimed at relating chemical composition with feedstock digestibility to more recent non-invasive and non-destructive spectroscopic techniques involving the emerging fields of chemometrics and PAT.

It is possible that the limitations of the previously discussed standard methods can be attributed to the large number of variables involved in the complicated fractionation process. Slight variations in season, plant species, and spatial location within the plant as well as fractionation conditions such as temperature, hydrolysis time, acid type, acid concentration, and even HPLC configuration can lead to irreproducible and contradictory results when attempting to compare compositional data across interdisciplinary research groups. Laborious and expensive techniques that result in compositional data that are difficult to compare with other research groups can be considered "standard" only insofar as a single, standardized laboratory does all of the analytical work. When the compositional analysis of a single sample can cost anywhere between \$800-\$2000[140] and can take days or weeks to complete, the cost associated with the elaboration of a thorough compositional analysis on a representative number of biomass samples can become laborious and expensive and is prohibitive to a profitable biorefinery.

The rapid prediction of biomass quality and composition can substantially reduce the cost of refining lignocellulosic materials. Furthermore, because infrared and Raman instruments are

relatively inexpensive and robust, they can be implemented as PAT in the biorefinery. It is likely that these robust methods will soon become commonplace for the analysis of lignocellulosic material in research and industrial settings.

2.6 Acknowledgements

The authors would like to acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC) for financial support through the Discovery Grant Program (Champagne, Cunningham). The authors would also like to acknowledge the Ministry of Research and Innovation for their support through the Early Researcher Award Program (Champagne) and Queen's University for their support by providing a Queen's Graduate Award and facilitating a G.E. Ted Courtnage Graduate Award in Engineering (Krasznai). The authors would like to thank the Courtnage family for establishing this award.

2.7 References

(1) Giger-Reverdin, S. *Animal Feed Science and Technology* **1995**, *55*, 295.
(2) Scheller, H. V.; Ulvskov, P. *Annual Review of Plant Biology* **2010**, *61*, 263.
(3) Vanholme, R.; Demedts, B.; Morreel, K.; Ralph, J.; Boerjan, W. *Plant Physiol.* **2010**, *153*, 895.
(4) Ahmad, M.; Taylor, C. R.; Pink, D.; Burton, K.; Eastwood, D.; Bending, G. D.; Bugg, T. D. H. *Molecular BioSystems* **2010**, *6*, 815.
(5) Barton Ii, F. E. *Animal Feed Science and Technology* **1988**, *21*, 279.
(6) Rubin, E. M. *Nature* **2008**, *454*, 841.
(7) FitzPatrick, M.; Champagne, P.; Cunningham, M. F.; Whitney, R. A. *Bioresource Technology* **2010**, *101*, 8915.
(8) Foyle, T.; Jennings, L.; Mulcahy, P. *Bioresource technology* **2007**, *98*, 3026.
(9) Sluiter, J. B.; Ruiz, R. O.; Scarlata, C. J.; Sluiter, A. D.; Templeton, D. W. *Journal of Agricultural and Food Chemistry* **2010**, *58*, 9043.
(10) Henneberg, W.; Stohmann, E. *J. Landwirtsch* **1859**, *3*, 485.
(11) Van Soest, P. J. *Nutritional Ecology of the Ruminant*; 2nd ed.; Comstock Publishing: Ithaca, NY, **1994**.
(12) Gorham, J. *New England Journal Medicine and Surgery* **1820**, *4*, 320.
(13) Horsford, E. N. *Philosophical Magazine Series 3* **1846**, *29*, 365.
(14) AOAC In *Official Methods of Analysis of AOAC International 13th Ed.*; AOAC International: Washington, DC, **1980**.
(15) Jung, H. J. G. *Journal of Nutrition* **1997**, *127*, 810S.

(16) Ralph, J. In *Proceedings of the 6th Brazilian Symposium Chemistry of Lignins and Other Wood Components* Guaratingueta, Brazil, **1999**, p 97.
(17) Rzedzicki, Z.; Sykut-Domanska, E.; Popielewicz, J. *Polish Journal of Food and Nutrition Sciences* **2008**, *58*, 307.
(18) Marton, J.; Adler, E. In *Lignin Structure and reactions*; Marton, J., Ed.; American Chemical Society Washington, DC: **1966**; Vol. 59, p vii.
(19) Klason, P. In *The Chemistry of Lignin*; Brauns, F. E., Ed.; Academic Press: New York, **1952**, p 151.
(20) Browning, B. L. *Methods of Wood Chemistry*; Wiley-Interscience: New York, **1967**.
(21) TAPPI In *Acid-Insoluble Lignin in Wood and Pulp*; Technical Association of the Pulp and Paper Industry: Atlanta, GA; Vol. T222 om-88.
(22) Sjöström, E.; Alén, R. *Analytical methods in wood chemistry, pulping, and papermaking*; Springer Verlag, **1999**.
(23) TAPPI In *Acid-Insoluble Lignin in Wood and Pulp*; Technical Association of the Pulp and Paper Industry: Atlanta, GA; Vol. T222 om-06.
(24) TAPPI In *Acid-Soluble Lignin in Wood and Pulp*; Technical Association of the Pulp and Paper Industry: Atlanta, GA; Vol. T250.
(25) Giger, S. *Ann.Zootech* **1985**, *34*, 85.
(26) Sun, R.; Song, X.; Sun, R.; Jian, J. *BioResources* **2011**, *6*, 317.
(27) Robertson, J. A.; I'Anson, K. J. A.; Treimo, J.; Faulds, C. B.; Brocklehurst, T. F.; Eijsink, V. G. H.; Waldron, K. W. *LWT - Food Science and Technology* **2010**, *43*, 890.
(28) Xu, Z.; Shen, G.; Lin, S. *Advanced Materials Research* **2010**, *171-172*, 261.
(29) Templeton, D. W.; Scarlata, C. J.; Sluiter, J.; Wolfrum, E. In *31st Symposium on Biotechnology for Fuels and Chemicals*; NREL: **2009**; Vol. San Francisco, CA.
(30) Saeman, J. F.; Bubl, J. L.; Harris, E. E.; Madison, Wis. : U.S. Dept. of Agriculture, Forest Service, Forest Products Laboratory: **1944**.
(31) Shaffer, P. A.; Somogyi, M. *Journal of Biological Chemistry* **1933**, *100*, 695.
(32) Saeman, J. F.; Bubl, J. L.; Harris, E. E. *Industrial & Engineering Chemistry Analytical Edition* **1945**, *17*, 35.
(33) Tsubaki, S.; Ozaki, Y.; Azuma, J. *Journal of Food Science* **2010**, *75*, C152.
(34) Hipsley, E. H. *International journal of food sciences and nutrition* **1953**, *7*, 168.
(35) DeVries, J. W.; Prosky, L.; Li, B.; Cho, S. *Cereal Foods World* **1999**, *44*, 367.
(36) Monro, J.; Burlingame, B. *Journal of Food Composition and Analysis* **1996**, *9*, 100.
(37) Saeman, J. F.; Moore, W. E.; Mitchell, R. L.; Millet, M. A. *Tappi J* **1954**, *37*, 336.
(38) Houghton, T. P.; Stevens, D. M.; Pryfogle, P. A.; Wright, C. T.; Radtke, C. W. *Applied Biochemistry and Biotechnology* **2009**, *153*, 4.
(39) Nelson, G. H.; Leming, J. A. *Tappi Journal* **1957**, *40*, 846.
(40) Wise, L. E.; Peterson, F. C.; Harlow, W. M. *Industrial & Engineering Chemistry Analytical Edition* **1939**, *11*, 18.
(41) Reid, J. D.; Nelson, G. H.; Aronovsky, S. I. *Industrial & Engineering Chemistry Analytical Edition* **1940**, *12*, 255.
(42) Norman, A. G.; Jenkins, S. H. *Biochemical Journal* **1933**, *27*, 818.
(43) Kürschner, K.; Hoffer, A. *Chem. Zeit.* **1931**, *55*, 161.
(44) Cunningham, R. L.; Detroy, R. W.; Carr, M. E. *Transactions of Illinois Academy of Sciences* **1982**, *75*, 283.
(45) Cunningham, R. L.; Detroy, R. W.; Bagby, M. O.; Baker, F. L. *Transactions of the Illinois Academy of Science* **1981**, *74*, 67.

(46) Cunningham, R. L.; Carr, M. E. *Biotechnology and Bioengineering Symposium* **1984**, *14*, 96.
(47) Cunningham, R. L.; Carlson, K. D.; Bagby, M. O. *Applied Biochemistry and Biotechnology* **1988**, *17*, 117.
(48) Claus, I.; Kordsachia, O.; Schröder, N.; Karstens, T. *Holzforschung* **2004**, *58*, 573.
(49) Banerjee, S.; Sen, R.; Pandey, R. A.; Chakrabarti, T.; Satpute, D.; Giri, B. S.; Mudliar, S. *Biomass and Bioenergy* **2009**, *33*, 1680.
(50) Tasman, J. E.; Berzins, V. *Tappi J* **1957**, *40*, 691.
(51) Tasman, J. E.; Berzins, V. *Tappi J* **1957**, *40*, 699.
(52) Hatfield, R.; Fukushima, R. S. *Crop Science* **2005**, *45*, 832.
(53) Fukushima, R. S.; Hatfield, R. D. *J.Agric.Food Chem* **2004**, *52*, 3713.
(54) Van Soest, P. J. *J. Assoc. Off. Anal. Chem.* **1963**, *46*, 825.
(55) Van Soest, P. J. *J. Assoc. Off. Anal. Chem.* **1963**, *46*, 829.
(56) Van Soest, P. J.; Wine, R. H. *Journal of the Association of Official Analytical Chemists* **1967**, *50*, 50.
(57) Goering, H. K.; Van Soest, P. J. *Agriculture Handbook* **1970**, *379*, 1.
(58) Van Soest, P. J. *Journal of the Association of Official Agricultural Chemists* **1965**, *49*, 546.
(59) Van Soest, P. J. In *Proceedings 23rd Southern Pasture Forage Crop Improvement Conference, Blacksburg, Virginia* 1966, p 24.
(60) Jung, H. G.; Mertens, D. R.; Payne, A. J. *Journal of dairy science* **1997**, *80*, 1622.
(61) Dien, B. S.; Jung, H. J. G.; Vogel, K. P.; Casler, M. D.; Lamb, J. A. F. S.; Iten, L.; Mitchell, R. B.; Sarath, G. *Biomass and Bioenergy* **2006**, *30*, 880.
(62) Cherney, D. J. R.; Patterson, J. A.; Cherney, J. H. *Journal of Dairy Science* **1989**, *72*, 3079.
(63) Mertens, D. R. *Journal of AOAC International* **2002**, *85*, 1217.
(64) Holechek, J. L.; Vavra, M. *Journal of Range Management* **1982**, *35*, 799.
(65) Chen, H.; Ferrari, C.; Angiuli, M.; Yao, J.; Raspi, C.; Bramanti, E. *Carbohydrate Polymers* **2010**, *82*, 772.
(66) Moore, W. E.; Johnson, D. B. In *Procedures for the Chemical Analysis of Wood and Wood Products (Determination of pentosans in pulp)*; Forest Products Laboratory Forest Service, US Department of Agriculture: **1967**.
(67) Moore, W. E.; Johnson, D. B. In *Procedures for the Chemical Analysis of Wood and Wood Products (Determination of wood sugars)*; Forest Products Laboratory Forest Service, US Department of Agriculture: **1967**.
(68) Fengel, D.; Wegener, G. In *Hydrolysis of Cellulose: Mechanisms of Enzymatic and Acid Catalysis Advances in Chemistry Series No. 181*; Brown, R. D., Jurasek, L., Eds.; American Chemical Society: **1979**, p 145.
(69) Morrison, I. M.; Stewart, D. *Phytochemistry* **1998**, *49*, 1555.
(70) Windeisen, E.; Strobel, C.; Wegener, G. *Wood Science and Technology* **2007**, *41*, 523.
(71) Grohmann, K.; Himmel, M.; Rivard, C.; Tucker, M.; Baker, J.; Torget, R.; Graboski, M. *Biotechnol. Bioeng. Symp.* **1984**, *14*, 137.
(72) Niu, K.; Chen, P.; Zhang, X.; Tan, W. S. *Journal of Chemical Technology & Biotechnology* **2009**, *84*, 1240.
(73) Prosky, L.; Asp, N. G.; Furda, I.; DeVries, J. W.; Schweizer, T. F.; Harland, B. F. *Journal- Association of official analytical chemists* **1984**, *67*, 1044.
(74) Prosky, L.; Asp, N. G.; Furda, I.; DeVries, J. W.; Schweizer, T. F.; Harland, B. F. *Journal - Association of Official Analytical Chemists* **1985**, *68*, 677.

(75) Theander, O.; Westerlund, E. A. *Journal of Agricultural and Food Chemistry* **1986**, *34*, 330.
(76) Theander, O.; Westerlund, E.; Andersson, R.; Aman, P. In *CRC handbook of dietary fibre in human nutrition*; Spiller, G. A., Ed.; CRC: **2001**; Vol. 3, p 87.
(77) Marlett, J. A. *Animal Feed Science and Technology* **1989**, *23*, 1.
(78) Theander, O.; Aman, P.; Westerlund, E.; Andersson, R.; Pettersson, D. *Journal of AOAC International* **1995**, *78*, 1030.
(79) Milne, T. A.; Chum, H. L.; Agblevor, F.; Johnson, D. K. *Biomass and Bioenergy* **1992**, *2*, 341.
(80) Del Río, J. C.; Marques, G.; Rencoret, J.; Martínez, Á. T.; Gutiérrez, A. *J.Agric.Food Chem* **2007**, *55*, 5461.
(81) Hatfield, R. D.; Grabber, J.; Ralph, J.; Brei, K. *J.Agric.Food Chem* **1999**, *47*, 628.
(82) Sluiter, A.; Hames, B.; Ruiz, R.; Scarlata, C.; Sluiter, J.; Templeton, D.; Crocker, D. *NREL, Golden, CO* **2004**.
(83) NREL 2010; Vol. 2010.
(84) ASTM; ASTM: West Conshohocken, PA, **2001**; Vol. E1758-01.
(85) Wolfrum, E. J.; Sluiter, A. D. *Cellulose* **2009**, *16*, 567.
(86) NREL 2006; Vol. 2010.
(87) Lee, S. H.; Doherty, T. V.; Linhardt, R. J.; Dordick, J. S. *Biotechnology and bioengineering* **2009**, *102*, 1368.
(88) Zhang, Y. H. P.; Ding, S. Y.; Mielenz, J. R.; Cui, J. B.; Elander, R. T.; Laser, M.; Himmel, M. E.; McMillan, J. R.; Lynd, L. R. *Biotechnology and bioengineering* **2007**, *97*, 214.
(89) Sannigrahi, P.; Ragauskas, A. J.; Miller, S. J. *Carbohydrate research* **2010**.
(90) Humbird, D.; Mohagheghi, A.; Dowe, N.; Schell, D. J. *Biotechnology progress* **2010**, *26*, 1245.
(91) Xu, J.; Cheng, J. J.; Sharma-Shivappa, R. R.; Burns, J. C. *Bioresource technology* **2010**, *101*, 2900.
(92) Shafiei, M.; Karimi, K.; Taherzadeh, M. J. *Bioresource technology* **2010**, *101*, 4914.
(93) Brijwani, K.; Oberoi, H. S.; Vadlani, P. V. *Process Biochemistry* **2009**, *45*, 120.
(94) Ververis, C.; Georghiou, K.; Danielidis, D.; Hatzinikolaou, D. G.; Santas, P.; Santas, R.; Corleti, V. *Bioresource Technology* **2007**, *98*, 296.
(95) Scarlata, C. J.; Hyman, D. A. *Journal of Chromatography A* **2010**.
(96) Templeton, D. W.; Scarlata, C. J.; Sluiter, J. B.; Wolfrum, E. J. *Journal of Agricultural and Food Chemistry* **2010**, 341.
(97) Remy, E. *Biochem.Z* **1931**, *236*.
(98) Williams, R. D.; Olmsted, W. H. *Journal of Biological Chemistry* **1935**, *108*, 653.
(99) Ferguson, W. S. *Biochemical Journal* **1942**, *36*, 786.
(100) Fales, F. W.; Russell, J. A.; Fain, J. N. *Clinical chemistry* **1961**, *7*, 289.
(101) Walford, S. N. In *Proc S Afr Sug Technol Ass* 2008; Vol. 81, p 266.
(102) Alsberg, B. K.; Clare, A. *Journal of Chemometrics* **2010**.
(103) Workman, J.; Koch, M.; Lavine, B.; Chrisman, R. *Analytical Chemistry* **2009**, *81*, 4623.
(104) Naumann, D. *Applied Spectroscopy Reviews* **2001**, *36*, 239.
(105) Mackanos, M. A.; Contag, C. H. *Trends in biotechnology* **2010**, *28*, 317.
(106) Teodor, E. S.; Teodor, E. D.; Virgolici, M.; Manea, M. M.; Truica, G.; Litescu, S. C. *Journal of Archaeological Science* **2010**, *10*, 2386.
(107) Yonenobu, H.; Tsuchikawa, S.; Oda, H. *Journal of Near Infrared Spectroscopy* **2003**, *11*, 407.

(108) Dayal, B. S.; MacGregor, J. F.; Taylor, P. A.; Kildaw, R.; Marcikic, S. *Pulp & Paper Canada* **1994**, *95*, 26.
(109) Antti, H.; Sjöström, M.; Wallbäcks, L. *Journal of Chemometrics* **1996**, *10*, 591.
(110) Benito, M. T. J.; Ojeda, C. B.; Rojas, F. S. *Applied Spectroscopy Reviews* **2008**, *43*, 452.
(111) Müller, G.; Schöpper, C.; Vos, H.; Kharazipour, A.; Polle, A. *BioResources* **2009**, *4*, 49.
(112) Durán, N.; Angelo, R. *Applied Spectroscopy Reviews* **1998**, *33*, 219.
(113) Schulz, H.; Baranska, M. *Vibrational Spectroscopy* **2007**, *43*, 13.
(114) Naumann, A.; Peddireddi, S.; Kües, U.; Polle, A. *Fourier transform microscopy in wood analysis*; Universitätsverlag, Göttingen: Göttingen, 2007.
(115) Kačuráková, M.; Wilson, R. H. *Carbohydrate Polymers* **2001**, *44*, 291.
(116) Adapa, P. K.; Karunakaran, C.; Tabil, L. G.; Schoenau, G. J. In *CSBE/SCGAB Annual Conference*; The Canadian Society of Bioengineering: Prince Edward Island, 2009.
(117) Gierlinger, N.; Schwanninger, M. *Plant Physiology* **2006**, *140*, 1246.
(118) Argyropoulos, D. S. *Research on Chemical Intermediates* **1995**, *21*, 373.
(119) Faix, O.; Böttcher, J. H. *Holzforchung* **1993**, *47*, 45.
(120) Tsuchikawa, S. *Applied Spectroscopy Reviews* **2007**, *42*, 43.
(121) Muteki, K.; MacGregor, J. F. *AICHE Journal* **2008**, *54*, 1554.
(122) Muteki, K.; MacGregor, J. F.; Ueda, T. *Ind.Eng.Chem.Res* **2006**, *45*, 4653.
(123) MacGregor, J. F.; Yu, H.; García Muñoz, S.; Flores-Cerrillo, J. *Computers & Chemical Engineering* **2005**, *29*, 1217.
(124) Nkansah, K.; Dawson-Andoh, B.; Slahor, J. *Bioresource technology* **2010**.
(125) Sellick, C. A.; Hansen, R.; Jarvis, R. M.; Maqsood, A. R.; Stephens, G. M.; Dickson, A. J.; Goodacre, R. *Biotechnology and bioengineering* **2010**, *106*, 432.
(126) Tuskan, G.; West, D.; Bradshaw, H. D.; Neale, D.; Sewell, M.; Wheeler, N.; Megraw, B.; Jech, K.; Wiselogel, A.; Evans, R. *Applied Biochemistry and Biotechnology* **1999**, *77*, 55.
(127) Malkavaara, P.; Alén, R. *Chemometrics and Intelligent Laboratory Systems* **1998**, *44*, 287.
(128) Wold, S.; Sjöström, M.; Eriksson, L. *Chemometrics and Intelligent Laboratory Systems* **2001**, *58*, 109.
(129) Boysworth, M. K.; Booksh, K. S. In *Handbook of near-infrared analysis*; Burns, D. A., Ciurczak, E. W., Eds.; CRC Press: Boca Raton, FL, **2008**, p 209.
(130) Nomikos, P.; MacGregor, J. F. *AICHE Journal* **2004**, *40*, 1361.
(131) Dhanoa, M. S.; Lister, S. J.; Sanderson, R.; Barnes, R. J. *Journal of Near Infrared Spectroscopy* **1994**, *2*, 43.
(132) Savitzky, A.; Golay, M. J. E. *Analytical Chemistry* **1964**, *36*, 1627.
(133) Krongtaew, C.; Messner, K.; Ters, T.; Fackler, K. *BioResources* **2010**, *5*, 2081.
(134) Helland, I. S.; Naes, T.; Isaksson, T. *Chemometrics and Intelligent Laboratory Systems* **1995**, *29*, 233.
(135) Martens, H.; Nielsen, J. P.; Engelsen, S. B. *Analytical Chemistry* **2003**, *75*, 394.
(136) Martens, H.; Stark, E. *Journal of Pharmaceutical and Biomedical Analysis* **1991**, *9*, 625.
(137) Wallbäcks, L.; Edlund, U.; Norden, B.; Berglund, I. *Tappi Journal* **1991**, *74*, 201.
(138) Liu, L.; Ye, X. P.; Womac, A. R.; Sokhansanj, S. *Carbohydrate Polymers* **2010**, *81*, 820.
(139) Wold, S.; Antti, H.; Lindgren, F.; Öhman, J. *Chemometrics and Intelligent Laboratory Systems* **1998**, *44*, 175.

(140) Hames, B. R.; Thomas, S. R.; Sluiter, A. D.; Roth, C. J.; Templeton, D. W. *Applied Biochemistry and Biotechnology* **2003**, *105*, 5.
(141) Shao, J. *Journal of the American Statistical Association* **1993**, *88*, 486.
(142) Xu, Q.-S.; Liang, Y.-Z. *Chemometrics and Intelligent Laboratory Systems* **2001**, *56*, 1.
(143) Kalivas, J. H. *Analytical Letters* **2005**, *38*, 2259.
(144) ASTM; ASTM: West Conshohocken, PA., **2005**; Vol. E1655-05.
(145) ASTM; ASTM: West Conshohocken, PA., **2004**; Vol. E2056-04.
(146) Cournoyer, A.; Simard, J. S.; Cartilier, L.; Abatzoglou, N. *Pharmaceutical development and technology* **2008**, *13*, 333.
(147) Brink, M.; Mandenius, C. F.; Skoglund, A. *Chemometrics and Intelligent Laboratory Systems* **2010**, *103*, 53.
(148) Vallander, L.; Eriksson, K. In *Bioprocesses and Applied Enzymology*; Fiechter, A., Ed.; Springer: **1990**; Vol. 42, p 63.
(149) Dubé, M. A.; Li, L. *Polymer-Plastics Technology and Engineering* **2010**, *49*, 648.
(150) Puskas, J. E.; Antony, P.; Kwon, Y.; Paulo, C.; Kovar, M.; Norton, P. R.; Kaszas, G.; Altstädt, V. *Macromolecular Materials and Engineering* **2001**, *286*, 565.
(151) Colom, X.; Carrillo, F. *Journal of Wood Chemistry and Technology* **2005**, *25*, 1.
(152) Stefke, B.; Windeisen, E.; Schwanninger, M.; Hinterstoisser, B. *Analytical Chemistry* **2008**, *80*, 1272.

Chapter 3

Quantitative Characterization of Lignocellulosic Biomass Using Multivariate Statistical Techniques

Significant portions of this chapter were accepted for publication (with revisions) in *Bioresource Technology* on January 17, 2011 (DOI: 10.1016/j.biortech.2012.01.089).

Abstract

In an effort to expedite the compositional analysis of lignocellulosic biomass (see Chapter 2), PLS regression models were developed using mixtures of cellulose, xylan, and lignin in a ternary mixture experimental design for multivariate model calibration. Mid-infrared spectra of these representative samples were recorded using Attenuated Total Reflectance (ATR) Fourier Transform Infrared (FT-IR) spectroscopy and regressed against their known composition using Partial Least Squares (PLS) multivariate techniques. The regression models were cross-validated and then used to predict the unknown compositions of two *Arabidopsis* cultivars, B10 and C10. The effect of various data preprocessing techniques on the final predictive ability of the PLS regression models was also evaluated.

The predicted compositions of B10 and C10 by the PLS regression model after second derivative data preprocessing were accurate compared to the results provided by a third-party analysis. This study suggests that mixture designs could be used as calibration standards in PLS regression for the compositional analysis of lignocellulosic materials if the infrared data is appropriately preprocessed.

3.1 Introduction

Determination of the relative quantities of cellulose, hemicellulose, and lignin in lignocellulosic biomass is often required in studies involving the fractionation[1,2,3,4], fermentation[5,6], or other modification[7,8,9] of this renewable plant-based material. However, despite its importance, the compositional analysis of lignocellulosic biomass remains a challenging and time-consuming task primarily due to its recalcitrant structure. Expensive fractionation and derivatization steps are often required, involving strong acids and exotic solvents that impose economic and environmental constraints when continually employed in research or industrial applications. In an effort to reduce costs and expedite the compositional analysis of lignocellulose, inexpensive and rapid techniques have been developed whereby infrared spectra of the raw biomass can be correlated, *via* multivariate techniques (i.e., a form of chemometrics), to compositional data gained from wet-chemical analyses.[10,11,12,13] Multivariate analysis, such as Projection to Latent Structures or Partial Least Squares (PLS) regression, as developed by Wold et al.[14], could also be utilized in tandem with other Process Analytical Technologies (PAT) where real-time compositional data could be potentially transmitted from upstream sensors to downstream unit operations, thus allowing for better control and efficiency in a biorefinery.[15]

Recent studies have used both near infrared (NIR) and mid infrared (MIR) data to calibrate predictive PLS regression models for the compositional analysis of lignocellulose.[10,16] Kadam et al.[17] compared NIR and MIR data of different crystallization systems and found that their particular MIR system with an Attenuated Total Reflectance (ATR) probe generally outperformed the NIR system. MIR showed decreased sensitivity due to fouling as well as improved peak selection due to the fundamental vibrational modes of molecules provided in the MIR electromagnetic region.

Traditional wet-chemical techniques have been typically used to calibrate multivariate models based on IR spectra of lignocellulosic biomass.[16] In our study, an alternative set of calibration samples was prepared by varying known amounts of cellulose, hemicellulose, and lignin in a series of powdered surrogate mixture samples covering a weight fraction range from 0 to 1 for each model compound, respectively. A PLS regression model was then constructed using FT-IR data of these surrogate mixtures as the training or calibration data set. FT-IR absorbance data of two genetically engineered *Arabidopsis* species was then used as the input variables in the regression model, which ultimately yielded compositional data for the two species in the form of cellulose, hemicellulose, and lignin weight fractions. We used a similar ATR FT-IR system to Kadam et al.[17] because it simplified the peak selection and sample preparation procedures. In addition, data preprocessing methods were applied to significantly improve the predictive capability of the multivariate model by reducing multiplicative, baseline, and scatter effects in the infrared data. To the authors' knowledge, this type of calibration approach has not yet been reported for the compositional analysis of lignocellulosic biomass. This type of calibration model involving carefully selected surrogate mixtures was selected because it could potentially reduce or eliminate the requirement for laborious wet-chemical analyses.

3.2 Materials and methods

3.2.1 Mixture design

A ternary mixture experimental design was developed using SAS JMP 8.0, as shown in Figure 3.1. The surrogate mixtures consisted of three model lignocellulosic compounds that are inexpensive and widely available: cellulose (Avicel PH-101, Sigma Aldrich), hemicellulose (Beechwood xylan, ≥ 90%, Sigma Aldrich), and lignin (Indulin AT, Kraft lignin, Meadwestvaco).

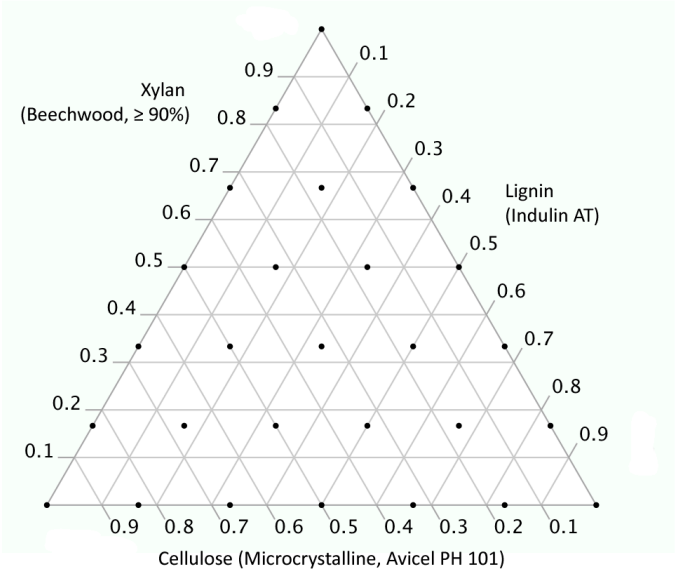


Figure 3.1. Ternary mixture plot of the model compounds used in the study. The surrogate mixtures comprised varying weight fractions of cellulose (Avicel PH-101), hemicellulose (Beachwood xylan), and lignin (Indulin AT, Kraft lignin). Mixtures are expressed as normalized weight fractions in the range of 0 to 1.

The surrogate mixtures shown in Figure 3.1 were intended to represent the entire concentration profile for a hypothetical sample of finely divided lignocellulosic biomass. Three components were added to a 20 mL glass screw-cap vial to make 1 g total of biomass from which aliquots were taken for ATR FT-IR spectroscopy. In addition to the materials used in the surrogate mixtures, two external samples were obtained from a company specializing in the genetic engineering of plant biomass, namely two samples of genetically engineered *Arabidopsis* species with unknown composition referenced as B10 and C10. Wet-chemical compositional analysis data for B10 and C10 were provided by a third-party after the generation of the predictive results (i.e., from PLS regression) described in this study. The following information was withheld from the authors during the development of the predictive models: B10 was a high yield line due to a constitutive promoter overexpressing a gene that regulates growth and yield. C10 was a standard control variety of the ecotype *Columbia*. With respect to propagation, seeds

were germinated on agar plates and transplanted at the two-leaf stage with three seedlings per 4 inch pot into standard soil mixture. Plants were grown at 22 °C with 18 hours of light at 200 microeinsteins intensity until senescence. The seeds were then removed and the remaining vegetative material was combined and processed through a Wiley mill equipped with a 40-mesh screen.

3.2.2 Attenuated Total Reflectance (ATR) FT-IR spectroscopy

Five replicate aliquots of the 28 calibration mixtures were individually analyzed using a Nicolet 6700 Fourier Transform Infrared (FT-IR) spectrometer with a diamond Attenuated Total Reflectance (ATR) accessory. Spectral data for the five replicates each comprised 64 co-added (i.e., each spectrum was added to the next) individual spectra at a resolution of 2 cm^{-1}. The co-added spectra for the five replicates were averaged (n = 5) to give 28 mean spectra, i.e., one for each surrogate mixture. Background spectra were collected between each replicate after the ATR crystal was cleaned with acetone to eliminate interferences caused by the cleaning agent or changes in the atmosphere. The calibration matrix was a 28 × 7024 matrix that comprised 28 averaged absorbance spectra for the surrogate calibration mixtures (rows), with wavenumbers ranging from 375 to 7399 cm^{-1} (columns). The calibration matrix was then augmented with the averaged absorbance spectra of five external validation mixtures with compositions within the experimental design and two unknown samples, making the final matrix dimension 35 × 7024. All spectral data was compiled into a single file prior to further processing using a MATLAB program. The final matrix was subjected to various data pretreatments in Unscrambler® X, including smoothing, first derivative, second derivative, baseline correction, standard normal variate (SNV), detrending, unit vector normalization, multiplicative signal correction (MSC), and orthogonal signal correction (OSC), as well as combinations of the aforementioned pretreatments.

3.2.3 Principal Component Analysis (PCA) and Partial Least Squares (PLS) regression

PCA and PLS regression were implemented on the pretreated spectral and loadings matrices using the ProSensus MultiVariate software package. The spectral intervals of 1800 - 400 and 4000 - 2500 cm^{-1} were used for PCA and PLS regression in preliminary studies. The spectra were then truncated to the less noisy and more representative range of 1800 - 800 cm^{-1}, as will be discussed in later sections.

3.2.4 Model validation

Cross-validation (CV) was used to determine the optimum number of principal components to use in the multivariate model. Since there is currently no consensus on when to apply a particular CV procedure, the procedure is largely determined by user preference (e.g., ASTM E 1655 does not stipulate a particular of CV procedure). Thus, the default CV procedure of ProSensus MultiVariate was employed whereby the calibration dataset is divided into 7 segments with 4 samples per segment (14.28% of data). The CV procedure then leaves out one of the 7 segments (4 samples) for validation of the regression model, which is built upon the remaining 6 segments (24 samples). In other studies, such as Hames et al.[13], a full CV procedure, a leave-one-out technique, which is not recommended according to Shao[18], as the leave-one-out (i.e. full cross-validation where the number of groups equals the number of samples) method can result in overfitting of the data and an underestimation of true predictive error.[19] As such, a more conservative "leave-four-out" technique was implemented in our study. Chemical fractionation followed by High Performance Liquid Chromatography (HPLC) on the resulting streams was used as an alternative method for the compositional analysis of lignocellulosic biomass and was performed by a third-party laboratory.

3.2.5 Model description

The multivariate techniques implemented in this study generally rely on reducing the amount of original data *via* matrix projections. In this way, the datasets are transformed into more meaningful or representative predictive models. Generally, one or two matrices are required for multivariate analysis of infrared spectra: an **X** matrix comprising spectral data and a **Y** matrix comprising the concentration profiles of the different samples under analysis. Let $\mathbf{X}_{(n\times k)}$ be a matrix consisting of n samples and k variables (wavenumbers). Let $\mathbf{Y}_{(n\times m)}$ be a matrix consisting of n samples and m variables (reference weight fractions). For all models, the number of samples (n) in both **X** and **Y** were equal to 35, comprising calibration (28), external validation (5), and prediction (2) samples. Three variables were included in **Y** for all models, corresponding to the known weight fractions of cellulose, hemicellulose, and lignin as depicted in Figure 3.2

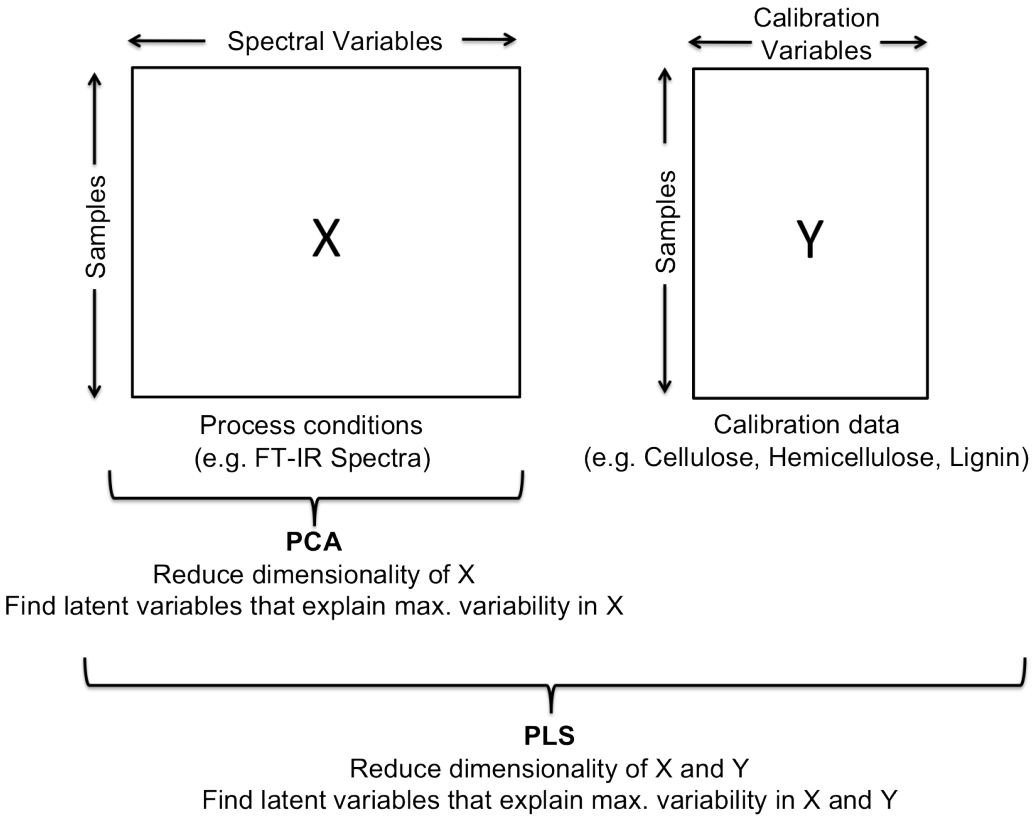


Figure 3.2. Illustration of the two main datasets used in this study. For PCA, only an X matrix is required while both X and Y are required for PLS regression.

In PCA, two matrices are formed, **T** and **P,** to approximate the original data matrix, **X**, according to $\mathbf{X} = \mathbf{TP}^T$, where **T** is the scores matrix and **P** is the loadings matrix ($\mathbf{P}^T$ is the matrix transpose of **P**). The matrix **X** is represented by **T** *via* a matrix projection, given by the new basis in **P**. Similarly, PLS regression involves matrix projections; however, this regression technique involves decomposing the **X** matrix using data in a new matrix, **Y**, such that the covariance between each latent variable in **X** and **Y** is maximized.[20] The resulting equations for PLS regression are $\mathbf{X} = \mathbf{TP}^T + \mathbf{E}$ and $\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F}$, where **T** and **U** are the scores matrices, **P** and **Q** are the loadings matrices, and **E** and **F** are the error matrices (residuals) of **X** and **Y**, respectively. Additional matrices involved in the PLS algorithm (i.e., NIPALS), such as the weight matrix **W**,

are required but not mentioned above since they are not necessary for the interpretation of the results in this chapter.

In this study, data preprocessing was applied to all samples in **X**, including external validation and prediction samples. All variables (columns) in **X** and **Y** were mean centred and unit variance scaled using the default procedure available in ProSensus MultiVariate. The dataset was truncated to variables in the range of 1800 to 800 cm^{-1} since cellulose, hemicellulose, and lignin are known to have strong infrared signals in this range, as summarized in Table 3.1.

Table 3.1. Important MIR wavenumbers corresponding to the model lignocellulosic components used in this study: cellulose, hemicellulose, and lignin.

Wavenumber (cm^{-1})	**Description**	**Reference**
1739	Wood xylan, C=O stretching in ketone, carbonyl and aliphatic groups	16
1640	Wood lignin, C=O stretching conjugated to aromatic ring	16
1595	Hardwood lignin, C=O stretching conjugated to aromatic ring	16
1510-1505	Lignin, aromatic skeletal stretching	16, 21
1463	Wood lignin and xylan, CH_2 deformation stretching	16
1429	Cellulose I, CH_2 scissoring motion	22
1425	Wood lignin, aromatic skeletal combined with C-H in-plane deformation and stretching	16
1420	Cellulose II, CH_2 scissoring motion	22
1321-1317	Wood lignin, condensation of guaiacyl and syringyl unit, syringyl unit and CH_2 bending and streching	16
1270	Lignin, guaiacyl ring breathing and C=O stretching	16
1240	Lignin, C_{aryl}-O	21
1163	Cellulose I and III C-O-C antisymmetrical bridge stretching mode	22
1156	Cellulose II and amorphous cellulose C-O-C antisymmetrical bridge stretching mode	23
1111	Cellulose I ring stretching or association of alcohol resulting in C-O stretching	22, 24
1102	Cellulose III ring stretching or association of alcohol resulting in C-O stretching	23
1090	Cellulose II and amorphous cellulose strong broad band, similar to 1111 and 1102 modes	23
897-893	Cellulose I – weak and broad centred at 897 cm^{-1} Cellulose II – strong and sharp at 893 cm^{-1} Cellulose III – less strong and located at 897 cm^{-1} as in I Amorphous cellulose – less strong and sharp than II at 897 cm^{-1}	24

The mid-infrared absorbance regions sensitive to water (4000 to 2600 cm^{-1}) and diamond interference (2400 to 1900 cm^{-1}) were discarded prior to performing multivariate analysis on the data. The representative FT-IR spectral region used for building the PLS regression models is shown graphically in Figure 3.3.

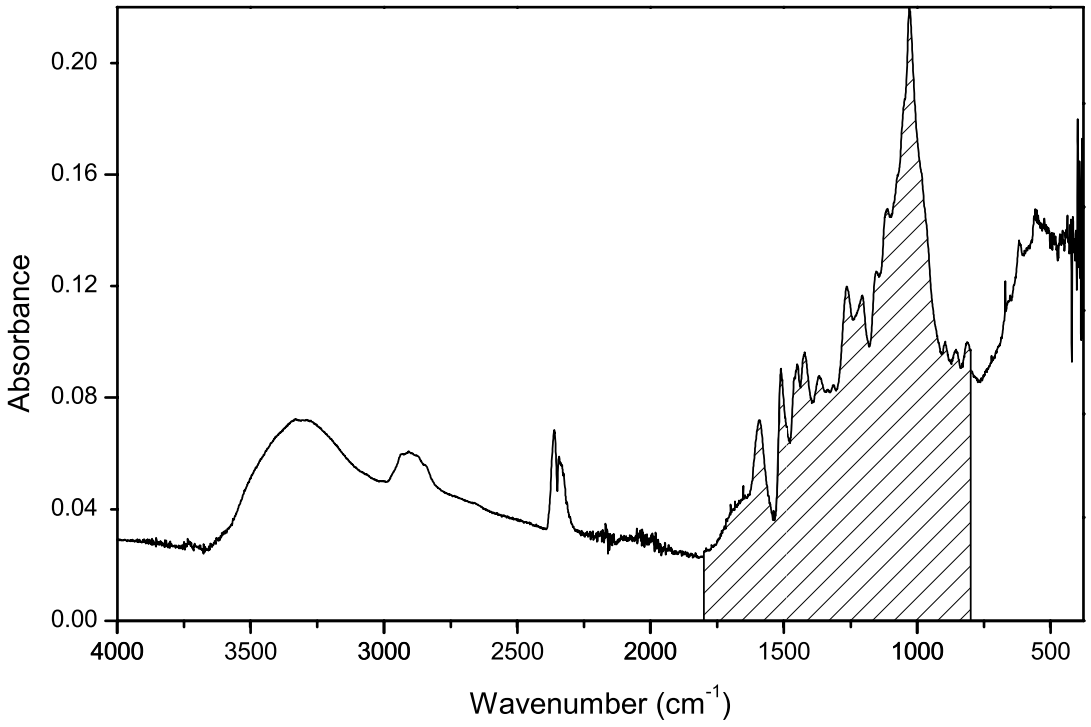


Figure 3.3. Typical FT-IR spectrum of a lignocellulosic biomass calibration sample prepared *via* the surrogate mixture design. No data preprocessing was applied to the above spectrum. Here, the shaded region shows those absorbances used for PLS regression models (i.e., absorbances corresponding to wavenumbers between 1800 and 800 cm^{-1}).

The models with the best predictive capability were those that used a truncated spectra in the lignocellulose "fingerprint" region (see Table 3.1), where only absorbances in the region 1800 to 800 cm^{-1} were used.

3.2.6 ATR FT-IR spectral data preprocessing

An investigation into various spectral data preprocessing techniques was conducted using CAMO Unscrambler® X. The data were then exported into ProSensus MultiVariate for PLS regression analysis.

3.2.6.1 Mean centering and unit variance scaling

Mean centering and unit variance normalizations on columns were used for every model developed in this study, as these techniques are commonly used with multivariate calibration methods. Mean centering can be accomplished by subtracting the mean of the sample's spectrum from each individual wavenumber of that sample. Likewise, the spectra of multiple samples can be mean centred based on their individual variables (i.e. wavenumbers). Graphically, mean centering the dataset tends to center the data about the origin, thus eliminating the intercept term in the regression model.[25] In this way, the mean centered model becomes easier to interpret since all of the samples in the dataset are centered at the origin. Extra caution must be exercised for regression models where the mean will drift over time, thereby introducing bias in the dataset, a topic covered in greater detail elsewhere[26]. For the purpose of this investigation a standard column-operating mean centering routine was implemented using ProSensus.

Unit variance scaling can also be applied to the rows (samples) or columns (variables) of a dataset. The technique is analogous to mean centering: each data point is divided by the standard deviation of the sample's spectrum or the standard deviation of a single variable for all samples. Unit variance scaling ultimately eliminates scaling effects; for example, by reducing the scale of very large variables and increasing the scale of very small variables. Unit variance scaling is important since multivariate techniques like PCA and PLS are very sensitive to scaling. A column-operating unit variance technique was implemented using ProSensus.

3.2.6.2 Standard Normal Variate (SNV)

Row-operating unit variance scaling is often employed in conjunction with mean scaling in what is referred to as standard normal variate (SNV) scaling. SNV reduces particle size (multiplicative, non-linear scattering) effects that are especially troublesome for solid mixtures.[27,28] During SNV scaling, each sample spectra is first mean centered. The standard deviation of the sample spectra is then calculated and the mean centred data is then divided by this value. Spectra treated using SNV have a mean of zero and a variance equal to unity. Detrending is sometimes used in conjunction with SNV-treated data since the latter are still susceptible to baseline curvature issues, which can be corrected by the former.

3.2.6.3 Smoothing

Savitzky-Golay smoothing is an averaging algorithm that fits a polynomial to the dataset and then predicts the point of interest from the resulting polynomial equation. Savitzky-Golay smoothing was implemented in Unscrambler® X using a polynomial degree of 2 and 9 smoothing points in total (4 points to the left and 4 points to the right of the data point of interest). Distortion to the spectra was minimal since the ATR FT-IR spectrometer was operated a resolution of 2 cm^{-1} and the smoothing algorithm smoothed over a relatively narrow range of 9 points (i.e., 18 cm^{-1}).

3.2.6.4 First derivative

In this study, first derivatives were calculated using a Savitsky-Golay algorithm whereby a polynomial with a user-selected degree and number of smoothing points (e.g., polynomial degree of 2 and 3 points smoothing used in this study) is fit to each curve segment, resulting in a spectrum that is smoothed in addition to differentiated. Taking the first derivative of a spectrum can reduce baseline offset and linear baseline issues. However, it will cause a shift in the

characteristic peaks, thus making the resulting spectrum difficult to interpret using peak-picking techniques.

3.2.6.5 Second derivative

The second derivative of a spectrum gives the rate of change in the slope of the curve and was calculated using the same Savitsky-Golay algorithm as for the first derivative (i.e., polynomial degree of 2 and 3 points smoothing). Similar to first derivative, second derivative treatment of spectra removes baseline offset and linear baseline issues. In addition, second derivative spectra retain the position of peaks such that the interpretation of second derivative spectra is generally easier than first derivative spectra. However, it should be noted that first and second derivative preprocessing might increase noise in the spectrum.

3.2.6.6 Multiplicative Signal/Scatter Correction (MSC)

Multiplicative Signal/Scatter Correction (MSC) is a method that can be used to reduce multiplicative and additive scatter effects in a dataset, such as those caused by path length variations, offset shifts, interferences, and particle size effects.[27] Mathematically, MSC is a transformation that first involves regressing the sample spectrum against an "ideal" spectrum, often estimated as the set-mean-spectra, $\overline{\mathbf{x}}(k)$, to fit parameters that describe additive and multiplicative effects.[29] The additive and multiplicative effects are estimated as the intercept and slope from the ordinary least squares solution, respectively. Once the intercept (a_i) and slope (b_i) parameters have been determined, each sample spectrum is corrected by subtracting a_i from each wavenumber absorbance then dividing that value by b_i. A detailed mathematical description of the MSC and Extended MSC (EMSC) techniques are described elsewhere.[29,30,31]

3.2.6.7 Orthogonal Signal Correction (OSC)

Orthogonal Signal Correction (OSC) is a data preprocessing technique that removes information from **X** (in this case, the matrix containing the FT-IR spectra data) that is unrelated, or orthogonal to, the response matrix, **Y** (in this case, the matrix containing the fractional compositions of the calibration mixtures). OSC is analogous to the PLS algorithm (NIPALS), with the exception that OSC attempts to *minimize* rather than *maximize* the covariance between **X** and **Y**.[32] Thus, each OSC component will remove information from **X** that shares minimal covariance (i.e., maximal orthogonality) with **Y**. Caution should be exercised when choosing the number of OSC components prior to PLS. If too many OSC components are used, the result will approach the multiple linear regression (MLR) solution and overfitting of the spectral data with respect to the response data may take place.[32] As a result, a conservative number of three OSC components were used for the data preprocessing in this study. A full mathematical description of the OSC process is outlined elsewhere.[32]

3.2.6.8 Baseline correction

Baseline correction seeks to minimize linear offset in the spectral data by subtracting either a minimum or user-defined value from the dataset. Constructing a line between two user-defined values can also impose a linear baseline. For this study, a linear baseline was constructed between the absorbance values at 800 and 1800 cm^{-1} and each subsequent wavenumber in the spectrum was normalized accordingly. For example, the general form of baseline correction is to subtract a constant absorbance value close to zero from the corresponding absorbance value at the same wavenumber for the spectrum to be corrected.

3.2.6.9 Detrending

Detrending is a spectral correction technique, similar to and often used in conjunction with SNV, that attempts to reduce nonlinear trends by fitting a polynomial equation to each spectra that is later used for baseline correction. The detrended spectra are calculated as the difference between the original spectrum and the polynomial equation describing the new baseline.

3.2.6.10 Unit vector normalization

Unit vector normalization transforms each sample spectrum to unit vectors. This vector normalization technique reduces the contribution of very intense absorbance values in the **X** matrix, thus increasing the importance of component ratios within a given surrogate mixture sample.[33]

A selection of the aforementioned preprocessing techniques that can be summarized in one or two equations are shown in Table 3.2.

Table 3.2. Formulae used to preprocess data for the twelve PLS regression models investigated in this study.

Technique	Equation(s)
Mean centering and unit variance scaling (autoscaling)	$x_{i,k}^{MC,row} = x_{i,k} - \frac{1}{K}\left(\sum_{k=1}^{K} x_{i,k}\right)$ or $x_{i,k}^{MC,column} = x_{i,k} - \frac{1}{i}\left(\sum_{i=1}^{i} x_{i,k}\right)$ Where $x_{i,k}^{MC}$ is the corrected spectrum and $x_{i,k}$ is the absorbance value of the kth wavenumber in the ith spectrum
Standard Normal Variate (SNV)	$x_{i,k}^{SNV} = \frac{1}{s}\left(x_{i,k} - \frac{1}{K}\sum_{k=1}^{K} x_{i,k}\right)$ Where $x_{i,k}^{SNV}$ is the corrected spectrum and $x_{i,k}$ is the absorbance value of the kth wavenumber in the ith spectrum and s is the calculated standard deviation of all k variables of the in the ith spectrum
Multiplicative Signal/Scatter Correction (MSC)	$\mathbf{x}_i = a_i\mathbf{1} + b_i\bar{\mathbf{x}} + e_i$ & $\mathbf{x}_i^{MSC} = \frac{\left(\mathbf{x}_i - \hat{a}_i\right)}{\hat{b}_i}$ Where $\mathbf{x}_i$ is the vector of spectral absorbances in the ith spectrum; a_i is the regression coefficient describing additive effects; $\mathbf{1}$ is a vector of 1's introduced for formality; b_i is the regression slope describing multiplicative effects; $\bar{\mathbf{x}}$ is the set-mean spectrum for all i samples; e_i is the regression model error; $\mathbf{x}_i^{MSC}$ is the MSC-corrected vector of spectral absorbances; $\mathbf{x}_i$ is the original sample spectrum; $\hat{a}_i$ is the estimated regression coefficient describing additive effects; and $\hat{b}_i$ is the estimated regression slope describing multiplicative effects
Baseline Correction	$x_{i,k}^{*} = x_{i,k} - f(k)$ Where $x_{i,k}^{*}$ is the baseline corrected absorbance at kth wavenumber in ith sample; $x_{i,k}$ is the absorbance value at kth wavenumber in ith sample; $f(k)$ is the absorbance value corresponding to some function of the wavenumber, k, for sample i (e.g., min(k) for traditional baseline correction)
Unit Vector Normalization	$\mathbf{x}_i^{*} = \mathbf{x}_i \Big/ \sqrt{\sum_{k=1}^{K} x_{i,k}^2}$ Where $\mathbf{x}_i^{*}$ is the unit vector normalized sample spectrum; $\mathbf{x}_i$ is the original sample spectrum; and $x_{i,k}$ is the absorbance value of the kth wavenumber in the ith spectrum

A total of twelve PLS regression models were developed, as summarized in Table 3.3.

Table 3.3. Summary of the twelve PLS regression models developed for this study showing the spectral range of the X matrix, preprocessing technique applied to the X matrix, and number of components selected after cross-validation of the model.

Model No.	Wavenumbers (cm^{-1})	Preprocessing Techniques	Components
1	1800 – 400 4000 – 2500	Unit variance Mean centred	3
2	1800 – 800	Unit variance Mean centred	3
3	1800 – 800	Smoothing Unit variance Mean centred	3
4	1800 – 800	Smoothing First derivative Unit variance Mean centred	3
5	1800 – 800	Smoothing Second derivative Unit variance Mean centred	3
6	1800 – 800	Smoothing MSC Unit variance Mean centred	3
7	1800 – 800	MSC Unit variance Mean centred	3
8	1800 – 800	SNV Unit variance Mean centred	3
9	1800 – 800	De-trending Unit variance Mean centred	3
10	1800 – 800	Linear baseline correction Unit variance Mean centred	3
11	1800 – 800	OSC Unit variance Mean centred	3
12	1800 – 800	Unit vector Unit variance Mean centred	3

Model 1 utilized the largest spectral domain, wavenumbers 4000 – 2500 cm^{-1} and 1800 – 800 cm^{-1}, and applied only the default unit variance and mean centering autoscaling prior to multivariate analysis. Model 2 was similar to Model 1 but wavenumbers from 1800 – 800 cm^{-1} were only considered. All subsequent models focused on applying different data pretreatments to the same wavenumber range of 1800 – 800 cm^{-1}. Model 3 employed a 9-point, 2^{nd} degree polynomial smoothing of the spectral data. Models 4 and 5 used first and second derivative spectra, respectively, with a default 9-point, 2^{nd} degree polynomial smoothing Savitsky-Golay algorithm. Model 6 employed MSC processed spectral data after smoothing while Model 7 used MSC spectral data alone. Model 8 utilized SNV corrected spectra while Model 9 utilized de-trended spectra. Model 10 was built using linear baseline corrected spectral data. Model 11 used OSC preprocessed data and finally Model 12 utilized unit vector preprocessed data for subsequent PLS regression analysis.

3.3 Results and Discussion

3.3.1 Principal Component Analysis (PCA)

In the first step of the model-building process, calibration and external validation datasets were subjected to PCA prior to data preprocessing. Three principal components were selected after cross-validation of the PCA model (PC-1, PC-2, and PC-3), in descending order of variance explained by each respective component vector. These principal components are linear combinations of the original spectral variables in **X** and are shown as the axes in Figure 3.4.

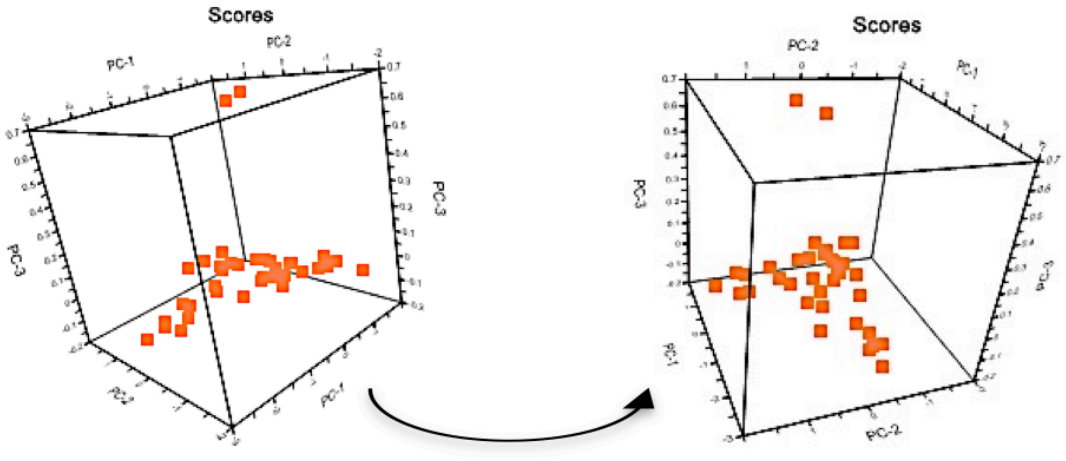


Figure 3.4. Principal component scores plot (two views) of the three components used in the analysis (PC-1, PC-2, and PC-3). The triangular cluster of datapoints represents the calibration data while the two points separated along PC-3 are the two external validation datapoints.

The triangular arrangement of the data points in Figure 3.4 reflects the expected variability between samples with varying amounts of cellulose, hemicellulose, and lignin. However, two outlier points are clearly separated from the calibration data along the third principal component (PC-3). These two points correspond to the two *Arabidopsis* samples with unknown composition. The separation between these two *Arabidopsis* samples ad the calibration samples along PC-3 indicates significant variability between these two types of spectra. This would indicate that the chemical composition of the model compounds used to calibrate the model were quite different from the chemical constituents comprising the *Arabidopsis* cultivars, which was to be expected when surrogate mixtures are invoked as calibration standards. While it is true that the physical mixtures of commercial cellulose, xylan, and lignin do not represent real plant biomass where cellulose, hemicellulose, and lignin interact and chemically cross-link in the plant cell walls, the calibration mixtures used for similar PLS regression models in literature involve fractionating the said components from real plant biomass, which ultimately disrupts these inter-polymer interactions anyway. As a result, we felt that while our dataset was not

necessarily representative of lignocellulose in its native state, it was not dissimilar to other studies where conventional wet-chemical techniques were used to isolate the cellulose, hemicellulose, and lignin fractions. This is not to say that the calibration mixtures utilized herein could not be improved - just that native state of the biomass within the cell wall is ultimately disrupted using our approach or previous literature studies. Despite the imperfect match, the calibration dataset was later used to construct a predictive PLS regression model. A different choice in model compounds could have led to better (or worse) model performance and will be investigated in future studies.

3.3.2 Partial Least Squares (PLS) regression

A total of twelve PLS regression models were built using the model compound calibration mixtures and validated using both cross-validation and the external *Arabidopsis* samples. The performance of each model was evaluated based on the number of components yielded by cross-validation; the coefficient of determination ($R^2 = 1 - (SSE/TSS)$; where SSE is the sum of squares error and TSS is the total sum of squares); the coefficient of prediction ($Q^2 = 1 - (PRESS/TSS)$; where PRESS is the predicted residuals sum of squares and TSS is the total sum of squares); the root mean square error of calibration or cross-validation (RMSEC or RMSECV); and the root mean square error of prediction (RMSEP). RMSEP was determined using the five external validation mixtures mentioned in the Mixture Design section – these mixtures were prepared in the same manner as the calibration samples, but with different compositions. Figure 3.5 shows the R^2 and Q^2 values for each of the models, which provide a relative estimate of the performance of the model calibrations and validations, respectively.

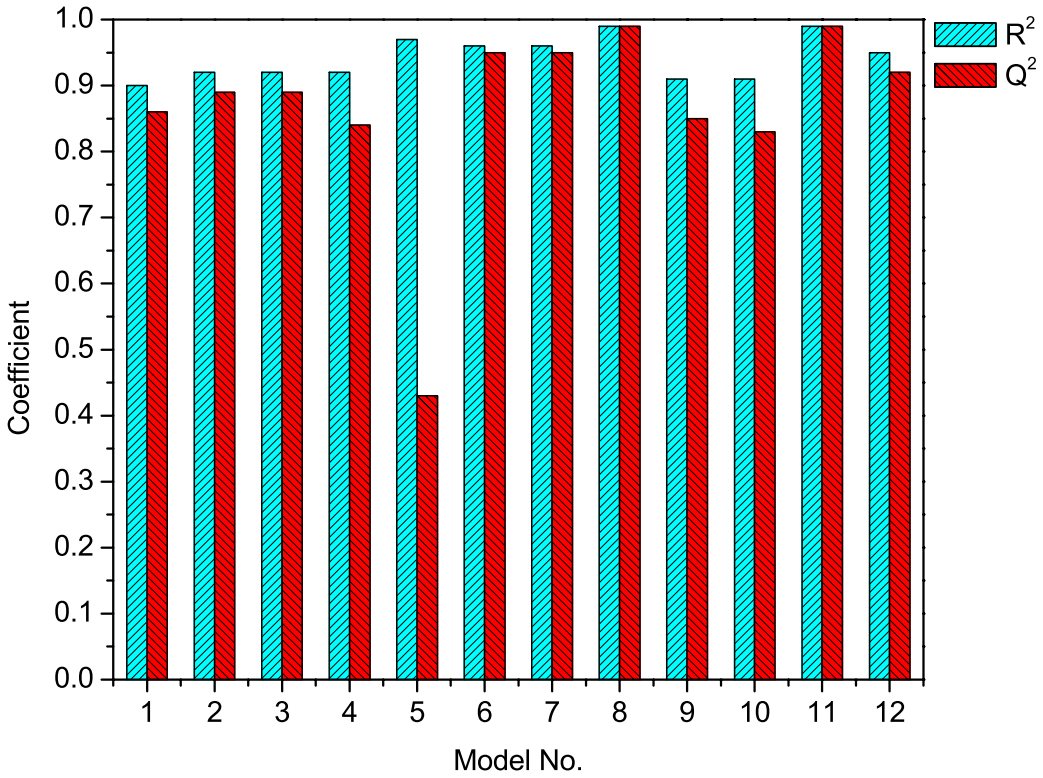


Figure 3.5. Performance of various data pretreatment regimes on the PLS regression model using R^2 and Q^2 as metrics.

From Figure 3.5, it can be noted that most of the PLS regression models showed satisfactory R^2 and Q^2 values, which means that significant variability was explained by both the calibration and prediction procedures. However, a clear exception is Model 5 (smoothing, second derivative) where taking the second derivative could have led to increased spectral noise and thus a poor Q^2 value (i.e., 0.43).

The RMSE of calibration and prediction (RMSEC and RMSEP) for each of the twelve models with respect to the cellulose, hemicellulose (xylan), and lignin fractions was also recorded in an effort to assess the performance of the data preprocessing techniques. Figure 3.6 illustrates the RMSEC and RMSEP of each model, where model performance was found to be inversely proportional to RMSE.

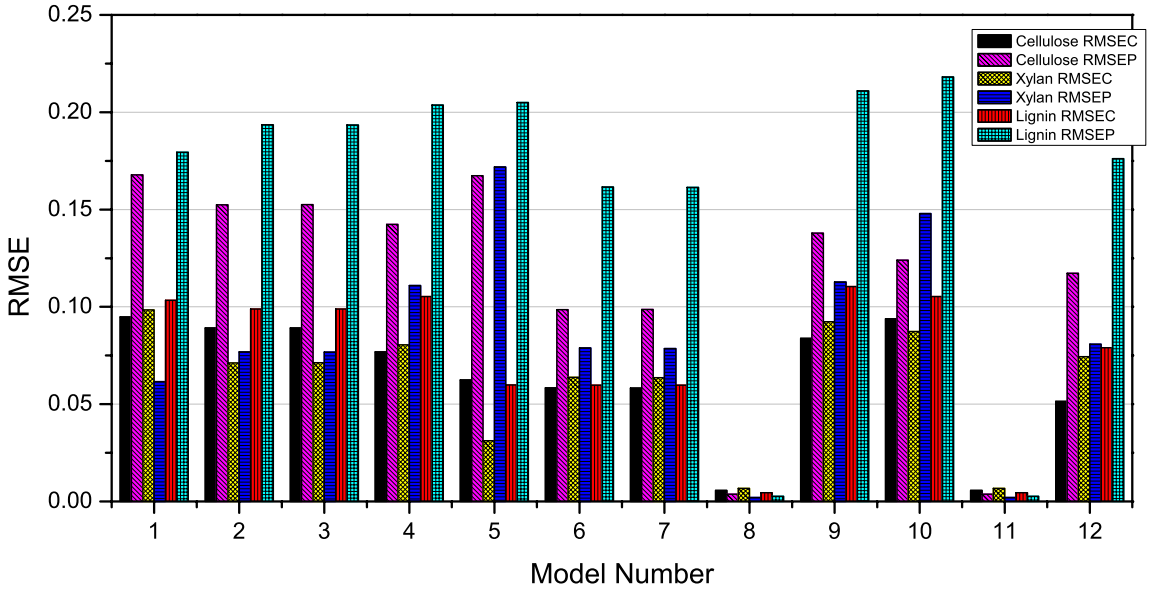


Figure 3.6. RMSE of calibration (cross-validation) and prediction of the twelve models evaluated in this study. In general, the lower the RMSE, the better the model's predictive capability.

The RMSE graph shows that Models 8 and 11, i.e. the models utilizing SNV and OSC preprocessing, respectively, gave the lowest error and best predictive performance with respect to compounds within the ternary surrogate mixture. Previous studies have reported that infrared reflectance spectra of mixtures can be subject to non-linear effects caused by scattering, which are minimized by applying SNV.[27,28] The low RMSEP of Models 8 and 11 suggest that SNV and OSC could be effective methods for correcting ATR FT-IR spectra of powdered samples prior to multivariate analysis. SNV is probably a more robust preprocessing technique because there are less user-selected parameters involved in SNV compared to OSC where a value for the OSC components is required.

It should be noted that these RMSE values were derived from models based solely on the model lignocellulosic compound mixtures obtained by mixing various fractions of the model cellulose, hemicellulose, and lignin compounds. Despite the fact that many of the models performed well against the aforementioned metrics (especially R^2 and Q^2), only two out of the

twelve models yielded plausible compositional results: those models were built using first and second derivative data pretreatment. Figure 3.7 shows the performance of Model 4 (smoothing, first derivative) with respect to predicting B10 and C10 composition values.

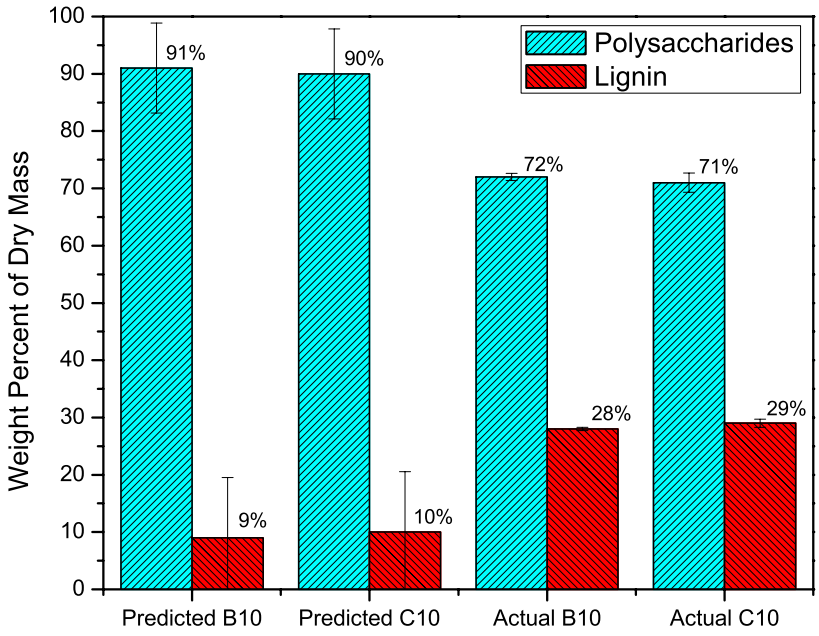


Figure 3.7. Performance of model 4 (smoothing, first derivative) with respect to predicting the composition of B10 and C10, assuming the entire plant is either polysaccharides or lignin. The predicted B10 and C10 values correspond to those from the multivariate model while the actual B10 and C10 correspond to values provided by a third-party wet chemical analysis of the two biomass samples. Error bars for the prediction values represent the RMSEC while the error bars for the actual values represent the RMSE; for polysaccharides, the average RMSEC of cellulose and xylan was used.

Cellulose and xylan (hemicellulose) values were lumped together as "Polysaccharides" for the sake of comparison with the conventional characterization data provided by the third-party, which gave values for total sugars and lignin. These values were then normalized to give polysaccharides and lignin content summing to 100%, i.e., assuming the entire plant was composed of either polysaccharides or lignin. Model 4 gave a B10 composition of 91 ± 8 wt.%

polysaccharides (cellulose + xylan) and 9 ± 11 wt.% lignin. Similarly for C10, the composition from the same PLS regression model was 90 ± 8 wt.% polysaccharides and 10 ± 11 wt.% lignin. The multivariate model overestimated the total polysaccharides content of B10 and C10, while lignin content was underestimated. Despite the discrepancy between absolute values, the model succeeded in predicting chemical similarity between B10 and C10, which was supported by the analyses results that showed increased biomass content of C10 over B10, but almost identical polysaccharides and lignin content. Model 5 (smoothing, second derivative) also predicted chemical similarity between B10 and C10 with improved accuracy. The predictive performance of Model 5 (second derivative) with respect to predicting the compositions of B10 and C10 are shown in Figure 3.8.

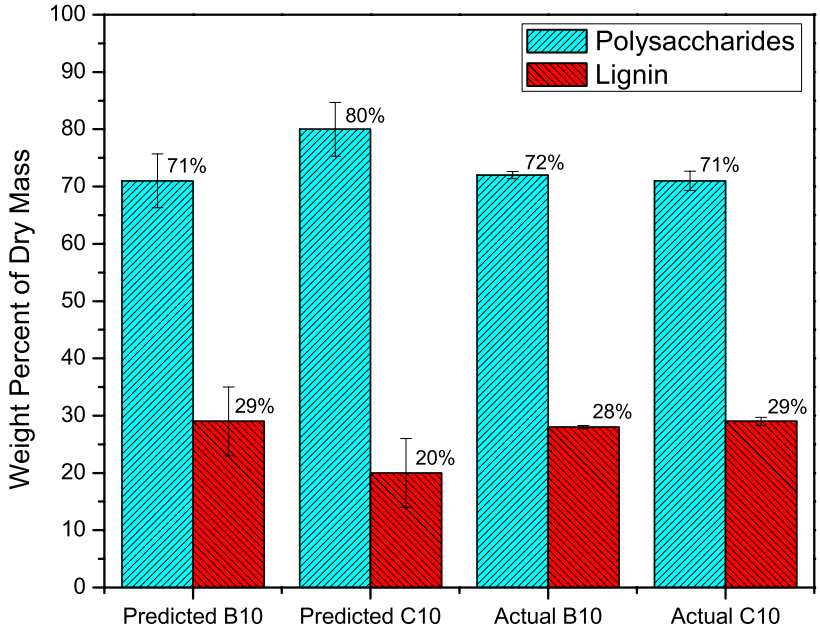


Figure 3.8. Performance of model 5 (smoothing, second derivative) with respect to predicting the composition of C10 and B10, assuming the entire plant is either polysaccharides or lignin. The predicted B10 and C10 values correspond to those from the multivariate model while the actual B10 and C10 correspond to values provided by a third-party wet chemical analysis of the two biomass samples. Error bars for the prediction values represent the RMSEC while the error bars for the actual values represent the RMSE; for polysaccharides, the average RMSEC of cellulose and xylan was used.

Model 5 gave a B10 composition of 71 ± 5 wt.% polysaccharides (cellulose + xylan) and 29 ± 6 wt.% lignin. Similarly for C10, the composition from the same PLS regression model was 80 ± 5 wt.% polysaccharides and 20 ± 6 wt.% lignin. Traditional wet chemical techniques gave a B10 composition of 72 ± 0.6 wt.% polysaccharides and 28 ± 0.3 wt.% lignin, and a C10 composition of 71 ± 1.7 wt.% polysaccharides and 29 ± 0.7 wt.% lignin. Model 5 exhibited higher predictive capabilities for the polysaccharide and lignin contents of B10 and C10 compared to the model employing first derivative spectra (Model 4). Dicotyledons (or dicots) like *Arabidopsis* contain roughly 45-50% cellulose, 25% hemicellulose as mostly xylans, 20% phenolics as lignin, and 0.1% pectins.[34,35] As much as 10% of the overall mass of biomass

samples can be attributed to trace materials like silica or protein. The results from PLS regression Model 5 suggests that surrogate mixtures can adequately predict, with reasonable certainty, the relative quantities of polysaccharides and lignin within a ground lignocellulosic biomass sample. The **P** loadings for the truncated spectra showed significant contributions to representing the variability of the underlying data over the selected "fingerprint" region of 1800-800 cm^{-1}. However, **P** loadings that were not significant in this region could have been easily removed using the software, but this was avoided since it would introduce user bias into the model building process. The two main issues encountered in this study were the non-linearities in the sample spectra and the choice of model compounds used in the calibration mixtures. The heterogeneous nature of the plant cell wall does indeed present some significant challenges to physical characterization techniques such as ATR FT-IR. The heterogeneity of the powdered samples that were placed on the ATR crystal had some variability associated with differences in particle size between the commercial cellulose, xylan, and lignin samples. The degree of over-interpretation due to sample heterogeneity is ultimately reflected in the predictive ability of the model. For this reason, replicate spectra were used on well-mixed powdered samples to reduce the effect of particle size on the predictive ability of the model. This model could be further improved by careful selection of the model compounds used for the calibration mixtures. Perhaps cellulose, xylan, or lignin extracted from a dicot species rather than Beechwood, for example, would further improve the predictive capability of the model, particularly if SNV (or OSC) preprocessing was applied to the ATR FT-IR spectra to reduce non-linear effects. Grinding and sieving of the calibration mixture components to produce a more homogeneous and monodisperse particle size distribution could also mitigate non-linear particle size effects, although Indulin AT (Kraft lignin), which is supplied as a fine powder, may have to be replaced with an appropriate alternative.

3.4 Conclusions

This study showed that surrogate mixtures could be used to build a predictive multivariate model for the compositional analysis of lignocellulosic biomass using ATR FT-IR spectroscopy. The effect of ATR FT-IR data preprocessing was also assessed to determine the most suitable technique for this particular type of multivariate model. Second derivative data preprocessing gave relatively accurate predictions of the composition of B10 and C10. SNV and OSC preprocessing showed the best performance with respect to RMSEC and RMSEP, but unfortunately these models could not accurately resolve the composition of B10 and C10 due to the nature of the calibration mixtures.

3.5 Acknowledgements

The authors would like to thank Dr. David Hyndman for providing the two *Arabidopsis* cultivars and Prof. Shawn Mansfield for conducting the third-party compositional analysis of these samples. The authors would like to acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC) for financial support through the Discovery Grant Program (Champagne, Cunningham). The authors would also like to acknowledge the Ministry of Research Innovation for their support through the Early Researcher Award Program (Champagne) and Queen's University for their support by providing a Queen's Graduate Award and facilitating a G.E. Ted Courtnage Graduate Award in Engineering (Krasznai). The authors would like to thank the Courtnage family for establishing this award.

3.6 References

(1) Allen, S. G.; Kam, L. C.; Zemann, A. J.; Antal, M. J. *Ind. Eng. Chem. Res* **1996**, *35*, 2709–2715.

(2) Gáspár, M.; Juhász, T.; Szengyel, Z.; Réczey, K. *Process Biochemistry* **2005**, *40*, 1183–1188.

(3) Zhang, Y.-H. P.; Ding, S.-Y.; Mielenz, J. R.; Cui, J.-B.; Elander, R. T.; Laser, M.; Himmel, M. E.; McMillan, J. R.; Lynd, L. R. *Biotechnol. Bioeng.* **2007**, *97*, 214–223.

(4) Shill, K.; Padmanabhan, S.; Xin, Q.; Prausnitz, J. M.; Clark, D. S.; Blanch, H. W. *Biotechnol. Bioeng.* **2010**, *108*, 511–520.
(5) Shafiei, M.; Karimi, K.; Taherzadeh, M. J. *Bioresource Technology* **2010**, *101*, 4914–4918.
(6) Qi, B.; Chen, X.; Wan, Y. *Bioresource Technology* **2010**, *101*, 4875–4883.
(7) Li, W. Y.; Jin, A. X.; Liu, C. F.; Sun, R. C.; Zhang, A. P.; Kennedy, J. F. *Carbohydrate Polymers* **2009**, *78*, 389–395.
(8) Sievers, C.; Valenzuela-Olarte, M. B.; Marzialetti, T.; Musin, I.; Agrawal, P. K.; Jones, C. W. *Ind. Eng. Chem. Res* **2009**, *48*, 1277–1286.
(9) Chen, W.; Yu, H.; Liu, Y.; Chen, P.; Zhang, M.; Hai, Y. *Carbohydrate Polymers* **2011**, *83*, 1804–1811.
(10) Nkansah, K.; Dawson-Andoh, B.; Slahor, J. *Bioresource Technology* **2010**, *101*, 4570–4576.
(11) Chen, H.; Ferrari, C.; Angiuli, M.; Yao, J.; Raspi, C.; Bramanti, E. *Carbohydrate Polymers* **2010**, *82*, 772–778.
(12) Krongtaew, C.; Messner, K.; Ters, T.; Fackler, K. *BioResources* **2010**, *5*, 2081–2096.
(13) Hames, B.; Thomas, S.; Sluiter, A.; Roth, C.; Templeton, D. *Appl Biochem Biotechnol* **2003**, *105*, 5.
(14) Wold, S.; Sjöström, M.; Eriksson, L. *Chemometrics and Intelligent Laboratory Systems* **2001**, *58*, 109–130.
(15) Workman, J.; Koch, M.; Lavine, B.; Chrisman, R. *Anal. Chem.* **2009**, *81*, 4623–4643.
(16) Chen, H.; Ferrari, C.; Angiuli, M.; Yao, J.; Raspi, C.; Bramanti, E. *Carbohydrate Polymers* **2010**, *82*, 772–778.
(17) Kadam, S. S.; van der Windt, E.; Daudey, P. J.; Kramer, H. J. M. *Crystal Growth & Design* **2010**, *10*, 2629–2640.
(18) Shao, J. *Journal of the American Statistical Association* **1993**, *88*, 486–494.
(19) Xu, Q.-S.; Liang, Y.-Z. *Chemometrics and Intelligent Laboratory Systems* **2001**, *56*, 1–11.
(20) Otto, M. *Chemometrics*; statistics and computer application in analytical chemistry; 2nd ed. Wiley-VCH: Weinheim, **2007**; p. 328.
(21) Rodrigues, J.; Puls, J.; Faix, O.; Pereira, H. *Holzforschung* **2001**, *55*, 265–269.
(22) Liang, C. Y.; Marchessault, R. H. *Journal of Polymer Science* **1959**, *39*, 269–278.
(23) Nelson, M. L.; O'Connor, R. T. *J. Appl. Polym. Sci.* **1964**, *8*, 1311–1324.
(24) Higgins, H. G.; Stewart, C. M.; Harrington, K. J. *Journal of Polymer Science* **1961**, *51*, 59–84.
(25) Boysworth, M. K.; Booksh, K. S. In *Handbook of near-infrared analysis*; Burns, D. A.; Ciurczak, E. W., Eds. CRC Press: Boca Raton, FL, **2008**; pp. 209–239.
(26) Nomikos, P.; MacGregor, J. F. *AIChE Journal* **2004**, *40*, 1361–1375.
(27) Dhanoa, M. S.; Lister, S. J.; Sanderson, R.; Barnes, R. J. *Journal of Near Infrared Spectroscopy* **1994**, *2*, 43–47.
(28) Blanco, M.; Coello, J.; Iturriaga, H.; Maspoch, S. *Chemometrics and Intelligent Laboratory Systems* **2000**, *50*, 75–82.
(29) Helland, I. S.; Naes, T.; Isaksson, T. *Chemometrics and Intelligent Laboratory Systems* **1995**, *29*, 233–241.
(30) Martens, H.; Nielsen, J. P.; Engelsen, S. B. *Anal. Chem.* **2003**, *75*, 394–404.
(31) Martens, H.; Stark, E. *Journal of Pharmaceutical and Biomedical Analysis* **1991**, *9*, 625–635.
(32) Wold, S.; Antti, H.; Lindgren, F.; Öhman, J. *Chemometrics and Intelligent Laboratory*

Systems **1998**, *44*, 175–185.
(33) Steinfath, M.; Groth, D.; Lisec, J.; Selbig, J. *Physiologia Plantarum* **2008**, *132*, 150–161.
(34) Ishii, T. *Plant Science* **1997**, *127*, 111–127.
(35) Gomez, L. D.; Bristow, J. K.; Statham, E. R.; McQueen-Mason, S. J. *Biotechnol Biofuels* **2008**, *1*, 15.

Chapter 4

Amphiphilic Block Copolymers Containing Polysaccharide and Synthetic Segments

Abstract

The synthesis of copolymers containing both renewable and synthetic materials has been regarded as a method to combine the diversity and versatility of natural biopolymers with the precise science of polymer chemistry. Atom Transfer Radical Polymerization (ATRP) and Single Electron Transfer Living Radical Polymerization (SET-LRP) are transition metal mediated polymerization techniques whereby well-defined synthetic monomers can be polymerized from the reducing end of a polysaccharide without the need for protecting group chemistry. Here, we report on the attempted synthesis of a novel water-soluble hydrazide alkyl halide initiator (2-bromo-2-methylpropane hydrazide hydrobromide) capable of coupling to the reducing end of polysaccharides *via* reductive amination. After this coupling reaction, ATRP or SET-LRP was attempted from the exposed alkyl bromide moiety at the reducing end in a mixed DMSO/water solvent system. In this manner, a synthetic polymer was to be grown from the reducing end of the polysaccharide, allowing for the final architecture and properties (e.g., self assembly) of the resulting glycoconjugate to be finely tuned. These hybrid natural/synthetic materials could potentially be used as graft-modified natural polymers, in environmental applications or as drug delivery vehicles. However, due to stability issues arising from the reactivity of the hydrazide and alkyl bromide components, the initiator that was synthesized and coupled to the reducing end of dextran could not initiate an ATRP or SET-LRP reaction. Despite this, the coupling reaction

between the hydrazide moiety of the initiator and the reducing end of dextran was quite efficient and should be further explored as a means to improve the troublesome reductive amination step.

4.1 Introduction

Hybrid polymers containing polysaccharides and synthetic blocks are becoming increasingly diverse in structure and application. Graft copolymers containing synthetic polymers and polysaccharides have been synthesized using a variety of different materials and polymer chemistries.[1,2,3,4] Regioselective approaches towards these synthetic/natural copolymers, such as block copolymers, are fewer in number, but a significant amount of work has recently been published on the subject.[5] Synthetic/polysaccharide block copolymers have the propensity, under certain conditions, to aggregate into stable colloidal systems with tuneable behaviour; a topic better summarized by Gryzbowski et al..[6] This process of self-assembly allows copolymers to aid in the encapsulation and transport of pharmaceuticals or therapeutics.[7,8] Because mammalian cells are covered in an extracellular matrix that contain both protein and polysaccharide components[9], the concept of designing polymer architectures where the surface integrates polysaccharide moieties has become an invaluable tool for polymer chemists interested in producing biocompatible materials.[10,11,12] Moreover, the abundant supply and renewability of plant-based polysaccharides, like cellulose, coupled with the finite reserves of petroleum has encouraged the development of new materials containing these naturally occurring sugar-based polymers.

Atom Transfer Radical Polymerization (ATRP) is a controlled/living radical polymerization that is initiated by an alkyl halide and mediated by a transition metal bound to an appropriate ligand.[13] ATRP can be used to produce a myriad of well-defined polysaccharide-containing copolymer architectures in organic solvents.[14,15] Because organic solvents are required

for conventional ATRP, polysaccharides are often protected before polymerization to facilitate dissolution in organic solvents and then later deprotected to give the final amphiphilic copolymer. Polymerizations akin to conventional ATRP have recently been demonstrated in polar solvents using elemental copper as the source of catalytic activity; a technique often referred to as Single Electron Transfer Living Radical Polymerization (SET-LRP).[16] Percec and coworkers assert that, in a polar solvent and in the presence of a ligand, SET-LRP proceeds with the activation of a sufonyl or alkyl halide initiator by Cu(0) to form Cu(I)X (where X = Br, Cl, I), which then rapidly disproportionates into pairs of Cu(0) and Cu(II)X_2 species.[16] Acrylic, methacrylic, or vinyl chloride monomers can be polymerized in this fashion with high end-group fidelity at room temperature.[17] The Cu(II)X_2 species acts as the deactivator, reversibly capping the growing polymer chain with a bromine group.[18] Cu(0)-mediated LRP, or SET-LRP, is claimed to proceed *via* an outer shell electron transfer (OSET) process, rather than an atom transfer process; however, the mechanistic differences between ATRP and SET-LRP are currently contested in literature.[19,20,21] Both ATRP and SET-LRP mechanisms rely on the rate of deactivation (k_{deact}) of the radical species being orders of magnitude higher than the rate of activation (k_{act}); thus, the polymerization proceeds in a controlled/living fashion because the dormant species is favoured over the active species.[13,16,22] Cu(0)-mediated LRP has some distinct advantages over ATRP including rapid polymerization kinetics, high-end group fidelity at high conversions and M_n, and improved removal and recycling of the catalyst system since copper wire can be used as the mediator.[16] Recent work by Chan et al.[23] has also provided a basis for the continuous production of polymers in a copper tube using SET-LRP. With respect to the current study, the most significant advantage of Cu(0)-mediated LRP over ATRP is the capability of using polar polymerization solvents that can solubilize polysaccharides (e.g. DMSO or ionic liquids), thus mitigating protection/deprotection of the polysaccharide block that would have been required for

conventional ATRP in non-polar solvents. Homogenous Cu(0)-mediated LRP of methyl acrylate (MA) and methyl methacrylate (MMA) using backbone-initiated hemicellulose from spruce trees has recently been demonstrated Voepel et al.[24] and Edlund et al.[25].

Amphiphilic block copolymers containing a polysaccharide and a synthetic segment were prepared by Houga et al. using a novel amine-terminated ATRP initiator (*N*-(2-aminoethyl)-2-bromo-2-methylpropanamide) to initiate dextran at its reducing-end aldehyde moiety.[26] Figure 4.1 shows the entire synthesis protocol for this particular amphiphilic block copolymer.

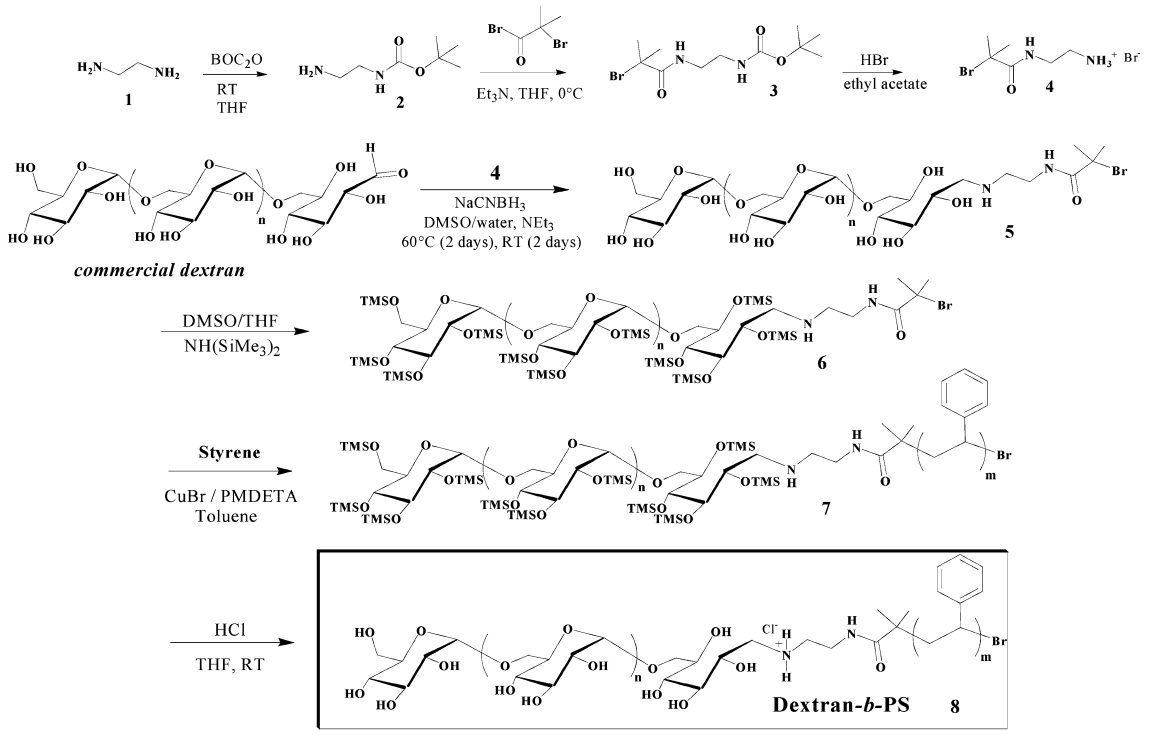


Figure 4.1. Synthesis procedure of dextran-*b*-polystyrene using a novel amine-terminated ATRP initiator. Reproduced with permission from Houga et al.[26].

The ATRP initiator was synthesized by first reacting ethylenediamine with BOC_2O to produce an amine-terminated, BOC-protected compound. The free amine group was then reacted with bromoisobutyryl bromide to give the BOC-protected initiator, which after treatment with HBr

solution gave the initiator as its bromide salt. The initiator was then coupled to dextran *via* reductive amination. The reducing-end initiated dextran was then silylated using $NH(SiMe_3)_2$ and subsequently dissolved in organic solvent for ATRP with styrene. The resulting block copolymer (silylated dextran-*b*-polystyrene) was fully soluble in organic solvent, which gave the amphiphilic block copolymer dextran-*b*-polystyrene upon hydrolysis of the trimethylsilyl (TMS) groups on the backbone of dextran using HCl.

Similarly, Yagi et al.[2] synthesized linear cellulose-*b*-polystyrene using a commercially available amide-terminated ATRP initiator, chloroacetamide. Reductive amination was implemented in a homogenous solvent system of dimethylacetamide (DMAc) LiCl with cellulose of varying degrees of polymerization (DP). After the cellulose-Cl macroinitiator was recovered, homogenous ATRP was conducted with styrene from the reducing end of cellulose-Cl in DMAc/LiCl. The synthesis procedure, from the preparation of cellulose-Cl to the homogenous ATRP of this macroinitiator with styrene, is illustrated in Figure 4.2.

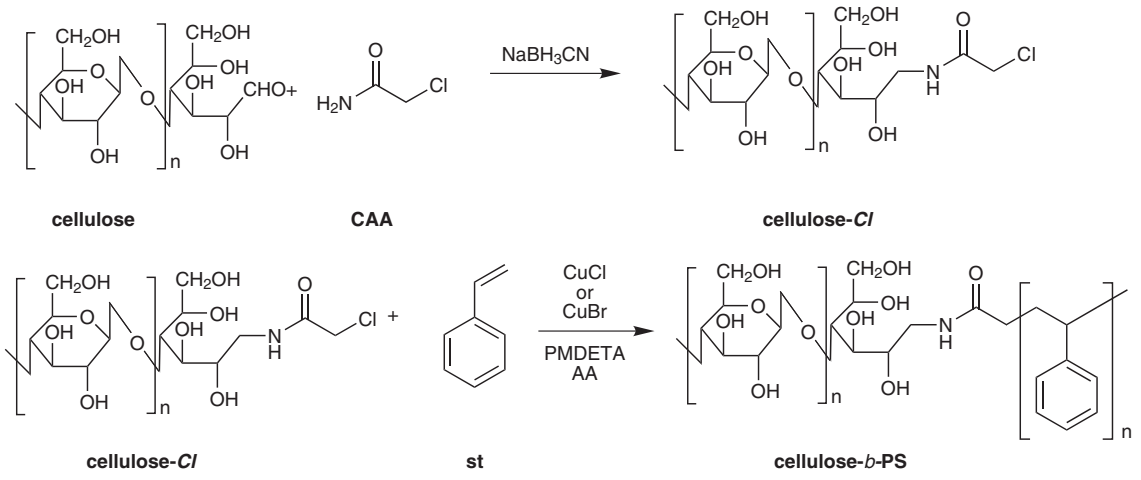
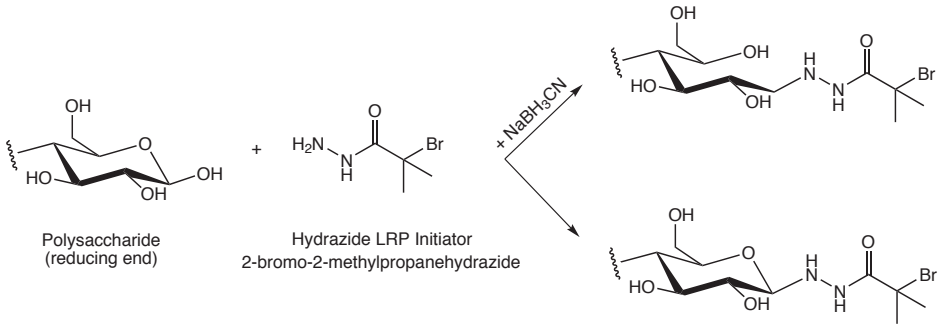


Figure 4.2. Synthesis of cellulose-*b*-polystyrene using chloroacetamide and reductive amidation to produce the cellulose-Cl macroinitiator (CA: chloracetamide, st: styrene, PS: polystyrene). Reproduced with permission from Yagi et al.[2].

Amides are less nucleophilic than amines due to a decreased electron density on the nitrogen atom from conjugation with the adjacent electrophilic carbonyl group, so the reductive amidation step employed by Yagi et al.[2] was much slower and less efficient than that of Houga et al.[26]. The weak nucleophilicity of chloroacetamide was reflected in the size exclusion chromatography (SEC) elution profiles, where a lower molecular weight peak from the cellulose-Cl macroinitiator and/or unreacted cellulose was present in addition to the expected higher molecular weight peak from cellulose-*b*-polystyrene.

A recent communication by Verma et al.[27] showed that microwave irradiation could expedite the reductive amination of dextran by generating aldehyde groups from the reducing-end hemiacetal moiety *via* non-thermal effects. The reductive amination reaction required 4 hours instead of 72 hours using BOC-ethylenediamine as the nucleophilic compound. Following reductive amination, the BOC-ethylenediamine bound to the reducing-end of dextran was deprotected with hydrochloric acid and the terminal -NH_2 group was used as a coupling agent with poly(D,L-lactide) to produce an amphiphilic dextran-*b*-poly(D,L-lactide) copolymer.

Here, we report on the synthesis of a novel hydrazide-containing alkyl bromide initiator for the homogenous Cu(0)-mediated LRP of synthetic monomers from the reducing-end of polysaccharides. Because hydrazide functionalities are much more nucleophilic than amides and amines, it was hypothesized that this route (Scheme 4.1) would improve both the reductive amination step and control over the subsequent polymerization.



Scheme 4.1. Proposed reaction of reducing-end with the LRP initiator (2-bromo-2-methylpropane hydrazide). This reaction should proceed in the absence of $NaBH_3CN$ to give the reducing-end initiated polysaccharide in its ring-closed form (bottom right). Adapted from Hermanson[28].

Furthermore, because hydrazides can form stable adducts with the reducing-end of polysaccharides in the absence of reducing agent (e.g., $NaBH_3CN$), this novel initiator could facilitate a greener approach to polysaccharide-containing block copolymers.

4.2 Experimental

4.2.1 Materials

2-Bromoacetamide (BA, Aldrich, 98%), isonicotinic hydrazide (isoniazid, Aldrich, 99%), ethyl α-bromoisobutyrate (EBiB, Aldrich, 98%), methyl 2-bromopropionate (MBiP, Aldrich, 98%) were used as received.

The acid bromide, α-bromoisobutyryl bromide (BriB, Aldrich, 98%); initiator synthesis solvent, tetrahydrofuran (THF, Aldrich, 99%); deprotection solvent, ethyl acetate (EtOAc, Fisher, reagent grade); polymerization solvent, deuterated dimethyl sulfoxide (DMSO-d6, Aldrich, 99.96%); protected hydrazide, *tert*-butyl carbazate (BOC-hydrazine, Aldrich, 98%); polysaccharide, dextran (*Leuconostoc* spp., M_W = 5678 g mol^{-1}, M_N = 3154, M_W / M_N = 1.80, Sigma-Aldrich); organic base, triethylamine (TEA, Sigma-Aldrich, 99%); deprotection agent, hydrogen bromide solution (33 wt.% HBr in acetic acid, Sigma-Aldrich); and reducing agent,

sodium cyanoborohydride ($NaBH_3CN$, Aldrich, 95%) were all used as received. The monomer, methyl acrylate (MA, Aldrich, 99%) was passed through a packed column containing basic aluminum oxide (Aldrich, ~150 mesh, 58 Å) to remove inhibitor and stored in a refrigerator at 5 °C prior to polymerizations. The ligand, tris[2-(dimethylamino)ethyl]amine (Me_6-TREN) was synthesized according to the methods of Britovsek et al[29] and stored in a refrigerator prior to use. The mediator, copper wire (Cu(0), 1.55 mm outer diameter), was purchased from The Home Depot Canada and washed with concentrated sulfuric acid for 10 minutes and dried under nitrogen immediately prior to use.

4.2.2 Methods

4.2.2.1 Synthesis of LRP initiator

Formation of the BOC-protected initiator

In a 250 mL three-neck round bottom flask containing a magnetic stir bar, BOC-hydrazine (7.0 g, 53 mmol) followed by approximately 100 mL (88.9 g, 1230 mmol) of anhydrous THF was added and stirred. After the BOC-hydrazine was fully dissolved, TEA (5.37 g, 7.40 mL, 53 mol) was added and the round bottom flask was connected to a condenser and placed inside an ice bath on a stir plate. Nitrogen was then introduced into the headspace of the flask and condenser *via* an 18-gauge needle in order to eliminate oxygen and residual moisture. BriB (12.18 g, 6.55 mL, 53 mmol) was then diluted with about 20 mL with THF and taken up into a 50 mL gas-tight syringe (Note: A glass-only plunger was used since rubber seals tended to degrade upon contact with BriB solution) and injected drop wise into the reaction flask over 20 minutes. The reaction flask was removed from the ice bath following the addition of BriB and allowed to reach room temperature and react over the next 48 hours. A white precipitate of triethylammonium bromide (TEAm-Br) was removed by vacuum filtration using a fritted glass

filter funnel containing Celite filter-aid equipped on a vacuum flask. The filtrate containing the BOC-protected initiator was then collected and concentrated on a rotary evaporator to give an off-white powder. Yield = 90%. ^{1}H NMR (400 MHz, $CDCl_3$, ppm): δ = 8.32 (-NH-, Br side), 6.54 (-NH-, BOC side), 1.98 (-CH_3, Br side), 1.47 (-CH_3, BOC side). ^{13}C NMR (400 MHz, $CDCl_3$, ppm): δ = 172 (-(*C*=O)-, initiator side), 155 (-(*C*=O)-, BOC side), 82 (-(C=O)-O-*C*-$(CH_3)_3$), 59 (-(C=O)-(*C*Br)-$(CH_3)_2$), 32 (-(C=O)-(CBr)-(*C*$H_3)_2$), 28 (-(C=O)-O-C-(*C*$H_3)_3$).

Deprotection of the BOC-protected initiator

In a 250 mL three-neck round bottom flask containing a magnetic stir bar, the BOC-protected initiator (5.9 g, 21 mmol) and EtOAc (60.0 mL, 54 g, 610 mmol) were added and stirred to give a clear solution. After allowing time for the BOC-protected initiator to dissolve, HBr solution (6.0 mL, 8.0 g, 2.6 g HBr, 33 mmol HBr) was introduced drop wise into the round bottom flask through an open neck using a long pipette since the HBr solution was found to either dissolve or react with the brown painted seal in glass syringes as well as the metal portions of Teflon syringes. A white precipitate of the initiator in its ammonium form was visible after about 10 minutes of stirring following the addition of HBr solution. The precipitated salt was then filtered over a 0.45 μm PTFE membrane on a Büchner funnel inside a vacuum flask. The precipitate was then collected from the filter membrane and washed in excess EtOAc to remove excess acid and unprotected BOC-initiator. Yield = 72%. ^{1}H NMR (400 MHz, D_2O, ppm): δ = 1.94 ($(CH_3)_2$). ^{13}C NMR (400 MHz, D_2O, ppm): δ = 172 (-(*C*=O)-), 55 (-(C=O)-(*C*Br)-$(CH_3)_2$), 30 (-(C=O)-(CBr)-(*C*$H_3)_2$).

4.2.2.2 Reductive amination of LRP initiator with dextran

In a 20 mL scintillation vial, the hydrazide initiator (1.15 g, 4.40 mmol, 25 eq.) in its ammonium form and TEA (0.61 mL, 4.40 mmol, 25 eq.) was added along with a 9:1 v/v solvent

mixture of DMSO/water (Millipore). The vial was mixed on a rotary shaker until cool to the touch and clear. This mixture was then transferred to a 100 mL three-neck round bottom flask containing a magnetic stir bar and dextran (1.00 g, 0.176 mmol, 1 eq.) was added followed by $NaBH_3CN$ (1.38 g, 22.0 mmol, 125 eq.) as the reducing agent. The reaction flask was then transferred to a hot plate and stirred at a constant temperature of 60 °C for 48 hours and room temperature for an additional 48 hours. The resulting reducing end initiated dextran was precipitated in excess ethanol and recovered as a fine precipitate after decantation and drying at room temperature under vacuum. Excess initiator dissolved in ethanol was removed *via* decantation prior to drying under vacuum.

4.2.2.3 Cu(0)-mediated living radical polymerization (LRP)

Copper wire (Cu(0)) was first activated for 10 minutes in concentrated sulfuric acid and then rinsed with copious amounts of acetone and dried under a stream of nitrogen prior to use. In a 100 mL Schlenk tube containing a magnetic stir bar, methyl acrylate (MA, 2 mL, 1.87 g, 22 mmol, 255 eq.), 3 mL of DMSO-d6, and activated Cu(0) wire (3 cm, 0.50 g) were added. A stock solution of $Cu(II)Br_2$ and Me_6TREN was prepared in advance and fully dissolved in DMSO-d6 using a sonication bath. Then, 1 mL of the $Cu(II)Br_2$ (1×10^{-3} g, 4.4×10^{-3} mmol, 0.052 eq.) and Me_6TREN (1.5×10^{-2} g, 6.4×10^{-2} mmol, 0.75 eq.) stock solution was added to the Schlenk tube. The order of addition into the Schlenk tube was MA, Cu(0), and then 1 mL of the $Cu(II)Br_2/Me_6TREN$ stock solution to give the molar amounts listed above. The mixture was then frozen with liquid nitrogen and subjected to three consecutive freeze-pump-thaw cycles. After the last freeze-pump-thaw cycle, a nitrogen blanket was introduced into the Schlenk tube. A stock solution of the initiator in DMSO-d6 was prepared in advance and bubbled with nitrogen for at least 30 minutes prior to addition to the Schlenk tube. After the Schlenk tube containing the degassed Cu(0), MA, $Cu(II)Br_2$, and Me_6TREN was fully thawed, 1 mL of the initiator stock

solution (0.0854 mmol, 1 eq.) was added *via* a degassed syringe to give a final DMSO-d6 volume of approximately 5 mL. The polymerization was allowed to proceed at 25 °C in a thermostatted oil bath with periodic sampling for ^{1}H-NMR (conversion) and SEC (molecular weight) analysis.

4.2.3 Characterization

4.2.3.1 Size Exclusion Chromatography (SEC)

Samples were withdrawn at recorded intervals from the polymerization flask with a degassed syringe, then injected into open 20 mL scintillation flasks prior to SEC analysis. Polymer samples were dried, weighed, and dissolved in approximately 2 mL of THF to give a concentration of about 20 mg·mL^{-1}. After complete dissolution, the samples were filtered through 0.2 μm nylon syringe filters into auto-sampler vials for Linear Size Exclusion Chromatography (LSEC).

LSEC was performed using a Waters 2690 separation module and a model 410 differential refractometer. Five Waters Styragel HR columns (HR5.0, HR4.0, HR3.0, HR1.0, and HR0.5) in series were used at a constant temperature of 40 °C. Distilled THF was used as eluent at a flow rate of 1.0 mL min^{-1}. The LSEC system was calibrated using narrow molecular weight poly(styrene) standards ranging from 374 to 400 × 10^{3} g mol^{-1}. The Mark-Houwink-Kuhn-Sakurada parameters used for the poly(styrene) standards were K = 1.14 × 10^{-4} dL g^{-1} and a = 0.716.[30] Similarly, for poly(methyl acrylate), K = 19.5 × 10^{-5} dL g^{-1} and a = 0.660.[31]

4.2.3.2 Nuclear Magnetic Resonance (NMR)

^{1}H-NMR and Heteronuclear Single-Quantum Correlation (HSQC) spectroscopy were performed on a 400 MHz Bruker Avance instrument (Bruker BioSpin GmbH) in $CDCl_3$, DMSO-d6, or D_2O. A total of 32 and 8 scans per spectrum were recorded for the ^{1}H and HSQC spectra,

respectively. For ^{13}C-NMR, 3000 scans and a relaxation delay of 3 seconds were used to enhance the intensity of tertiary carbon centres. Conversion was estimated by comparing the integral of the vinyl bonds from approximately 6.1 and 5.2 ppm to the constant amount of methoxy protons (of MA and pMA) between approximately 3.5 and 2.9 ppm. Samples were withdrawn at recorded intervals from the polymerization flask with a degassed syringe, then injected into an NMR tube and frozen in liquid nitrogen (to stop the polymerization) prior to ^{1}H-NMR analysis.

4.2.3.3 Attenuated Total Reflectance Fourier Transform Infrared (ATR FT-IR) Spectroscopy

Fourier Transform Infrared (FT-IR) spectroscopy was performed on a Thermo Scientific Nicolet 6700 instrument using an Attenuated Total Reflectance (ATR) accessory equipped with a diamond crystal. A total of 64 scans were co-added per spectrum with a background spectrum of air taken between successive samples. Constant pressure was applied to the sample on the diamond crystal using a hand-tightened anvil.

4.3 Results and Discussion

4.3.1 Activity of amides, amines, and hydrazides: 2-bromoacetamide model experiment

In order to introduce an alkyl halide initiating moiety at the reducing-end of polysaccharides, reductive amination (or amidation) has typically been performed using bifunctional alkyl halide initiators. A model NMR experiment was first conducted to assess the rapidity and ease to which 2-bromoacetamide (i.e., an amide) or isoniazid (i.e., a hydrazide) could be attached to cellobiose (Table 4.1). Cellobiose is a β(1-4) linked anhydroglucose oligomer consisting of two units. Higher molecular weight anhydroglucose polysaccharides, like cellulose,

have a lower quantity of reducing-ends for the same mass of polymer, resulting in poor resolution of the reducing-end signals in ^{1}H-NMR.

Table 4.1. Summary of NMR experiments conducted with 2-bromoacetamide (BA) and isoniazid (a hydrazide).

Entry	**Stoichiometry (NU:RA:CB)**	**Duration**	**$\frac{\text{mol } \beta\text{-OH}}{\text{mol } \alpha\text{-OH}}$ (%)***	**Comments**
BA1	1:1:1	24 hours	18	10 wt.% CB in water
BA2	2:2:1	24 hours	55	10 wt.% CB in water
BA3	1:1:1	7 days	80	10 wt.% CB in water
Isoniazid	1:0:1	24 hours	83	5 wt.% CB in water

All reactions conducted at 70 °C and magnetic stirring. NU: 2-bromoacetamide or isoniazid, BA: bromoacetamide, RA: $NaBH_3CN$, CB: cellobiose.

** Estimated via ^{1}H-NMR as a percentage of the reducing end β-hydroxyl substitution compare to α-hydroxyl substitution (integral β divided by integral α); due to the anomeric effect, the amine or hydrazide attached to the reducing end of cellobiose will prefer the axial position, thus increasing the ratio of remaining beta/alpha (equatorial/axial) hydroxyls.*

Quantitative conversions of cellobiose reducing-ends to cellobiose-amide and cellobiose-hydrazide conjugates were observed *via* ^{1}H-NMR by monitoring the reducing-end protons[32] of cellobiose, equatorial (β) hydroxyl at δ = 6.7 ppm and axial (α) hydroxyl at δ = 6.3 ppm as shown in Figure 4.3.

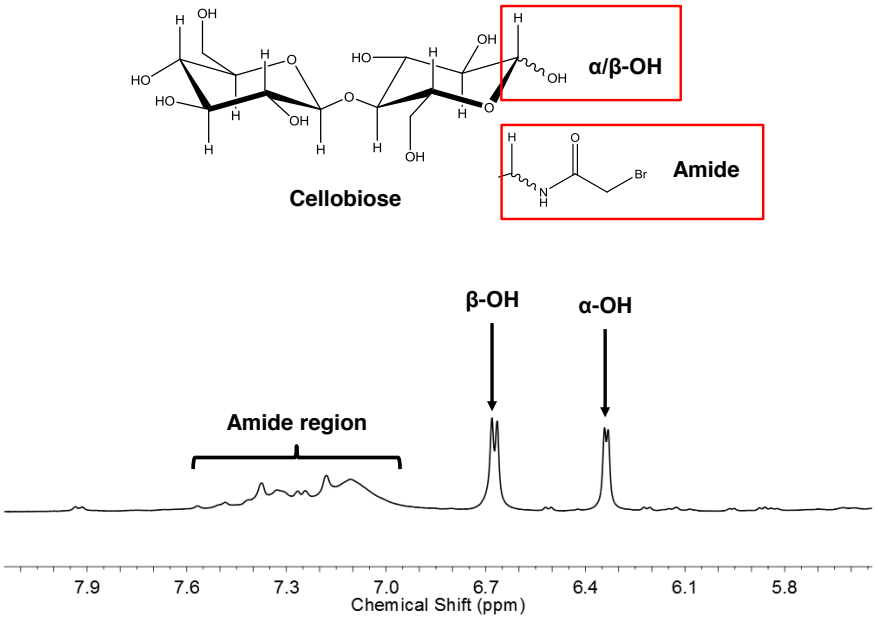
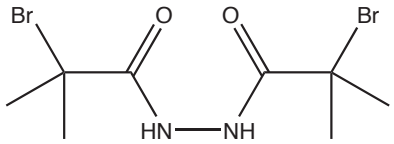


Figure 4.3. Chemical structure and representative ^{1}H-NMR spectrum of cellobiose and its 2-bromoacetamide derivative. Conversion of the reducing end hydroxyl groups to amide derivatives can be evidenced by the decreased intensity of the α hydroxyl substituent.

According to the "reverse" anomeric effect, the equatorial (β) hydroxyl substituent is more likely to be substituted by a bulky or positively charged amine or hydrazide adduct due to electronic or steric effects.[33] However, Perrin[33] showed that this was not the case and that most amine-containing substituents still preferred the axial (α) position only insofar as the effect of steric hindrance forces them to migrate towards equatorial. This conclusion was confirmed by early experiments where the chemical shift due to the α hydroxyl markedly decreased due to substitution with 2-bromoacetamide or isoniazid. As expected, isoniazid was observed to react much more readily than 2-bromoacetamide, even in the absence of reducing agent. These preliminary experiments provided the motivation to explore a hydrazide-containing Cu(0)-mediated LRP initiator.

4.3.2 Initiator synthesis

Several routes were attempted towards the synthesis of a nucleophilic initiator for Cu(0)-mediated LRP (see Table 4.2). The first route attempted to substitute the ethoxy, methoxy, or acid bromide functionality of EBiB, MBiP, and BriB with anhydrous hydrazine. Due to highly reactive nature of these reagents, extra caution was exercised while performing these experiments. The substitution of the ethoxy group of EBiB or the methoxy group of MBiP with hydrazine at low temperatures showed no formation of the desired product. The reaction of EBiB and MBiP with hydrazine at higher temperatures (e.g. 30 to 70 °C) also failed due to the preferential substitution of the alkyl bromide with hydrazine. Syntheses involving BriB and hydrazine tended to fail because of the uncontrolled formation of the disubstituted compound 2-bromo-*N*′-(2-bromo-2-methylpropanoyl)-2-methylpropane hydrazide (Scheme 4.2).

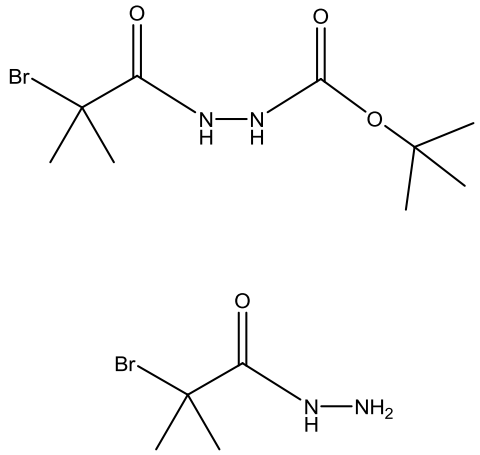


Scheme 4.2. Undesired disubstituted compound 2-bromo-*N*′-(2-bromo-2-methylpropanoyl)-2-methylpropane hydrazide.

To mitigate the formation of this disubstituted product, molar equivalents of hydrazine to BriB were used, in addition to reducing the reaction temperatures using ice or frozen EtOH. Despite these efforts, the formation of the desired hydrazide-based initiator remained difficult using BriB and hydrazine.

The use of *tert*-butyl carbazate (BOC-hydrazine) was ultimately chosen due to its high conversions and safe handling. Reactions involving molar equivalents of BOC-hydrazine with

BriB gave excellent yields of the BOC-protected initiator, *tert*-butyl 2-(2-bromo-2-methylpropanoyl)hydrazine carboxylate (Scheme 4.3)



Scheme 4.3. Chemical structure of the BOC protected initiator (*tert*-butyl 2-(2-bromo-2-methylpropanoyl)hydrazine carboxylate, top) and the desired hydrazide-containing initiator (2-bromo-2-methylpropane hydrazide, bottom).

Typical ^{1}H-NMR and ^{13}C-NMR spectra of the BOC-protected initiator are shown in Figure 4.4 and Figure 4.5, respectively.

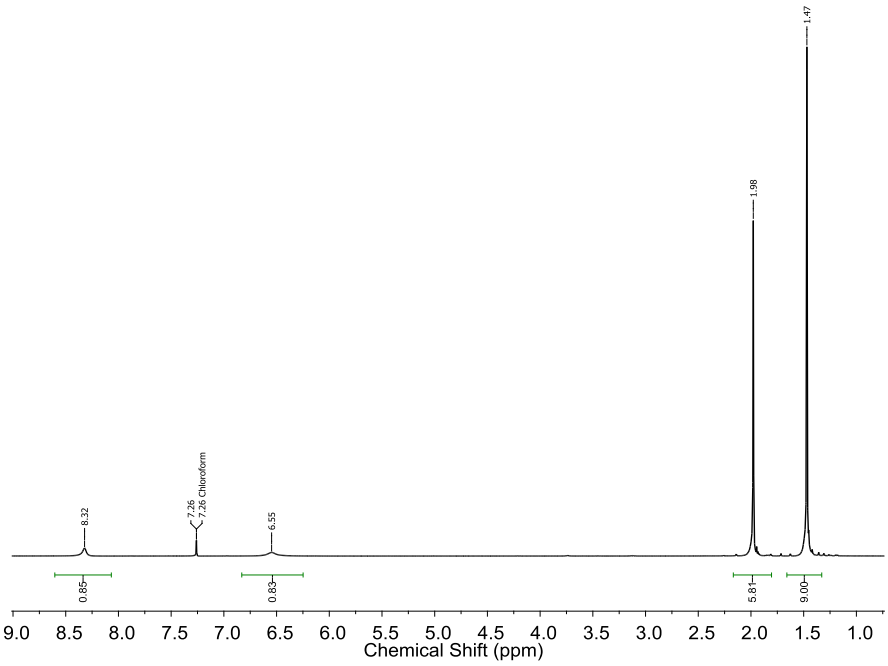


Figure 4.4. ^{1}H-NMR (400 MHz, $CDCl_3$) spectrum of the BOC-protected LRP initiator, *tert*-butyl 2-(2-bromo-2-methylpropanoyl)hydrazine carboxylate synthesized using the IS17 method.

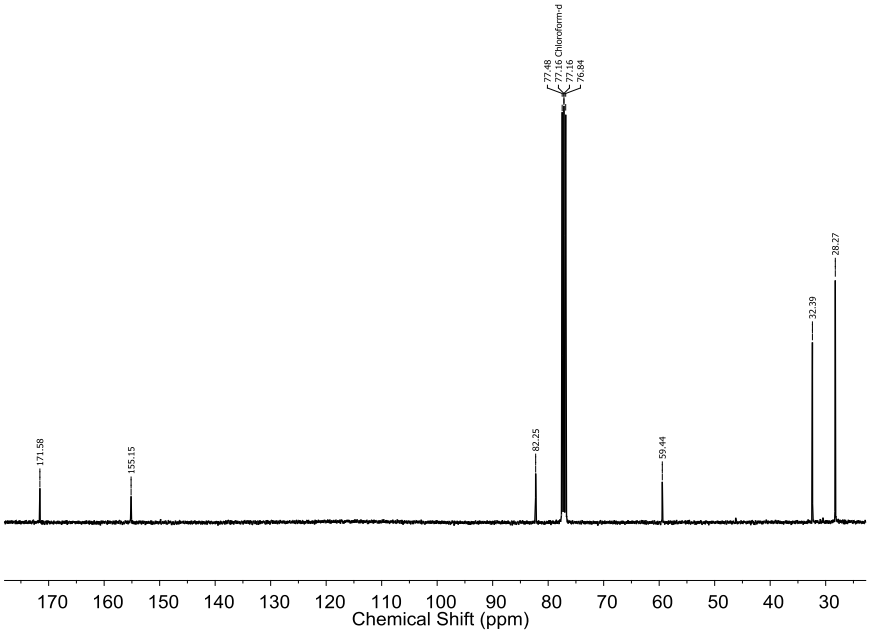


Figure 4.5. ^{13}C-NMR (400 MHz, $CDCl_3$) spectrum of the BOC-protected LRP initiator, *tert*-butyl 2-(2-bromo-2-methylpropanoyl) hydrazine carboxylate synthesized using the IS17 method.

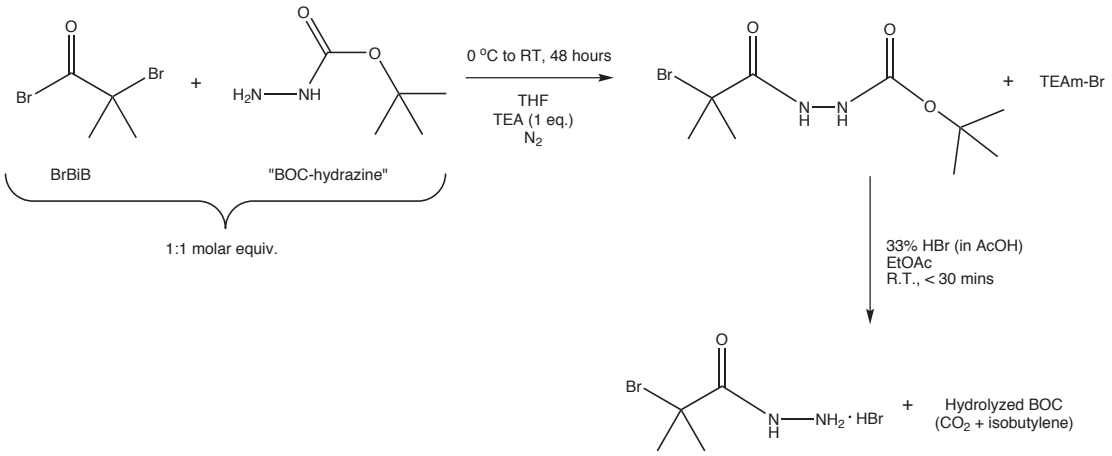
In addition to a high yield of around 90%, excellent purity of the BOC-protected initiator was attained, as evidenced by the integral of the methoxy region at 1.94 ppm in Figure 4.4. This methoxy peak corresponds to the (-CH_3)$_2$ protons of the alkyl bromide, with an integral of 5.81 compared to 6.00 for a 100% purity (~97% purity).

Table 4.2. List of the various initiator syntheses attempted in this study. Entry IS17 was chosen as the most appropriate route towards the hydrazine-containing LRP initiator.

Entry	Alkyl bromide	Nucleophile	Stoichiometry	Conditions	Comments
IS1	EBiB	Hydrazine	1:1	EtOH 0 °C to RT 24 hours	No reaction
IS2	EBiB	Hydrazine	1:1	MeOH 0 °C to RT 24 hours	No reaction
IS3	BriB	Hydrazine	1:1	MeOH 0 °C to RT 24 hours	Uncontrolled reaction of BriB with MeOH gave majority of undesired product
IS4a	EBiB	Hydrazine	1:1	MeOH 30 °C 24 hours	Gave yellow solution, preferential reaction with alkyl bromide
IS4b	EBiB	Hydrazine	1:1	EtOH 30 °C 24 hours	Gave yellow solution, preferential reaction with alkyl bromide, precipitate formed, likely hydrazine bromide salt
IS5	EBiB	Hydrazine	1:1	MeOH/water (2:1) 30 °C 24 hours	Small amount of product formed, evidenced by carbonyl region in ^{13}C-NMR
IS6	EBiB	Hydrazine	1:1	Water 0 °C to RT 24 hours	No methyl peaks indicative of alkyl bromide; preferential substitution at alkyl bromide with hydrazine
IS7a	EBiB	Hydrazine	1:1	EtOH RT to 30 °C 24 hours	Broad peak due to hydrazine salt formation, alkyl bromide likely substituted
IS7b	EBiB	Hydrazine	1:1	iPrOH RT to 30 °C 24 hours	Precipitate formed, broad peak due to hydrazine salt formation, alkyl bromide likely substituted
IS8	MBiP	Hydrazine	1:1	EtOH RT 24 hours	Large amount of white precipitate formed in flask of hydrazine salt
IS9	MBiP	Hydrazine	1:1	EtOH RT 24 hours	Larger scale duplicate of IS8, large amount of white precipitate formed in flask of hydrazine salt
IS10	EBiB	Hydrazine	1:1	THF	No reaction, even after 3 days

				RT 24 hours	
IS11	BriB	Hydrazine	1:1	THF, TEA 0 °C to RT 30 hours	Limited solubility of hydrazine in THF, uncontrolled amount of disubstituted produced, multiple methyl proton environments
IS12	BriB	Hydrazine	1:1	THF, TEA 0 °C to RT 24 hours	Larger scale duplicate of IS11, limited solubility of hydrazine in THF, uncontrolled amount of disubstituted produced, multiple methyl proton environments
IS13	BriB	Hydrazine	1:1	THF 0 °C to RT 24 hours	Duplicate of IS12 without TEA, salt formed with amine of initiator, limited solubility of hydrazine in THF, uncontrolled amount of disubstituted produced, multiple methyl proton environments
IS14	EBiB	Hydrazine	1:1	THF Reflux at 70 °C under N_2 24 hours	Duplicate of IS13 but under reflux at 70 C, definite substitution of alkyl bromide with hydrazine and/or formation of hydrazine bromide salt
IS15	EBiB	Hydrazine	1:1	THF Reflux at 70 °C under N_2 24 hours	Larger scale duplicate of IS14, definite substitution of alkyl bromide with hydrazine and/or formation of hydrazine bromide salt
IS16	BriB	BOC-hydrazine	1:1	THF, TEA 0 °C to RT 24 hours	Quantitative formation of BOC-protected initiator formed under controlled conditions, retention of alkyl bromide functionality
IS17	BriB	BOC-hydrazine	1:1	THF, TEA 0 °C to RT 24 hours	Duplicated of IS16, quantitative formation of BOC-protected initiator formed under controlled conditions, retention of alkyl bromide functionality
IS19	BriB	Hydrazine	1:1	THF, TEA 0 °C to RT 3 hours	Limited solubility of hydrazine in THF, uncontrolled amount of disubstituted produced reduced by short reaction time, multiple methyl proton environments, product oil was cloudy when dissolved in $CDCl_3$ – suggests formation of desired product
IS20	BriB	Hydrazine	1:1	CH_2Cl_2, TEA 0 °C to RT 24 hours	THF added to precipitate TEA-salt from solution, rotavap of solution gave very viscous brown product (more so than IS19), large distribution of methyl peaks due to uncontrolled reaction with dichloromethane
IS21	BriB	Ethylene-diamine	1:1	THF, TEA 0 °C to RT 20 hours	Uncontrolled formation of disubstituted product

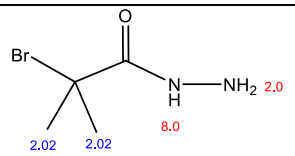
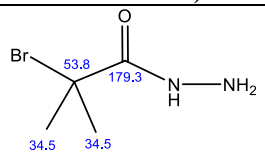
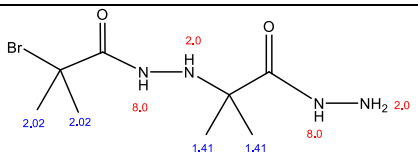
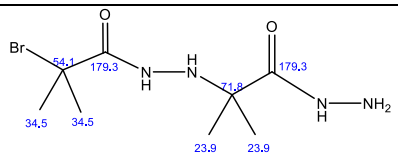
The BOC-protected initiator required a strong acid for deprotection. Various deprotection acids were attempted, including hydrochloric, phosphoric, and hydrobromic acid. Phosphoric acid was used as a deprotection acid because of its relatively mild nature and specificity towards BOC hydrolysis.[34] Unfortunately, the product recovery step was difficult since the aqueous system (85 wt.% aqueous H_3PO_4) supported the dissolution of the initiator in its salt form, and since aqueous phosphoric acid has a low vapour pressure and cannot be readily removed *in vacuo*. Hydrochloric acid was also tried as the deprotection acid because it was readily available in the lab, but concerns over the retention of the labile bromide functionality in the presence of chloride anions caused the migration to hydrobromic acid. Hydrobromic acid (33 wt.% HBr in acetic acid) was therefore used since it cannot introduce an alkyl chloride functionality and will yield the initiator in its HBr salt form as a white precipitate in ethyl acetate, along with the liberation of CO_2 and isobutylene from the hydrolyzed BOC group. This approach using HBr solution to deprotect the initiator was borrowed from Houga et al.[26]. After deprotecting, the hydrazide initiator in its HBr form was dried over a Teflon filter membrane. The final synthesis protocol for the novel hydrazide-containing initiator 2-bromo-2-methylpropanehydrazide HBr is summarized in Scheme 4.4.



Scheme 4.4. Synthesis protocol for the novel hydrazide-containing initiator 2-bromo-2-methylpropane hydrazide in its HBr salt form.

On one occasion, the initiator turned a deep orange colour upon contact with a stainless steel spatula (possibly carrying some moisture from the humid air in the laboratory). ^{1}H-NMR of the orange substance confirmed that the alkyl halide bond was breaking, thus releasing free bromine giving the otherwise white precipitate an orange colour. In this case, the alkyl halide bond was likely reacted upon by the hydrazide functionality of a different initiator molecule (see Table 4.3).

Table 4.3. Structures and NMR chemical shift predictions of the desired and undesired products observed following the deprotection of the BOC-protected initiator.

Compound	Structure (^{1}H-NMR Prediction)	Structure (^{13}C-NMR Prediction)
Desired product		
Major undesired side-product		

Note: The numbers next to protons and carbons are the ChemDraw predicted ^{1}H-NMR and ^{13}C-NMR chemical shifts, respectively. Predictions in blue correspond to a good estimate while red corresponds to a rough estimate; chemical shifts associated with amines are more difficult to predict.

The chemical shifts at 1.4 ppm in the ^{1}H-NMR spectrum of the hydrazide-containing initiator in D_2O (see Figure 4.6) can be assigned to the methyl protons adjacent to the tertiary carbon that underwent substitution of bromine with a hydrazide from a different initiator molecule. The ^{13}C-NMR spectrum of the initiator confirms the presence of these molecules that underwent inter-initiator coupling, where a chemical shift corresponding to the hydrazide-substituted tertiary carbon centre appears at 73 ppm. The methyl groups of this major side-product can also be observed in the ^{13}C-NMR spectrum at 26 ppm and likewise the carbonyl of this undesired product at 177 ppm. In an effort to determine whether this unwanted side-reaction could be avoided, an NMR experiment was conducted whereby the BOC-protected initiator was first dissolved in MeOD and then deprotected using *p*-toluenesulfonic acid in an NMR tube. In this manner, the effect of solvent and acid could be elucidated. Unfortunately, the same inter-initiator coupling pattern was observed in both the ^{1}H and ^{13}C-NMR spectra of this deprotection reaction using *p*-toluenesulfonic acid. A hydrazide-substituted tertiary carbon centre was observed at 74 ppm and a new set of methyl environments between 1.5 and 1.0 ppm. This NMR experiment suggests that the inter-initiator coupling between the hydrazide functionality of one initiator molecule and the alkyl bromide of a neighbouring initiator molecule occurs independent of the chosen deprotection acid or solvent. The recovered HBr salt of 2-bromo-2-methylpropane hydrazide was readily soluble in D_2O, giving the ^{1}H-NMR and ^{13}C-NMR spectra shown in Figure 4.6 and Figure 4.7, respectively.

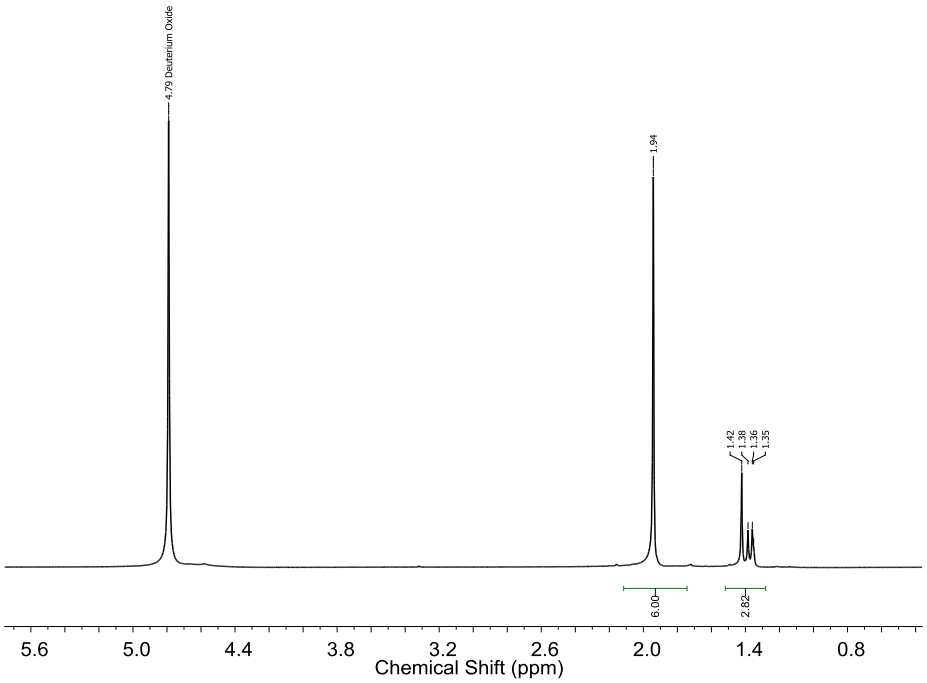


Figure 4.6. ^{1}H-NMR (400 MHz, D_2O) of the hydrazine-containing LRP initiator, 2-bromo-2-methylpropane hydrazide. Two methyl proton environments were evidenced, -CBr-(CH_3)$_2$ at 1.9 ppm and -NH-CH-(CH_3)$_2$ at 1.4 ppm.

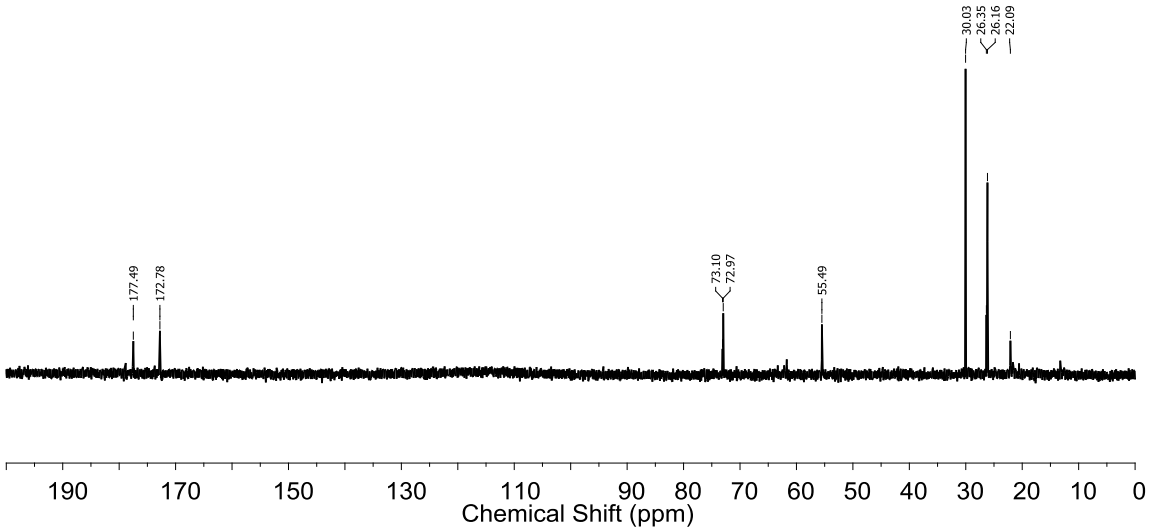


Figure 4.7. ^{13}C-NMR (400 MHz, D_2O) of the hydrazine-containing LRP initiator, 2-bromo-2-methylpropane hydrazide showing chemical shifts of both the desired and undesired products following deprotection. See experimental section for assignments.

Despite the aforementioned stability issues, 2-bromo-2-methylpropane hydrazide appeared (*via* NMR) to retain some of its alkyl bromide functionality following deprotection with HBr solution.

Infrared spectroscopy was also used to observe the fundamental vibration modes of the compounds described above. ATR FT-IR spectra of the BOC-protected initiator, deprotected initiator, reducing-end initiated dextran, and virgin dextran are shown in Figure 4.8.

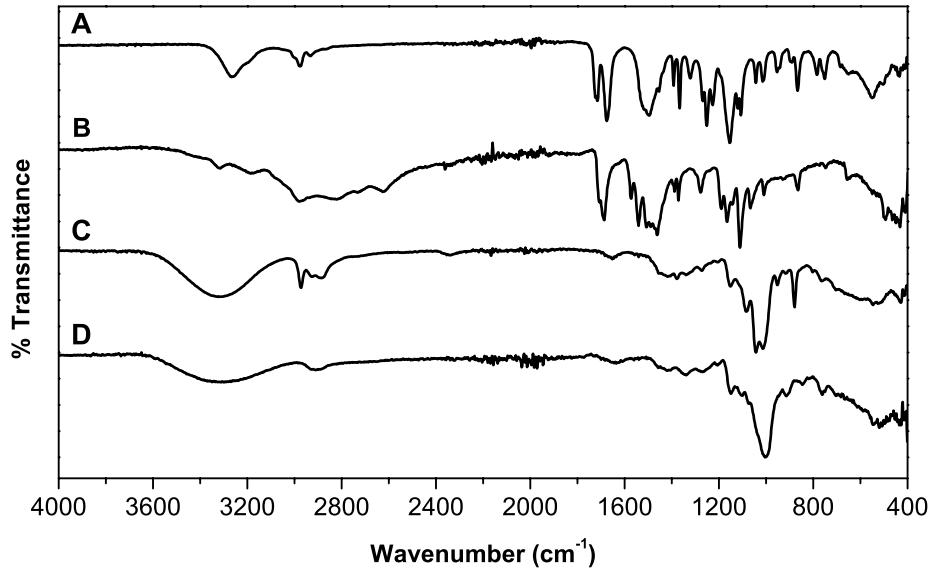


Figure 4.8. ATR FT-IR spectra of A: the BOC-protected initiator, B: the deprotected initiator in its HBr salt form, C: dextran initiated at its reducing-end with the deprotected initiator, and D: virgin dextran.

The ATR FT-IR spectrum of the BOC-protected initiator (Figure 4.8, spectrum A) shows two carbonyl vibration modes around 1700 cm^{-1} that collapse into a single carbonyl vibration mode (albeit with a shoulder peak to the left of the major carbonyl peak due to the inter-initiator coupling) after the initiator was deprotected (Figure 4.8, spectrum B). A broad C-H and N-H stretching mode can be observed for the BOC-protected initiator between 3400 and 2800 cm^{-1} (Figure 4.8, spectrum A) whereas for the deprotected initiator, a broad peak can be observed from

3500 to 2400 cm^{-1} due to the N-H stretching of the $-NH_3^+$ cation produced after treatment with the HBr solution. The alkyl bromide functionality can be observed for both the BOC-protected and deprotected initiator in the region spanning 690 to 515 cm^{-1}. For the reducing-end initiated dextran (Figure 4.8, spectrum C) and the virgin dextran (Figure 4.8, spectrum D), these alkyl bromide vibration modes are unfortunately obscured by the vibration modes of dextran. However, a strong aliphatic C-H stretching mode can be observed just under 3000 cm^{-1} in the spectrum for the reducing-end initiated dextran (Figure 4.8, spectrum C) whilst this strong peak is absent from the virgin dextran sample (Figure 4.8, spectrum D). This suggests the presence of additional alkane functionalities in the reducing-end initiated dextran, possibly due to the successful coupling of the initiator with dextran. In addition, a strong peak can be observed around 900 cm^{-1} for all spectra, except for virgin dextran, which is likely due to the C-N stretching mode of materials containing a hydrazide functionality. These ATR FT-IR results support the proposed structure of the hydrazide-containing initiator. Moreover, these data also suggest that a hydrazide functionality was successfully introduced at the reducing-end of dextran, although the retention of the alkyl halide functionality cannot be confirmed. The success of the reductive amination was further studied through ^{1}H-NMR experiments, as discussed in the following section.

4.3.3 Reductive amination

Reductive amination was carried out using the HBr salt of the hydrazide initiator shown in Scheme 4.4. Experimental details are provided in the experimental section above. In general, excess initiator and excess reducing agent ($NaBH_3CN$) were added to a solution containing organic base (triethylamine) and the polysaccharide (dextran). Organic base was needed since the initiator was in an acidic salt form. The progress of the reductive amination was monitored *via* ^{1}H-NMR by observing the disappearance of the anomeric proton signals, as shown in Figure 4.9.

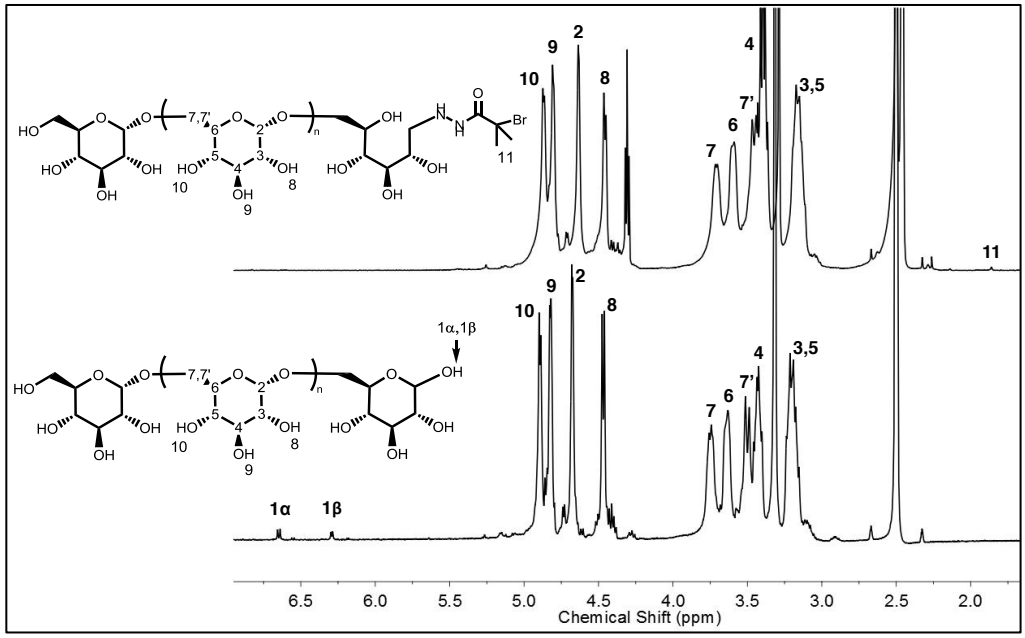


Figure 4.9. ^{1}H-NMR spectra of virgin dextran before (bottom) and dextran after (top) reductive amination with 2-bromo-2-methylpropane hydrazide.

The disappearance of the α and β hydroxyl protons at 6.6 and 6.3 ppm, respectively, can be observed from the ^{1}H-NMR spectrum following reductive amination. The ^{1}H-NMR of the reductively aminated dextran also shows slight evidence of alkyl halide methyl protons in the vicinity of 1.9 ppm. Because this methyl proton environment was similar to that observed for the initiator alone in D_2O, it was assumed that the hydrazide-based initiator was successfully attached to the reducing end of dextran. However, this methyl peak was much smaller than that of observed by Houga et al.[26], which suggests that while the reducing-end of dextran was successfully substituted, the attached compound had markedly less alkyl bromide character than the initiator prior to reductive amination. In the presence of strong reducing agents like NaH, hydrazide-like compounds have been found to undergo cyclization, which results in the substitution of the alkyl bromide with the terminal -NH_2 of the hydrazide (see compound 14, Scheme 4 in Greenhill et al.[35]). A similar reaction might occur with 2-bromo-2-

methylpropanehydrazide during reductive amination in the presence of $NaBH_3CN$. Unfortunately, further analysis has not yet been undertaken to provide evidence in favour or against this proposed side-reaction. The dextran-based initiators used in the subsequent Cu(0)-mediated LRP step were synthesized in the presence of $NaBH_3CN$.

In addition to the standard reductive amination conditions outlined above, an NMR experiment was conducted whereby 2-bromo-2-methylpropane hydrazide was attached to the reducing-end of dextran in the absence of reducing agent ($NaBH_3CN$). Under reducing-agent free conditions, the equatorial beta position has been shown to be preferred by hydrazide compounds.[36] The HSQC spectra of the reductive amination of dextran with 2-bromo-2-methylpropane hydrazide after 4 days in the presence and absence of $NaBH_3CN$ are shown in Figure 4.10.

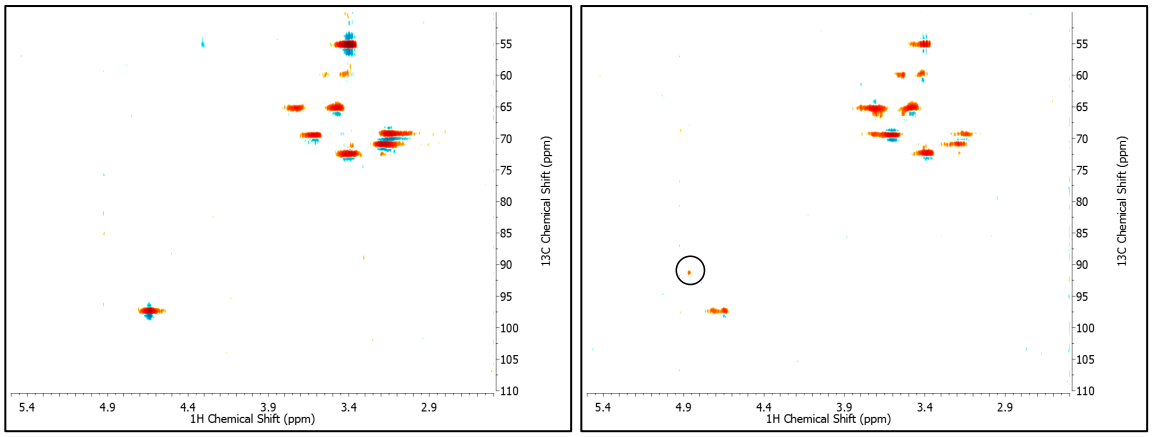


Figure 4.10. HSQC spectra of dextran initiated at its reducing-end with 2-bromo-2-methylpropane hydrazide conducted in the presence (left) and absence (right) of $NaBH_3CN$. Protons corresponding to the hydrazide-functionalized reducing-end in its closed-ring form are circled in the spectrum on the right.

For both experiments in Figure 4.10, the reducing-end -OH groups were absent in ^{1}H-NMR (data not shown), indicating quantitative coupling of the hydrazide group. It can be observed from the 2D HSQC spectra of the reductive amination of dextran with of 2-bromo-2-

methylpropanehydrazide that, in the absence of $NaBH_3CN$, a new proton environment at δ = {4.9, 91} ppm appears. This new proton environment can be attributed to the reducing-end proton of glucopyranose in the ring-closed form adjacent to the *N'* substituted hydrazide initiator (see Scheme 4.1). The fidelity of the alkyl bromide group following reductive amination in the absence of $NaBH_3CN$ was the subject of the next experiment.

An additional experiment was conducted involving the reductive amination of dextran (1 mol eq.) with 2-bromo-2-methylpropane hydrazide (5 mol eq.) in the absence of $NaBH_3CN$ and at room temperature. This reductive amination was monitored *via* ^{1}H-NMR, as shown in Figure 4.11.

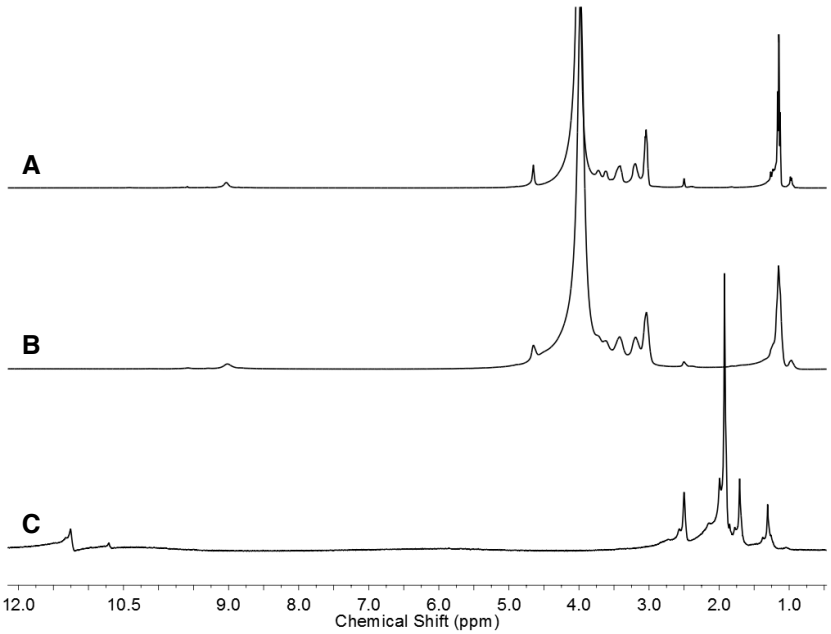


Figure 4.11. ^{1}H-NMR spectra showing the reductive amination of dextran with 2-bromo-2-methylpropane hydrazide in the absence of $NaBH_3CN$ and at room temperature. ^{1}H-NMR spectra are given, (400 MHz, DMSO-d6), A: 2-bromo-2-methylpropane hydrazide, triethylamine, dextran, DMSO-d6 after approximately 4 days of reaction time; (300 MHz, DMSO-d6), B: 2-bromo-2-methylpropane hydrazide, triethylamine, dextran, DMSO-d6 after approximately 2 hours of reaction time, C: 2-bromo-2-methylpropane hydrazide HBr.

After about 2 hours of reaction time, it was observed (Figure 4.11) that the methyl environment between 2.5 ppm and 1.5 ppm was absent after adding 2-bromo-2-methylpropane hydrazide to the solution containing dextran and triethylamine in DMSO-d6. After 4 days, the spectrum of the reductive amination was almost unchanged from the same reaction after just 2 hours. Because this experiment was conducted in DMSO-d6, it is possible that the reduction of the alkyl bromide occurred as outlined by Hutchins et al..[37] Furthermore, the hydrazide functionality could undergo substitution reaction with the alkyl halide, as described by Hendrickson and Sternback.[38] Regardless of the source of instability, the possibility remains that the reactivity of the hydrazide functionality could lead to numerous possible side-reactions involving the alkyl bromide moiety.

The reductive amination of dextran with 2-bromo-2-methylpropane hydrazide proved to be relatively successful, although the retention of the active alkyl bromide functionality could not be fully confirmed. Despite these troubles, Cu(0)-mediated LRP of this reducing-end initiated dextran was attempted, as will be discussed in the subsequent section.

4.3.4 Cu(0)-mediated Living Radical Polymerization (LRP)

4.3.4.1 Control experiment with ethyl α-bromoisobutryate (EBiB)

A control experiment was designed using a commercially available alkyl bromide initiator (EBiB) to mimic the conditions under which the dextran-based initiator would be polymerized. A commercially available initiator was chosen because the dextran-based initiator was not synthesized in high quantities. In this manner, optimization of the polymerization conditions could be expedited without requiring great quantities of the more expensive hydrazide-based initiator. Furthermore, Cu(0)-mediated LRP in DMSO using EBiB has been widely reported in literature.[16,39,17] The appropriate amount of solvent and monomer was determined by adding the dextran-based initiator (approx. 285 mg) to 2 mL of MA (the monomer) to give a

heterogeneous mixture. DMSO-d6 (the polymerization solvent) was then added until the solution became clear. The appropriate amount of DMSO-d6 was determined to be 5 mL. DMSO-d6 was used since samples were periodically withdrawn from the polymerization solution for conversion estimation using ^{1}H-NMR analysis (neat). Moreover, DMSO-d6 is much drier than regular DMSO because it comes in a sealed vial. Dry DMSO-d6 ensured that the Cu(0)-mediated LRP proceeded in the absence of large quantities of water, which also eliminated the broad chemical shift of water at 3.3 ppm in the ^{1}H-NMR spectra. The final concentration of initiator in both experiments containing EBiB as the initiator and the reducing-end initiated dextran were approximately 0.0121 mol·L^{-1}. Full experimental details are given in the corresponding section above. It should be noted that a similar polymerization was attempted in the absence of Cu(II)Br_2, but the reaction terminated at low conversions (~42%) due to the lack of deactivating species that would otherwise control the polymerization.

The molecular weight distributions (MWDs) of the Cu(0)-mediated LRP of EBiB in the presence of initial quantities Cu(II)Br_2 are shown in Figure 4.12.

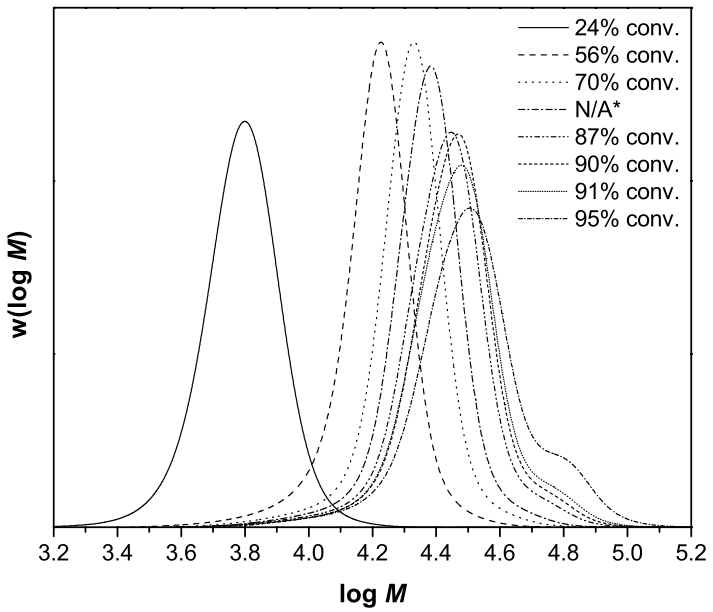


Figure 4.12. Progression of the MWD with respect to conversion for the EBiB initiated Cu(0)-mediated LRP of MA. *N/A: The NMR tube containing sample 4 broke, so conversion data was not determined for this sample.

The MWD of the polymer initiated by EBiB increased with conversion up until about 87% where termination began to manifest as a higher molecular weight shoulder. Figure 4.13 shows the M_n and dispersity (M_w / M_n) calculated *via* GPC as a function of conversion.

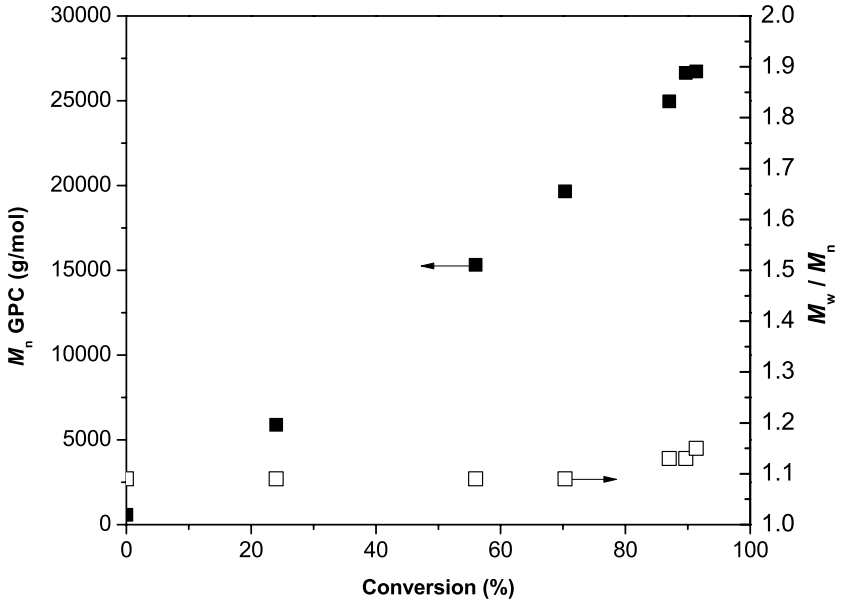


Figure 4.13. M_n and M_w / M_n as a function of conversion for the EBiB initiated Cu(0)-mediated LRP of MA.

For this Cu(0)-mediated LRP with EBiB, the dispersity remained low and the M_n grew linearly with conversion. Good control was observed since the dispersity remained well below 1.2 for the duration of the polymerization. The M_n and dispersity of the polymer at approximately 91% conversion was 2.67×10^4 g·mol^{-1} and 1.15, respectively. Significant termination was evident after about 66 minutes from the point of 91% conversion where the conversion only increased to 95%. The final M_n and dispersity of the polymer at 95% conversion was 2.92×10^4 g·mol^{-1} and 1.20, respectively. The Cu(II)Br_2 added at the beginning of the polymerization appeared to improve control, which mitigated termination until very high conversions of about 90% were reached. A plot describing the kinetics of the polymerization is shown in Figure 4.14.

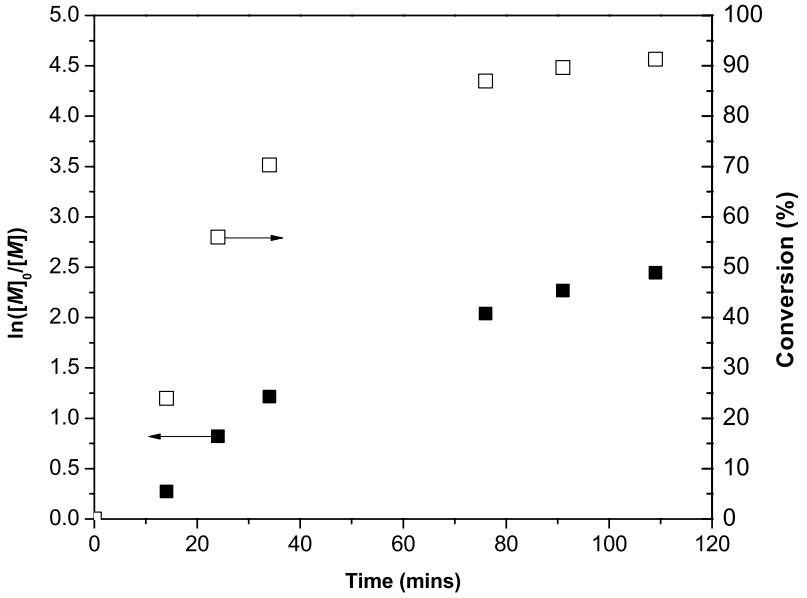


Figure 4.14. Semi-logarithmic plot of ln($[M]_0/[M]$) as a function of time. Conversion as a function of time is also shown on the secondary axis. The last data point was omitted due to significant effect of termination on the concentration of radicals at 95% conversion.

The polymerization was observed to reach over 90% conversion in less than 2 hours. Finally, the initiator efficiency (I_{eff}) was estimated to be 75% by dividing the final M_n, calculated by GPC, with the theoretical final M_n, estimated by $[M]_0/[I]$ where $[M]_0$ and $[I]$ are the initial monomer (MA) and initiator (EBiB) concentrations, respectively. This comparison of GPC M_n and theoretical M_n are summarized in Figure 4.15.

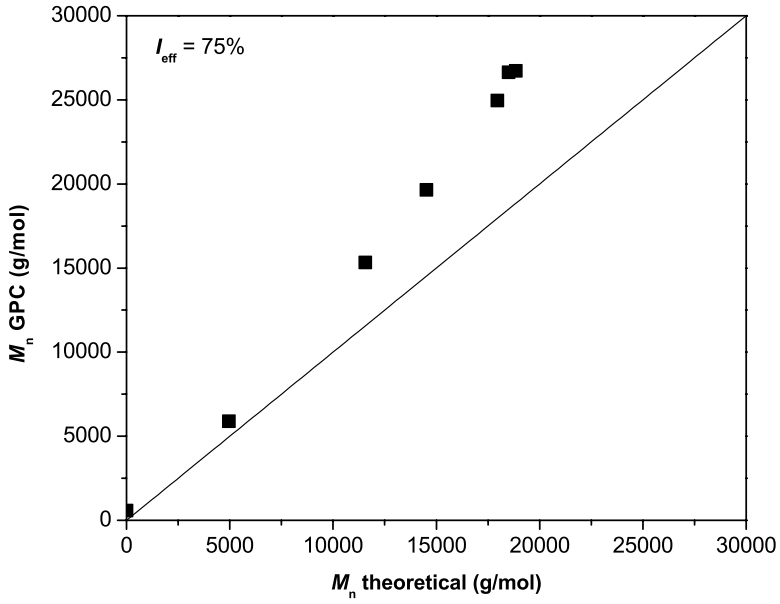


Figure 4.15. Comparison of M_n estimated by GPC and theoretical M_n estimated by the ratio $[M]_0/[I]$. The line corresponds to perfect agreement between both M_n GPC and M_n theoretical.

The M_n was observed to follow the theoretical M_n quite closely at low molecular weights, but deviated as M_n increased. A reasonable initiator efficiency of 75% was observed. This preliminary control experiment involving low concentrations of EBiB showed good control and livingness, even at high conversions. Motivated by these results, the reducing-end initiated dextran was used instead of EBiB, at the same concentration and under identical reaction conditions.

4.3.4.2 Attempted polymerization with reducing-end initiated dextran

The same procedure that was used for EBiB was used for the reducing-end initiated dextran. No appreciable conversion of monomer to polymer was detected over the course of 5 hours. A duplicate experiment was conducted that also gave no quantifiable amount of polymer. However, a visible amount of precipitate formed after leaving the reaction mixture for over 3 days exposed to air. The copper wire was left in the polymerization solution during this extended period of time and it was noticed that a corona of precipitate had formed on the surface of the

copper wire, which appeared to be polymer. This substance was insoluble in DMSO, DMF/LiBr, and THF. A sample was taken for ^{1}H-NMR in DMSO-d6 whereby a dark green precipitate settled to the bottom of the NMR tube. In $CDCl_3$, a cloudy mixture formed and showed evidence of polymeric methoxy protons as well as backbone CH_2 protons, as shown in Figure 4.16.

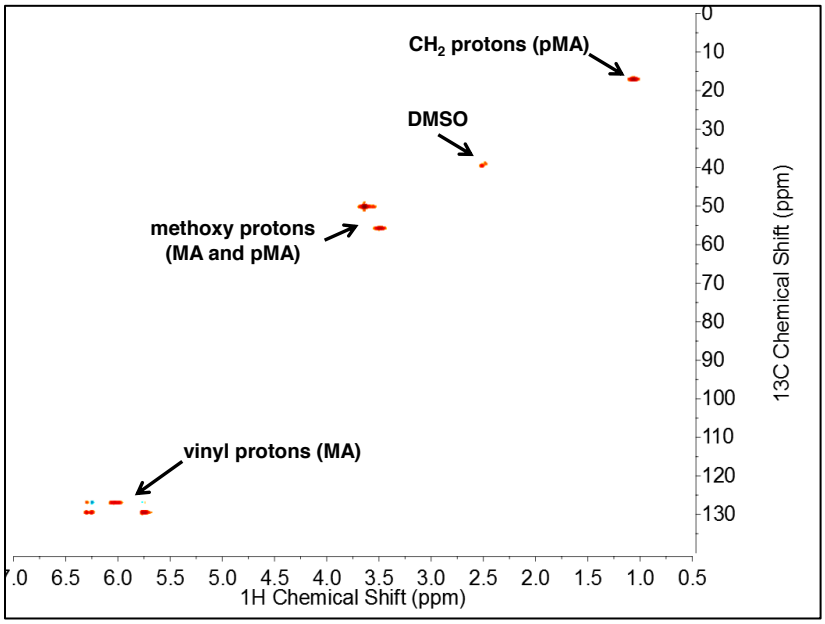


Figure 4.16. HSQC (400 MHz, $CDCl_3$) spectrum of the precipitate recovered from polymerization containing the reducing-end initiated dextran after leaving the flask open to air for several days.

While the above NMR spectrum of the recovered precipitate from the polymerization showed evidence of pMA polymer, it cannot be concluded whether or not this pMA was formed as a copolymer with the reducing-end initiated dextran or if it was simply due to autopolymerization of the uninhibited MA under quiescent conditions. Because this material was insoluble in DMF/LiBr or THF, SEC analysis could not be performed.

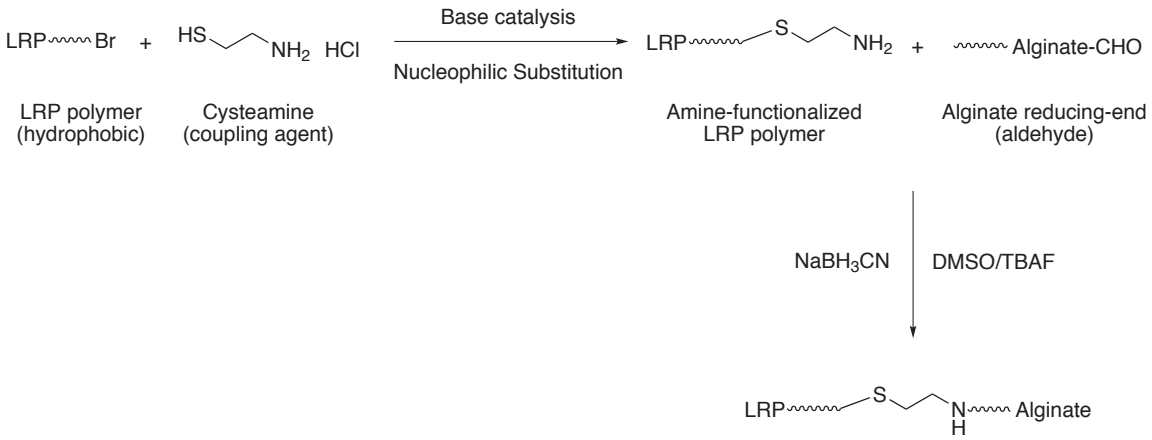
4.4 Conclusion and Recommendations

The attempted synthesis of a novel hydrazide containing LRP initiator, 2-bromo-2-methylpropane hydrazide was ultimately unsuccessful due to the absence of an alkyl bromide moiety required for initiating LRP. Although the resulting hydrazide could not initiate a polymerization, it was found to react quite effectively at the reducing end of dextran. Quantitative yields were observed for the reaction between the hydrazide and the reducing end, but the presence of the terminal alkyl bromide moiety was not evidenced. Cu(0)-mediated LRP of MA was successfully conducted using relatively low concentrations of the commercially available initiator EBiB. This LRP with EBiB showed controlled/living character towards high conversions (~90%), giving low dispersity polymers in under two hours. The same polymerization conditions were used for the dextran-based initiator; however, no conversion was observed and the polymerization was deemed to be a failure. A precipitate was observed in the Schlenk flask and around the copper wire after letting the polymerization solution stand for days open to air. NMR analysis of this material showed protons reminiscent of pMA, but whether this pMA was generated by the 2-bromo-2-methylpropane hydrazide moiety attached to the reducing end of dextran or whether this pMA was simply an artifact of autopolymerization is still open to debate.

Despite the inconclusive nature of the Cu(0)-mediated LRP experiment, the hydrazide-containing LRP initiator developed herein still holds potential to expedite the reductive amination step, which currently hinders the facile production of polysaccharide-containing block copolymers. A better understanding of the kinetics of dextran-based initiator system would help to design a successful Cu(0)-mediated polymerization experiment. The activation kinetics of this new initiator system could be elucidated by following the methods of Tang et al..[40] Once the appropriate k_{act} value of the reducing-end initiated dextran has been determined in relation to EBiB, a controlled Cu(0)-mediated LRP could potentially be designed. Simulations based on

different polymerization conditions and recipes could also be implemented using software such as PREDICI® (Wulkow CiT GmbH).

An alternative approach could be also implemented to produce the desired polysaccharide-containing amphiphiles. Linear synthetic polymers produced using Cu(0)-mediated LRP (or any other polymerization technique that produces electrophile-terminated polymers) can be modified under base catalysis with cysteamine hydrochloride to give an amine terminated synthetic polymer, as shown in Scheme 4.5.



Scheme 4.5. Alternative route towards the synthesis of block copolymers containing synthetic and polysaccharide segments. Cysteamine hydrochloride is used as a coupling agent between the synthetic and polysaccharide blocks in a grafting approach.

This "grafting" approach would use the commercially available compound cysteamine hydrochloride, thus eliminating specialized initiators. Possible weaknesses of this grafting approach, however, are the slow reaction kinetics between the chain-end of two large polymers and the requirement of DMSO/TBAF in the last step, which might limit the solubility of some synthetic polymers. Despite these weaknesses, the alterative approach outlined above is generally more facile than grafting synthetic polymers from the reducing-end of a polysaccharide and could produce a library of amphiphilic copolymers with tuneable self-assembly behaviour.

4.5 Acknowledgements

The authors would like to thank Dr. Ralph Whitney for his guidance and Nicky Chan for synthesizing the ligand (Me_6TREN) and for his advice on conducting Cu(0)-mediated LRP. The authors would like to acknowledge the financial support provided by the Government of Ontario through the Ontario Research Chair Program (Cunningham). The authors would also like to acknowledge the Ministry of Research and Innovation for their support through the Early Researcher Award Program (Champagne) and Queen's University for their support by providing a Queen's Graduate Award and facilitating a G.E. Ted Courtnage Graduate Award in Engineering (Krasznai). The authors would like to thank the Courtnage family for establishing this award.

4.6 References

(1) Roy, D.; Semsarilar, M.; Guthrie, J. T.; Perrier, S. *Chem. Soc. Rev.* **2009**, *38*, 2046–2064.
(2) Yagi, S.; Kasuya, N.; Fukuda, K. *Polym. J.* **2010**, *42*, 342–348.
(3) Tizzotti, M.; Charlot, A.; Fleury, E.; Stenzel, M.; Bernard, J. *Macromol. Rapid Commun.* **2010**, *31*, 1751–1772.
(4) Malmström, E.; Carlmark, A. *Polym. Chem.* **2011**, DOI: 10.1039/c1py00445j.
(5) Schatz, C.; Lecommandoux, S. *Macromol. Rapid Commun.* **2010**, *31*, 1664–1684.
(6) Grzybowski, B. A.; Wilmer, C. E.; Kim, J.; Browne, K. P.; Bishop, K. J. M. *Soft Matter* **2009**, *5*, 1110.
(7) Francis, M.; Cristea, M. *Pure and applied chemistry* **2004**, *76*, 1321–1335.
(8) Lemarchand, C.; Gref, R.; Couvreur, P. *European Journal of Pharmaceutics and Biopharmaceutics* **2004**, *58*, 327–341.
(9) Beaty, N. B.; J Mello, R. *Journal of Chromatography B: Biomedical Sciences and Applications* **1987**, *418*, 187–222.
(10) Liu, Z.; Jiao, Y.; Wang, Y.; Zhou, C.; Zhang, Z. *Advanced Drug Delivery Reviews* **2008**, *60*, 1650–1662.
(11) Hoare, T. R.; Kohane, D. S. *Polymer* **2008**, *49*, 1993–2007.
(12) Kumar, J.; Bousquet, A.; Stenzel, M. H. *Macromol. Rapid Commun.* **2011**, DOI: 10.1002/marc.201100331.
(13) Wang, J.-S.; Matyjaszewski, K. *Macromolecules* **1995**, *28*, 7572–7573.
(14) Narumi, A.; Kakuchi, T. *Polym. J.* **2008**, *40*, 383–397.
(15) Spain, S. G.; Gibson, M. I.; Cameron, N. R. *J. Polym. Sci. A Polym. Chem.* **2007**, *45*, 2059–2072.
(16) Percec, V.; Guliashvili, T.; Ladislaw, J. S.; Wistrand, A.; Stjerndahl, A.; Sienkowska, M. J.; Monteiro, M. J.; Sahoo, S. *Journal of the American Chemical Society* **2006**, *128*, 14156–14165.

(17) Nyström, F.; Soeriyadi, A. H.; Boyer, C.; Zetterlund, P. B.; Whittaker, M. R. *J. Polym. Sci. A Polym. Chem.* **2011**, DOI: 10.1002/pola.25010.
(18) Nguyen, N. H.; Rosen, B. M.; Lligadas, G.; Percec, V. *Macromolecules* **2009**, *42*, 2379–2386.
(19) Rosen, B. M.; Percec, V. *J. Polym. Sci. A Polym. Chem.* **2007**, *45*, 4950–4964.
(20) Lin, C. Y.; Coote, M. L.; Gennaro, A.; Matyjaszewski, K. *Journal of the American Chemical Society* **2008**, *130*, 12762–12774.
(21) Isse, A. A.; Gennaro, A.; Lin, C. Y.; Hodgson, J. L.; Coote, M. L.; Guliashvili, T. *Journal of the American Chemical Society* **2011**, *133*, 6254–6264.
(22) Fischer, H. *Chem. Rev.* **2001**, *101*, 3581–3610.
(23) Chan, N.; Cunningham, M. F.; Hutchinson, R. A. *Macromol. Rapid Commun.* **2011**, *32*, 604–609.
(24) Voepel, J.; Edlund, U.; Albertsson, A.-C.; Percec, V. *Biomacromolecules* **2011**, *12*, 253–259.
(25) Edlund, U.; Albertsson, A.-C. *J. Polym. Sci. A Polym. Chem.* **2011**, *49*, 4139–4145.
(26) Houga, C. M.; Meins, J.-F. O. L.; Borsali, R.; Taton, D.; Gnanou, Y. *Chemical Communications* **2007**, 3063.
(27) Verma, M. S.; Gu, F. X. *Carbohydrate Polymers,* **2011***,* DOI: 10.1016/j.carbpol.2011.11.025.
(28) *Bioconjugate Techniques*; Hermanson, G. T., Ed. 2nd ed. Elsevier, **2008**.
(29) Britovsek, G. J. P.; England, J.; White, A. J. P. *Inorg. Chem.* **2005**, *44*, 8125–8134.
(30) Beuermann, S.; Paquet, D. A.; McMinn, J. H.; Hutchinson, R. A. *Macromolecules* **1996**, *29*, 4206–4215.
(31) Hutchinson, R. A.; Paquet, D. A., Jr; McMinn, J. H.; Beuermann, S.; Jackson, C. *DECHEMA Monogr.* **1995**, *131*, 467.
(32) Zhang, J.; Zhang, H.; Wu, J.; Zhang, J.; He, J.; Xiang, J. *Phys. Chem. Chem. Phys.* **2010**, *12*, 1941–1947.
(33) Perrin, C. L. *Pure & Appl. Chem.* **1995**, *67*, 719–728.
(34) Li, B.; Berliner, M.; Buzon, R. *J. Org. Chem.* **2006**, *71*, 9045-9050.
(35) Greenhill, J. V.; Taylor, E. C. *Heterocycles* **1991**, *32*, 2417–2427.
(36) Ojala, C.; Ostman, J. *Carbohydrate Research* **2002,** *337*, 21-29.
(37) Hutchins, R. O.; Hoke, D.; Keogh, J.; Koharski, D. *Tetrahedron Letters* **1969**, *10*, 3495–3498.
(38) Hendrickson, J. B.; Sternbach, D. D. *J. Org. Chem* **1975**, *40*, 3450–3452.
(39) Soeriyadi, A. H.; Boyer, C.; Nyström, F.; Zetterlund, P. B.; Whittaker, M. R. *Journal of the American Chemical Society* **2011**, DOI: 10.1021/ja205080u.
(40) Tang, W.; Matyjaszewski, K. *Macromolecules* **2007**, *40*, 1858–1863.

Chapter 5

Polysaccharide-stabilized Core Cross-linked Polymer Micelle Analogues

Significant portions of this chapter were accepted for publication in *Polymer Chemistry* on January 28, 2011 (DOI: 10.1039/C2PY00601D).

Abstract

A novel approach is presented for the synthesis of block-copolymers that resemble the structural architecture of a core cross-linked micelle. Because the approach in Chapter 4 was unsuccessful, an alternative approach towards the synthesis of amphiphilic polysaccharide-containing copolymers was attempted using commercially available materials. The polymers are synthesized from a combination of catalytic chain transfer polymerization (CCTP), thiol addition chemistry and reductive amination. A hydrophobic hyperbranched core is synthesized *via* CCTP of methyl methacrylate (MMA) and ethylene glycol dimethacrylate (EGDMA), which affords control over the polymer architecture and the degree of chain end-functionality. The vinyl unsaturations of the hyperbranched polymers are converted in nucleophilic pendant amines by thiol addition using cysteamine hydrochloride. A polysaccharide shell is grafted onto the hyperbranched core *via* reductive amination with dextran (DEX). The synthesized poly(MMA-*co*-EGDMA)-*b*-DEX polymers possess an amphiphilic character, are colloidally stable and resemble the topology of a core cross-linked micelle. The presented methodology provides a robust, modular, and tuneable approach towards the synthesis of amphiphilic core cross-linked micelle analogues without the need for elaborate chemistry and self-assembly.

5.1 Introduction

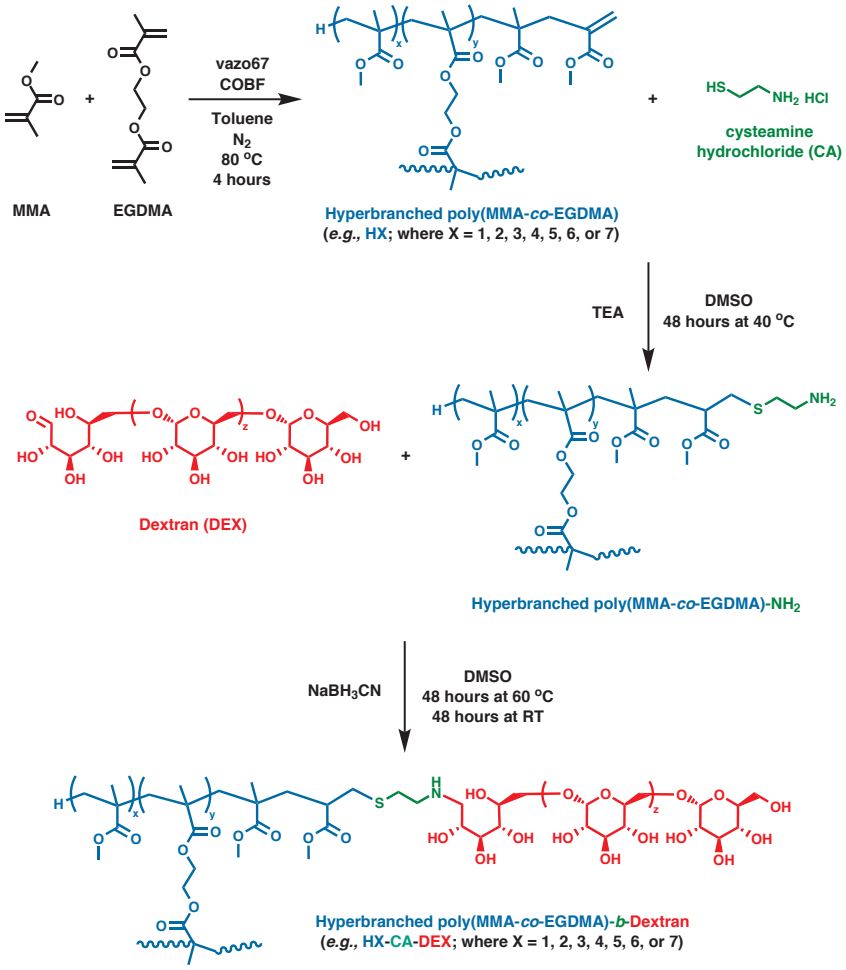
Tailoring of the size, architecture, and chemical functionality of macromolecules *via* synthetic polymer chemistry can produce polymeric materials with unique properties and, consequently, of stunning diversity.[1] One particular area that has received considerable attention is the synthesis of polymer colloids as drug and/or gene delivery vehicles for poorly water-soluble therapeutics.[2-5] An example is the exploration of linear amphiphilic block copolymers for self-assembly of micellar structures consisting of a hydrophobic core decorated with a hydrophilic shell.[6,7] Despite their potential, these colloidal structures suffer from elaborate chemistry followed by time-consuming self-assembly.[8] Furthermore, upon dilution these structures can disintegrate if the block-copolymer concentration falls below the critical micelle concentration (CMC). Although, cross-linking of the hydrophobic core or hydrophilic shell is able to prevent disintegration, it further adds to the multi-step synthesis of these materials.[9] We propose an alternative methodology for the synthesis of core cross-linked micelles by utilizing a hyperbranched polymer architecture, which circumvents many of the elaborate synthesis steps and self-assembly required in the synthesis of polymer micelles.

Thiol-ene chemistry has demonstrated to be a versatile tool for the post-polymerization modification of vinyl groups. Thiol-ene chemistry generally refers to the addition of nucleophilic thiols (-SH) to electrophilic -enes (C=C) *via* radical addition or base catalysis.[30-32] Thiol additions are particularly attractive since no additional copper catalyst is required, contrary to copper(I)-catalyzed azide-alkyne cycloadditions (CuAAc).[33] As a result, thiol additions have grown in popularity, in particular in combination with CCTP.[34-37] The use of functional thiols, such as cysteamine (CA, 2-aminoethanethiol), further enhances the versatility of thiol additions as it can convert an electrophilic -ene into a nucleophilic amine ($-NH_2$) as demonstrated by Haddleton's

group.[35] The pendant amine groups on hyperbranched macromolecules can potentially serve as reaction points for the reductive amination of aldehydes or ketones, thus allowing for the introduction of a shell-functionality to the structure. For example, van der Ende et al.[38] synthesized dendritic polyester macromolecules that were further functionalized using thiol-ene chemistry and reductive amination to produce core-shell drug delivery particles capable of targeting and transport across biological barriers. Reductive amination is also a versatile tool for coupling amines with polysaccharides or carbohydrates through the hemiacetal (aldehyde) moiety at the reducing chain-end.[39] These techniques have been widely used to functionalize glycopolymers[40] and glycodendrimers[41] to yield materials that have biocompatible and responsive properties.

Functionalization of the extremities of hyperbranched polymers with hydrophilic fragments can afford a core-shell-like structure that can mimic an amphiphilic block-copolymer micelle. Examples have been reported where the surface of a micelle or dendrimer has been modified to contain polysaccharide,[42,43] fluorescent[44] or poly(ethylene glycol) moieties.[45] Polysaccharides are particularly useful for drug and/or gene delivery applications because they are biodegradable and biocompatible.[46] Furthermore, in vivo polysaccharide-protein interactions are strengthened by the multivalency effects of polysaccharides.[47,48] Colloidal structures containing carbohydrate moieties have been used to create nanoparticles for biomedical imaging[49] and temperature responsive pH-degradable core cross-linked micelles.[50]

Here we report on a new approach for the synthesis of core-shell hyperbranched polymers using CCTP, thiol addition chemistry and reductive amination (see Scheme 5.1).



Scheme 5.1. Three step synthesis protocol for the formation of the core-crosslinked micelles poly(MMA-*co*-EGDMA)-*b*-DEX. The first step involves the copolymerization of conventional vinyl monomers in the presence of the chain transfer agent COBF to produce a THF-soluble hyperbranched core with pendant vinyl moieties. These external double bonds are then modified *via* thiol addition chemistry with the commercially available functional thiol cysteamine hydrochloride to give an amine-functional hyperbranched core poly(MMA-*co*-EGDMA)-NH_2. The peripheral amine groups are then used as nucleophiles in the subsequent reductive amination reaction with the reducing end of dextran.

This three step synthetic approach first involves CCTP of methyl methacrylate (MMA) and ethylene glycol dimethacrylate (EGDMA) in the presence of bis(boron difluorodimethylglyoximate) cobalt(II) (COBF) to produce hyperbranched poly(MMA-*co*-

EGDMA) of varying size, degrees of branching and fraction of pendant vinyl groups. Thiol addition chemistry is subsequently used to convert the pendant vinyl bonds (-ene group) to give quantitative yields of the hyperbranched poly(MMA-*co*-EGDMA)-NH_2. Finally, dextran (DEX) is covalently attached to the amine groups *via* reductive amination at the reducing end of the polysaccharide, yielding a poly(MMA-*co*-EGDMA)-*b*-DEX. These branched core-shell structures are amphiphilic and mimic the topology of a core cross-linked micelle. The influence of the COBF and EGDMA concentration on hyperbranched architecture was investigated. The final colloidal properties of the poly(MMA-*co*-EGDMA)-*b*-DEX were also studied and related to the characteristics (i.e. the degrees of branching and unsaturations in the polymer structure) of the hyperbranched core.

5.2 Experimental

5.2.1 Materials

Methyl methacrylate (MMA, Aldrich, 99%) and ethylene glycol dimethacrylate (EGDMA, Aldrich, 98%) were passed over a column containing basic aluminum oxide (Aldrich, ~150 mesh, 58 Å) to remove the inhibitor and stored at 5 °C prior to polymerization. Diethyl ether (Fisher, reagent grade), toluene (Fisher, reagent grade) and dimethyl sulfoxide (DMSO, Fisher, reagent grade) were used as received. Tetrahydrofuran (THF, Aldrich, 99%) was either distilled or used as received. Bis(boron difluorodimethylglyoximate) cobalt(II) (COBF) was prepared according to literature methods[51] and had an intrinsic activity (C_T) of 20×10^3, as determined in MMA solution polymerization at 80 °C. The initiator, 2,2'-azobis(2-methylbutyronitrile) (vazo67, DuPont) was recrystallized from methanol and stored below 0 °C prior to use. Triethylamine (TEA, Sigma-Aldrich, 99%), cysteamine hydrochloride (CA, Sigma,

98%), sodium cyanoborohydride ($NaBH_3CN$, Aldrich, 95%), and dextran (Sigma-Aldrich, M_w = 5678 g·mol^{-1}, M_n = 3154 g·mol^{-1}, $D = M_w/M_n$ = 1.80) were used as received.

5.2.2 Synthesis of poly(MMA-*co*-EGDMA)

The poly(MMA-*co*-EGDMA) polymers were synthesized in homogenous solution CCTP. All polymerizations were performed in a three-neck 100 mL round bottom flask equipped with a magnetic stir bar. A catalyst stock solution (0.05 mg·mL^{-1}) was prepared by adding an appropriate amount of deoxygenated monomer (MMA) into a degassed Schlenk tube containing an accurate amount of COBF. This solution was then subjected to three vacuum-pump cycles with nitrogen to remove any residual oxygen. A typical poly(MMA-*co*-EGDMA) synthesis was performed as follows (Table 5.1, entry H6): vazo67 (30 mg, 0.16 mmol), toluene (25 mL), MMA (19.2 mL, 0.23 mol), and EGDMA (2.2 mL, 12 mmol) were added to the reaction flask and purged for at least 30 minutes with nitrogen.

Table 5.1. CCTP of MMA and EGDMA for the synthesis of the hyperbranched core.

Entry	COBF (ppm)	f_{tr} (s^{-1})[a]	MMA (mol%)	EGDMA (mol%)	φ (s^{-1}·mol%$^{-1}$)[b]
L1	0.2	48	100.0	0.0	N/A
L2	2.5	598	100.0	0.0	N/A
H1	0.5	120	99.0	1.0	120
H2	2.0	478	97.8	2.2	217
H3	1.0	239	98.6	1.4	171
H4	3.0	717	95.8	4.2	171
H5	1.0	239	98.0	2.0	120
H6	3.0	717	94.9	5.1	141
H7	2.0	478	96.8	3.2	149

[a] The chain transfer frequency (f_{tr}) is a measure for the chain transfer activity in a polymerization, expressed as $f_{tr} = k_{tr}$[Co], [b] φ is the ratio of the chain transfer frequency and the cross-linker concentration in mol%, defined as $\varphi = f_{tr}$ / [EGDMA].

Subsequently, an appropriate amount of the COBF stock solution (5.8 mL, 0.05 mol MMA, 7.0 ×

10^{-4} mmol COBF) was transferred using a degassed syringe into the reaction flask. The total amount of MMA used for each experiment was kept constant at 25 mL (0.23 mol). For linear poly(MMA) the same procedure outlined above was used, with the exception that no EGDMA was added to the polymerization. The reaction flask was then attached to a condenser and placed in an 80 °C thermostatted oil bath under magnetic stirring and allowed to polymerize for approximately 4 hours. The polymer in solution was then diluted with THF and passed through a column containing basic aluminum oxide to remove COBF. Subsequently, the solvent and remaining monomers were removed and the polymer dried at 85 °C under atmospheric pressure. The polymers were recovered as films and then further characterized as described below.

5.2.3 Thiol addition: poly(MMA-*co*-EGDMA)-NH_2

In a 20 mL test tube, hyperbranched poly(MMA-*co*-EGDMA) polymer (1.0 g) and DMSO (10 mL) were added and stirred at room temperature overnight. Following the complete dissolution of the hyperbranched polymer, 2 equivalents of thiol (CA) and 4 equivalents of TEA with respect to the quantity of vinyl bonds are added. The solution was stirred at 40 °C for 48 hours until the complete disappearance of the proton signals of the vinyl bonds as observed by ^{1}H-NMR, e.g., δ(400 MHz; $CDCl_3$) 5.3 (1H, d, C=CH_b) and 6.3 (1H, d, C=CH_a). The reaction mixture was then precipitated into excess cold diethyl ether, decanted and then washed with excess diethyl ether and finally dried in vacuo. ^{1}H-NMR in $CDCl_3$ was used to confirm the full conversion of vinyl bonds to amines. ^{1}H-NMR: δ(400 MHz; $CDCl_3$): 2.9 (2H, -SH-CH_2-CH_2-NH_2), 3.2 (2H, -SH-CH_2-CH_2-NH_2), 3.6 (3H, -O-CH_3), 8.4 (2H, -NH_2).

5.2.4 Reductive amination: poly(MMA-*co*-EGDMA)-*b*-DEX

In a 20 mL test tube, poly(MMA-*co*-EGDMA)-NH_2 (0.1 g) and DMSO (10 mL) were added and stirred at room temperature overnight. Following complete dissolution of poly(MMA-*co*-EGDMA)-NH_2, 1 equivalent of dextran to the amount of thiol added in the previous step was added, followed by 3 equivalents of sodium cyanoborohydride. The mixture was magnetically stirred at 60 °C for 48 hours and room temperature for an additional 48 hours. The reaction mixtures containing linear polymers were then funnelled into 6000-8000 $g \cdot mol^{-1}$ molecular weight cutoff (MWCO) dialysis tubing and, similarly, the hyperbranched polymer solutions were poured into 12000 $g \cdot mol^{-1}$ MWCO dialysis tubing and dialyzed against distilled deionized water (DDW, Millipore Synergy water filtration unit) for approximately one week with daily water replacement to purify the resulting poly(MMA-*co*-EGDMA)-*b*-DEX. The final copolymers were lyophilized to remove water. Henceforth, poly(MMA-*co*-EGDMA)-*b*-DEX copolymers will be referred to as HX-CA-DEX where X ranges from 1 to 6, corresponding to the 6 hyperbranched cores synthesized *via* CCTP.

5.2.5 Size Exclusion Chromatography (SEC)

Linear size exclusion chromatography (LSEC) was performed using a Waters 2690 separation module and a model 410 differential refractometer. Five Waters Styragel HR columns (HR5.0, HR4.0, HR3.0, HR1.0, and HR0.5) in series were used at a constant temperature of 40 °C. Distilled THF was used as eluent at a flow rate of 1.0 mL min^{-1}. The LSEC system was calibrated using narrow molecular weight poly(styrene) standards ranging from 374 to 400 × 10^3 g mol^{-1}. The Mark-Houwink-Kuhn-Sakurada parameters used for the poly(styrene) standards were $K = 1.14 \times 10^{-4}$ $dL \cdot g^{-1}$ and $a = 0.716$.[52] Similarly, for poly(methyl methacrylate), $K = 0.944 \times 10^{-4}$ $dL \cdot g^{-1}$ and $a = 0.719$.[53]

Triple detector size exclusion chromatography (TDSEC) was performed using a Viscotek 270max separation module with a triple detector configuration comprising refractive index (RI); light scattering (low angle light scattering (LALS) and right angle light scattering (RALS)); and viscosity (IV) detectors. Two porous PolyAnalytik columns with an exclusion molecular weight of 20×10^6 g·mol^{-1} were used in series at a constant temperature of 40 °C. Distilled THF was used as eluent at a flow rate of 1.0 mL·min^{-1}. The TDSEC system was calibrated using a narrow poly(styrene) standard of 99×10^3 g·mol^{-1} with a refractive index of 0.185 mL·g^{-1} and an intrinsic viscosity of 0.477 dL·g^{-1}. Traditionally, viscometry has been used to characterize hyperbranched polymers by comparison to linear analogues, according to the Mark-Houwink equation[54], as shown in Equation 5.1.

$$[\eta] = KM_v^a \qquad (5.1)$$

where $[\eta]$ is the intrinsic viscosity, K and a are the Mark-Houwink constants, and M_v is the viscosity-average molecular weight. Branching values for the hyperbranched polymers were determined using a standard sample of linear poly(MMA) synthesized *via* CCTP (LSEC: M_w = 31284; M_n = 8780; D = 3.6) and a random, tri-functional polydisperse branching calculation.[55] All TDSEC analyses were performed using Viscotek OmniSEC software (version 4.6.1).

5.2.6 Nuclear Magnetic Resonance (NMR)

^{1}H-NMR and Heteronuclear Single-Quantum Correlation (HSQC) spectroscopy were performed on a 400 MHz Bruker Avance instrument (Bruker BioSpin GmbH) in $CDCl_3$, DMSO-d_6, or D_2O. A total of 32 and 8 scans per spectrum were recorded for the ^{1}H and HSQC spectra, respectively. The mol% of vinyl bonds ($C=CH_2$) in the linear and hyperbranched polymer structures was estimated from the integral of the signals of the vinyl and methoxy protons. ^{1}H-

NMR: δ(400 MHz; $CDCl_3$): 3.8 – 3.2 (2H, C=CH_2), 6.3 – 5.3 ppm (3H, O-CH_3).

5.2.7 Attenuated Total Reflectance Fourier Transform Infrared (ATR FT-IR) Spectroscopy

Fourier Transform Infrared (FT-IR) spectroscopy was performed on a Thermo Scientific Nicolet 6700 instrument using an Attenuated Total Reflectance (ATR) accessory equipped with a diamond crystal. A total of 64 scans were co-added per spectrum with a background spectrum of air taken between successive samples. Constant pressure was applied to the sample on the diamond crystal using a hand-tightened anvil.

5.2.8 Differential Scanning Calorimetry (DSC)

Differential Scanning Calorimetry (DSC) was conducted on a Q2000 DSC device (TA Instruments) under nitrogen flow. The temperature was varied between -20 °C and 150 °C at a rate of 10 °C·min^{-1}.

5.2.9 Dynamic Light Scattering (DLS)

Particle size analysis was performed on a Malvern Zetasizer Nano ZS at 25 °C using a DTS0012 disposable cuvette. The lyophilized hyperbranched polymers were diluted with DDW to a concentration of 0.1 wt% prior to measurement. Refractive indexes used for the measurements were 1.330 for water and 1.440 for polymer samples.

5.2.10 Transmission Electron Microscopy (TEM)

TEM was performed on a Hitachi H-7000 instrument at an operating voltage of 75 kV. Solutions of 1 wt% lyophilized hyperbranched polymer in distilled deionized water were prepared for TEM analysis.

5.3 Results and Discussion

5.3.1 The core: Synthesis and properties of the hyperbranched polymers

The core of the poly(MMA-*co*-EGDMA)-*b*-DEX block-copolymers consists of a hyperbranched poly(MMA-*co*-EGDMA) copolymer synthesized *via* the "Strathclyde methodology"[15] using COBF as the chain transfer agent. The architecture of the polymer (i.e. hyperbranched or cross-linked) is dependent on the concentration of EGDMA and the intrinsic activity of the chain transfer agent (i.e., COBF).[18] The chain transfer activity in CCTP can be expressed in terms of the chain transfer frequency (f_{tr}): the product of the rate coefficient of chain transfer (k_{tr}) and the COBF concentration ([Co]), as shown in Equation 5.2.

$$f_{tr} = k_{tr}[\mathrm{Co}] \qquad (5.2)$$

Expressing the chain transfer activity in terms of f_{tr} allows for an empirical correlation between the chain transfer activity and the cross-linker concentration to predict the polymer architecture and the degree of branching. Previously it has been reported that in the copolymerization of MMA and a divinyl monomer in the presence of a chain transfer agent the ratio (φ) of the f_{tr} (s^{-1}) and the cross-linker concentration in mol% should equal $\varphi > 85$ to avoid macro-gelation.[18] Furthermore, the structure of the hyperbranched polymer can be controlled by varying the value of φ, i.e. by changing either the EGDMA or COBF concentration (or both) as long as the criterion of $\varphi > 85$ is satisfied.

The recipes used for the synthesis of the hyperbranched poly(MMA-*co*-EGDMA) polymers are collected in Table 5.1. The COBF concentrations range from 0.5 to 3.0 ppm (defined as moles of COBF per 10^6 moles of MMA and EGDMA), corresponding to a f_{tr} ranging from 120 to 720 s^{-1}. The EGDMA concentration is subsequently varied from 1.0 to 5.0 mol%

(Table 5.1), yielding values for φ ranging from 120 to 217 $s^{-1} \cdot mol\%^{-1}$. The positioning of the recipes H1 to H7 with respect to the cross-linking threshold of $\varphi > 85$ is presented in Figure 5.1.

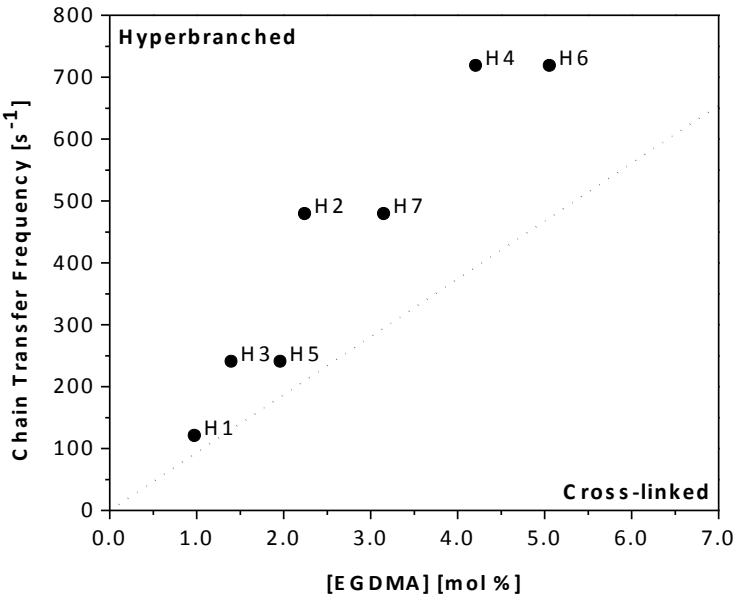


Figure 5.1. Correlation between the chain transfer frequency and cross-linker concentration for the synthesized hyperbranched cores. The solid line represents the threshold between hyperbranched and cross-linked ($\varphi = 0.85$).

All of the hyperbranched poly(MMA-*co*-EGDMA) polymers were fully soluble in THF and DMSO, indicating that hyperbranched polymers were indeed successfully synthesized.

Information on the polymer architecture was obtained from ^{1}H-NMR spectroscopy and triple detection size exclusion chromatography (TDSEC). It can be seen from Table 5.2 that the mol% of vinyl bonds in the hyperbranched polymer structure increases with increasing f_{tr} and/or EGDMA concentration.

Table 5.2. Molecular weight, branching, and thermal properties of linear poly(MMA) and hyperbranched poly(MMA-*co*-EGDMA) samples *via* TDSEC and DSC analyses.

Entry	C=CH$_2$ (mol%)[a]	M_n (10^3 g·mol^{-1})[b]	M_w (10^3 g·mol^{-1})[b]	D (-)[b]	[η] (mL·g^{-1})[c]	R_H (nm)[c]	Branch (#)[c]	Branch Freq.(-)[c]	g' (-)[c]	T_g (^{o}C)
L1	3.0	15.3	49.3	3.2	14.3	4.4	0.0	0.0	1.00	72
L2	1.0	1.4	3.4	2.4	112	3.7	0.0	0.0	1.00	73
H1	3.0	22.1	127	5.7	12.3	5.4	8.2	1.1	0.65	56
H2	8.0	5.5	23.8	4.4	5.3	2.4	4.9	2.2	0.63	52
H3	3.0	14.0	79.7	5.7	11.5	4.5	4.2	0.3	0.67	61
H4	12.0	5.6	23.1	4.2	5.1	2.4	5.3	2.6	0.64	51
H5	5.0	36.0	1050	29.2	20.0	11.3	13.5	0.4	0.16	58
H6	17.0	8.9	30.8	3.5	3.1	2.2	30.6	18.7	0.37	51
H7	10.0	183	2360	12.9	29.7	17.2	59.0	0.1	0.03	56

[a] Determined from ^{1}H-NMR, [b]Determined from linear-SEC, [c] Determined from triple detector SEC.

An increase in f_{tr} results in more chain transfer events and consequently the formation of more vinyl bonds. Likewise, an increase in the EGDMA concentration increases the probability of branching which generates more chain-ends that can undergo CCTP. Consequently, the chain-end functionality of the hyperbranched polymers can be controlled *via* the CCTP conditions.

LSEC was used to determine the number and weight-average molecular weight (M_n and M_w, respectively) and dispersity ($D = M_w / M_n$) of the synthesized linear and hyperbranched polymers (Table 5.2). Information about the polymer architecture of the hyperbranched polymers was obtained from TDSEC. Figure 5.2A presents the molecular weight distributions (MWDs) of a linear poly(MMA) L2 and the hyperbranched poly(MMA-*co*-EGDMA) H2, H4, H6 and H7, which are polymerized in the presence of the highest COBF and EGDMA concentrations.

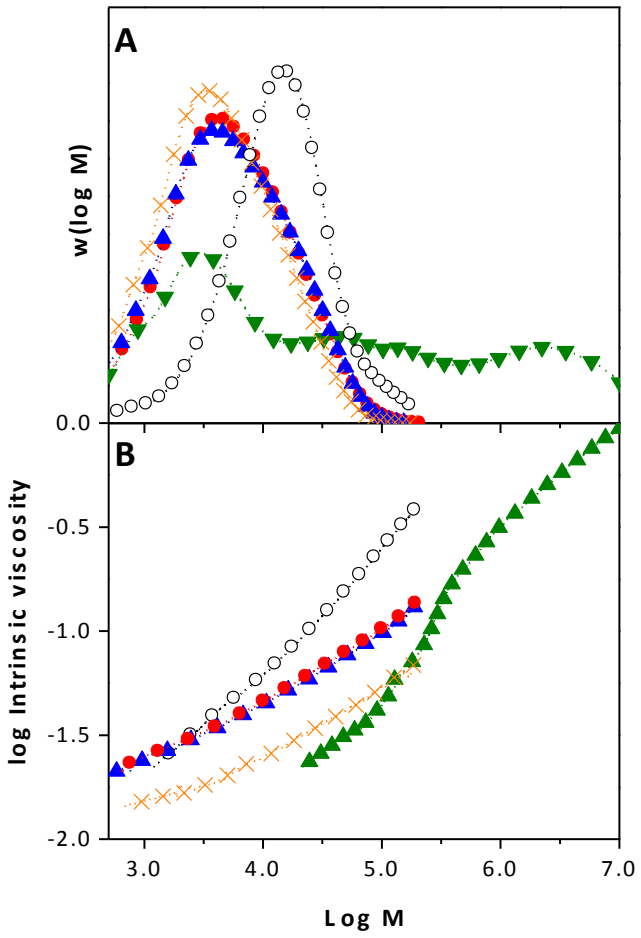


Figure 5.2. TDSEC molecular weight analysis of the hyperbranched polymers. A. The MWDs (RI signal) and B. Mark-Houwink plots of the intrinsic viscosity both as a function of log molecular weight. L2 (○), H2 (●), H4 (▲), H6 (✖), H7 (▼).

In CCTP typically monomodal MWDs are obtained with a *D* of 2 (see entry L2, Table 5.2). It can be seen from Figure 5.2A that the MWDs of H2, H4, H6 and H7 are significantly broader than the MWD of L2. The width of the MWD is quantified by the *D* in Table 5.2, which is significantly larger than 2 for all the hyperbranched polymers.

The Mark-Houwink plots for L2, H2, H4, H6 and H7 are compared in Figure 5.2B. It can

be observed that the signals in Figure 5.2B are lower over the entire molecular weight range for the hyperbranched polymers. This indicates a lower intrinsic viscosity and, the presence of branches in the polymer structure. The average intrinsic viscosity ([η]) of the linear poly(MMA) L2 is higher than that of the hyperbranched poly(MMA-*co*-EGDMA) (Table 5.2). This can be attributed to the decreased hydrodynamic radius of hyperbranched polymers.

The hyperbranched architecture of the poly(MMA-*co*-EGDMA) polymers can be characterized by the number of branches, the branching frequency and the average Zimm branching factor (g').[54] The average Zimm branching factor per definition equals 1 for a linear polymer and ranges between 0 and 1 for hyperbranched polymers. It can be seen from Table 5.2 that the poly(MMA-*co*-EGDMA) polymers become increasingly hyperbranched as the EGDMA concentration is increased for a given f_{tr}. An increase in the EGDMA concentration promotes the formation of more branches and consequently the number of branches and the branching frequency increase, whereas g' decreases. This clearly demonstrates that hyperbranched polymer architecture of the core can be controlled by varying the COBF and EGDMA concentrations. Furthermore, control over the chain-end functionality (i.e. the mol% of pendant vinyl groups) is readily available and allows for control over the grafting density of the shell in the final polymer architecture.

5.3.2 The core: Thiol addition

Prior to the attachment of the hydrophilic shell, the chain-ends of the hyperbranched poly(MMA-*co*-EGDMA) polymers have to be functionalized with a nucleophilic amine functionality. Thiol addition is performed using cysteamine hydrochloride (CA). The poly(MMA-*co*-EGDMA) polymers are reacted with CA in the presence of 2 mole equivalents of TEA (1 mole equivalent for the Thiol addition and 1 mole equivalent for the deprotonation of the

cysteamine hydrochloride). The reaction was monitored by ^{1}H-NMR (Figure 5.3).

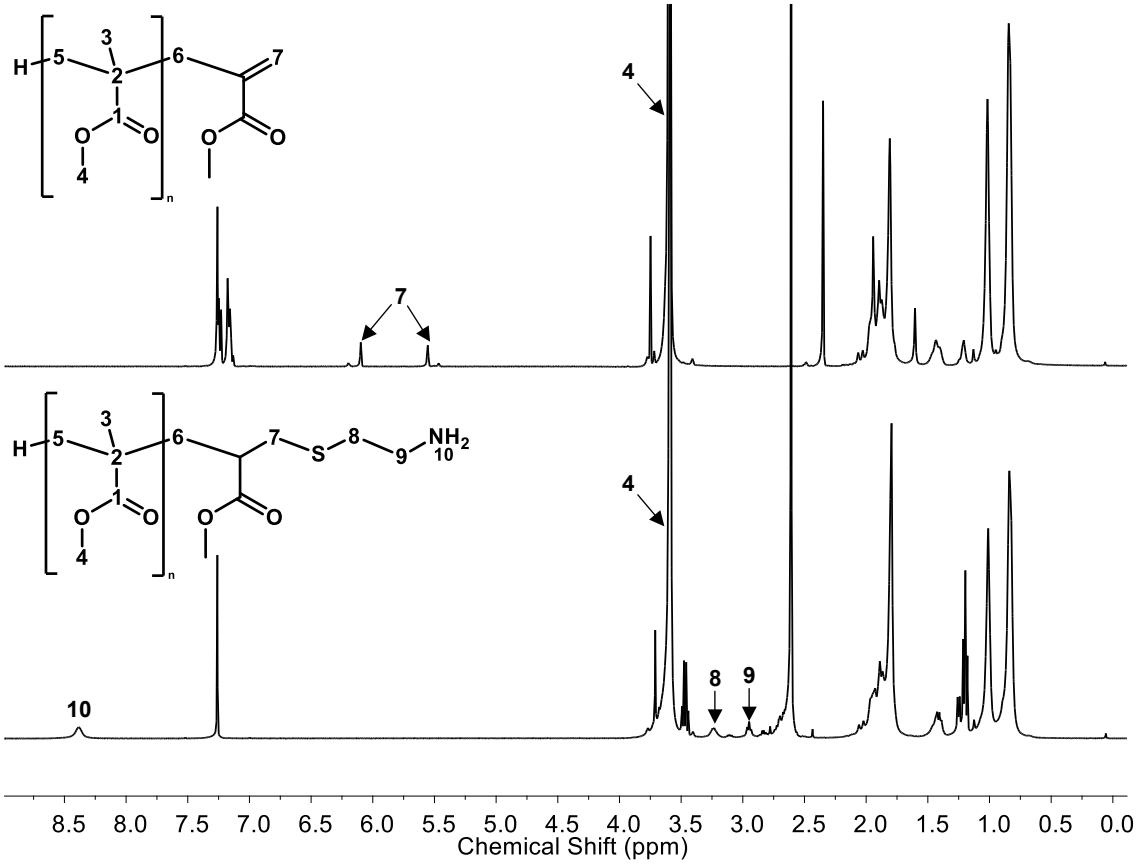


Figure 5.3. ^{1}H-NMR ($CDCl_3$) spectra of L1 before (top) and after (bottom) thiol addition chemistry with cysteamine hydrochloride.

It can be seen that the characteristic signals of the vinyl protons at δ = 5.5 ppm (1H, d, C=CH_b) and δ = 6.3 ppm (1H, d, C=CH_a) disappear and that a new signal appears at δ = 8.4 ppm (2H, s, -NH_2) due to the introduction of a pendant amine group. The thiol addition attained nearly quantitative conversions in 24 hours. The pendant amine groups serve as nucleophiles in the subsequent reductive amination reaction with dextran for the attachment of the shell.

5.3.3 The shell: Reductive amination of poly(MMA-*co*-EGDMA)-NH_2 with dextran

The hydrophilic shell, and therewith the amphiphilic character of the polymer, was introduced by a reductive amination between the hyperbranched poly(MMA-*co*-EGDMA)-NH_2 polymer and the hemiacetal functionality of dextran. The reductive amination was carried out for approximately 2 days at 60 °C and 2 days at room temperature in DMSO. Subsequently, the solutions were dialyzed for a week to remove low molecular weight impurities, unreacted dextran and to displace the DMSO by distilled deionized water.

During dialysis a number of the synthesized polymers formed translucent colloidal suspensions inside the dialysis tubing (see Figure 5.4).

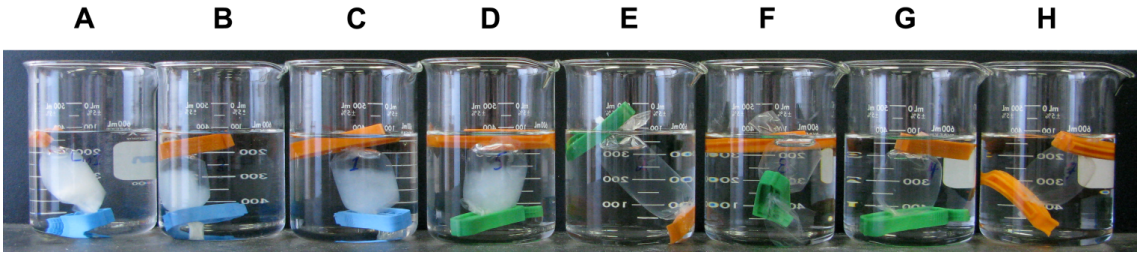


Figure 5.4. Photograph taken during dialysis of linear poly(MMA)-*b*-DEX and hyperbranched poly(MMA-*co*-EGDMA)-*b*-DEX against DDW. The dialysis tubes labelled A, B, C, and D contain insoluble copolymers, while E, F, G, and H contain colloidally stable systems. A: L1-CA-DEX, B: H3-CA-DEX, C: H1-CA-DEX, D: H5-CA-DEX, E: H2-CA-DEX, F: H7-CA-DEX, G: H4-CA-DEX, H: H6-CA-DEX.

This was a first indication of the successful grafting of dextran to the amine-functionalized polymer. The amine functionalized poly(MMA-*co*-EGDMA)-NH_2 polymer is hydrophobic and if there were no dextran (or an insufficient amount of dextran) attached, the polymer will simply precipitate once the DMSO is replaced by water (see for example A, B, C, and D in Figure 5.4). However, if a sufficient level of dextran is grafted onto the core, it will render the polymer

amphiphilic and in some cases colloidally stable in water (see for example E, F, G, and H in Figure 5.4).

Quantitative proof of the successful reductive amination was obtained from HSQC, ^{1}H-NMR, and ATR FT-IR spectroscopy

Figure 5.5 presents the 2D HSQC spectrum of H6-CA-DEX in DMSO-d_6 recorded after extensive dialysis and lyophilization.

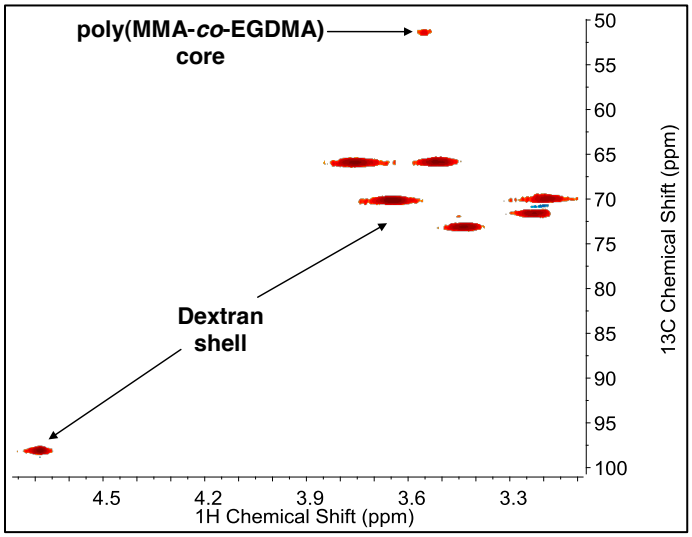


Figure 5.5. 2D HSQC NMR (DMSO-d6) spectrum of H6-CA-DEX after extensive dialysis and lyophilization showing carbon-coupled protons from both the core (methoxy protons) and shell (backbone dextran protons) components.

The HSQC spectrum displays the proton and carbon signals for the methoxy group (-O-CH_3) of poly(MMA-*co*-EGDMA) at $\delta\{3.5, 51.5\}$ (i.e. $\delta\{^{1}H, ^{13}C\}$). Furthermore, proton and carbon signals are observed that originate from the dextran C-C backbone at $\delta\{3.2, 70.1\}$, $\delta\{3.2, 71.6\}$, $\delta\{3.4, 73.1\}$, $\delta\{3.5, 65.9\}$, $\delta\{3.6, 70.2\}$, $\delta\{3.8, 66.0\}$, and $\delta\{4.7, 98.1\}$. This confirms the presence of dextran in the polymer structure, in addition to the poly(MMA-*co*-EGDMA). And

since unreacted dextran has been removed through extensive dialysis, it suggests that dextran is covalently attached to the hyperbranched core. The HSQC spectrum also provides some information about the structure of the poly(MMA-*co*-EGDMA)-*b*-DEX polymer. The intensity of the methoxy peak (i.e. the protons displaying the highest intensity in the ^{1}H-NMR of the core) is suppressed when compared to signals of the dextran backbone protons. This suggests a core-shell structure, where the proton signals from dextran have a higher ^{1}H intensity due to their peripheral location.

Further evidence of the core-shell structure was provided by ^{1}H-NMR of H6-CA-DEX (see Figure 5.6) in three NMR solvents that selectively solubilize the dextran shell (D_2O), the synthetic core ($CDCl_3$), and both the shell and core materials together (DMSO-d_6).

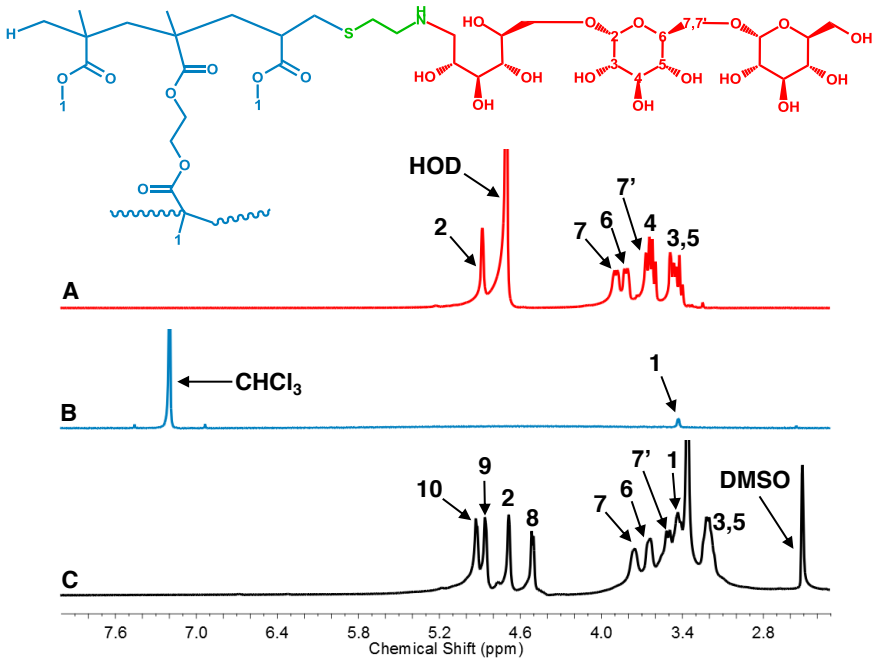


Figure 5.6. ^{1}H-NMR spectra of H6-CA-DEX in three different solvents that selectively solubilize the shell (A: D_2O), the core (B: $CDCl_3$), and the entire core cross-linked micelle (C: DMSO-d_6).

It was observed from the spectra in Figure 5.6 that the chemical shifts belonging to the dextran backbone protons are visible in D_2O, suggesting that a dextran shell is present, while in $CDCl_3$ only a small peak belonging to the methoxy protons of the poly(MMA-*co*-EGDMA) core are visible. The relatively weak intensity of the three methoxy protons of the core in $CDCl_3$ suggests that the synthetic groups inside the core cross-linked micelle have limited mobility due to steric effects. In DMSO-d_6, the spectrum shows peaks arising from the protons of both dextran and the methoxy groups of poly(MMA-*co*-EGDMA). The absence of the reducing-end protons of dextran at δ = 6.6 and δ = 6.3 ppm in DMSO-d_6 confirms that the dextran chains were covalently linked, *via* reductive amination, to the pendant -NH_2 groups of the poly(MMA-*co*-EGDMA) core. The linear poly(MMA)-*b*-DEX also showed a similar absence of reducing end protons. These NMR studies lend credence to the expected core-shell structure of the final poly(MMA-*co*-EGDMA)-*b*-DEX material.

ATR FT-IR spectra of the linear and hyperbranched copolymers after reductive amination, dialysis, and lyophilization were recorded and compared to similar spectra of pure dextran and poly(MMA). The characteristic ATR FT-IR vibration mode of poly(MMA) is the C=O carbonyl stretching peak around 1727 cm^{-1}. For dextran, the C-O glycosidic linkage stretching peak around 1012 cm^{-1} is the characteristic vibration mode. It can be seen from Figure 5.7 that the C=O stretching as well as the C-O stretching are both visible in the poly(MMA-*co*-EGDMA)-*b*-DEX copolymer spectra. Similar observations in FT-IR spectra were made for a comparable copolymer system of poly(MMA)-*g*-cellulose triacetate.[49] A comparison of the intensities of the carbonyl absorbance peak arising from the poly(MMA-*co*-EGDMA) core and the glycosidic absorbance peak of the dextran shell provides further evidence for the core-shell structure. The carbonyl absorbance peak of the poly(MMA-*co*-EGMDA)-*b*-DEX polymers

decreases when compared to the glycosidic absorbance peak with increasing amounts of vinyl bonds in the hyperbranched core. ATR FT-IR probes the fundamental chemical vibrations close to the surface of the sample and, as a result, samples with more vinyl bonds and (consequently more dextran) should display a suppressed carbonyl character in ATR FT-IR due to the core-shell structure. For example, H2-CA-DEX (Figure 5.7A) and H4-CA-DEX (Figure 5.7B) differed only in the amount of vinyl bonds in the core and consequently the amount of dextran in the shell; so the expected suppression of carbonyl signature of the sample with more surface dextran (i.e., H4-CA-DEX, Figure 5.7) is clearly visible in the ATR FT-IR signatures in Figure 5.7.

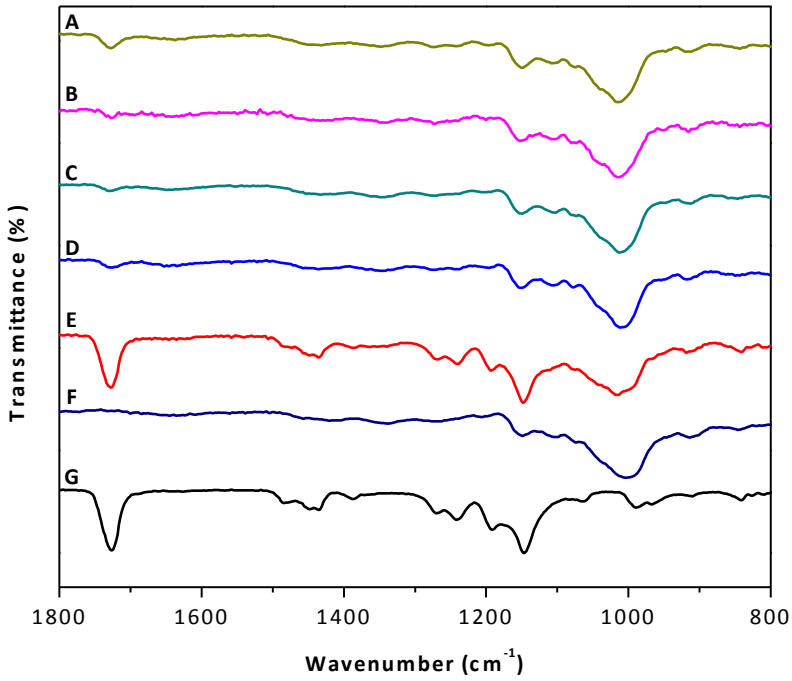


Figure 5.7. ATR FT-IR spectra of the linear (L1-CA-DEX) and hyperbranched poly(MMA-*co*-EGDMA)-*b*-DEX polymers synthesized in this study showing both poly(MMA) (C=O at 1727 cm^{-1}) and dextran (C-O at 1012 cm^{-1}) characteristic peaks. A: H2-CA-DEX, B: H4-CA-DEX, C: H6-CA-DEX, D: H7-CA-DEX, E: L1-CA-DEX, F: DEX, G: poly(MMA).

Furthermore, the linear poly(MMA)-*b*-DEX copolymer showed the expected glycosidic carbonyl absorbance peaks; however, no suppression of the carbonyl vibration mode can be observed due to the linear structure of this particular copolymer.

The HSQC, ^{1}H-NMR, and ATR FT-IR spectra in combination with the visual observations shown in Figure 5.4 suggest the formation of a core-shell structure after dextran was successfully grafted to the surface of the otherwise hydrophobic hyperbranched cores. Only those poly(MMA-*co*-EGDMA) cores with a fraction of vinyl bonds > 8 mol% (i.e. H2, H4, H6, and H7) were colloidally stable in aqueous solution after dextran was attached to the surface. The amount of vinyl bonds on the surface of the core cross-linked polymer ultimately limits the amount of dextran that can be attached. Conveniently, since the amount of vinyl bonds is essentially dependent on the CCTP recipe (i.e. the f_{tr} and the EGDMA concentration), the colloidal properties of final core cross-linked micelle can be tailored based on the initial CCTP polymerization conditions. A stable colloidal system can be produced due to the steric stabilization effect of the dextran groups[56] on the surface of the hyperbranched core polymer.

5.3.4 Colloidal properties of the poly(MMA-*co*-EGDMA)-*b*-DEX polymers

The colloidally stable poly(MMA-*co*-EGDMA)-*b*-DEX core cross-linked micelles were analyzed by dynamic light scattering (DLS) (Figure 5.8) and transmission electron microscopy (TEM) (Figure 5.9) to discern the size and morphology of the synthesized colloids. An analysis of the particle size distribution (PSD), in terms of the Z-average (Z-avg.), number-average ($D_{p(N)}$), volume-average ($D_{p(V)}$) particle sizes and the polydispersity index (PDI = $D_{p(V)}$ / $D_{p(N)}$) is presented in Table 5.3.

Table 5.3. Particle size properties determined by DLS of 0.1 wt.% solutions of core-shell polymers in DDW

Entry	*Z*-avg. (nm)	$D_{p(N)}$ (nm)	$D_{p(V)}$ (nm)	PDI[a]
H2-CA-DEX	50.2	31.1	74.2	2.37
H4-CA-DEX	28.3	29.3	35.5	1.21
H6-CA-DEX	18.1	15.9	20.0	1.26
H7-CA-DEX	240.3	16.9	349.5	20.6

[a] The averages are calculated from the primary distribution only.

It can be seen from Figure 5.8 that the poly(MMA-*co*-EGDMA)-*b*-DEX polymers generally form colloidal nanoparticles that have a monomodal PSD with particle sizes below 100 nm.

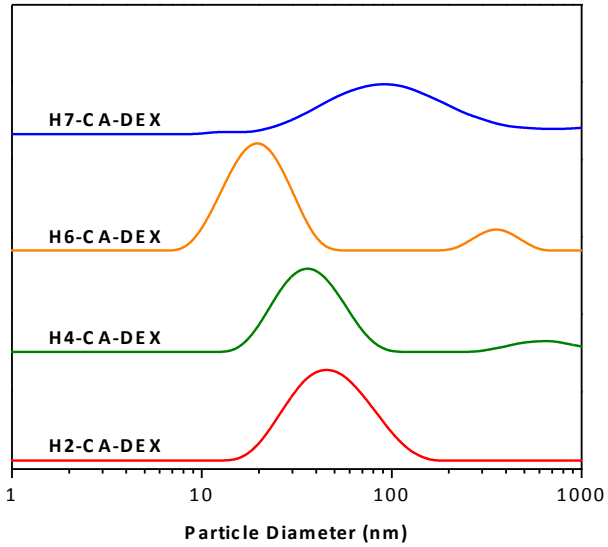


Figure 5.8. The Z-average particle size distribution as measured by DLS of four hyperbranched poly(MMA-*co*-EGDMA)-*b*-DEX samples 0.1 wt.% in DDW.

The H7-CA-DEX polymer is an exception where a very broad PSD is obtained with particle sizes > 100 nm. Similar trends are observed in the number- and volume-average PSDs. The PSDs of H4-CA-DEX and H6-CA-DEX can be considered monodisperse as the PDI value is approximately 1.20 – 1.30 (for a monodisperse sample PDI = 1.0, in practice a value between 1.0

and 1.5 can be considered monodisperse). The PSDs of H2-CA-DEX and H7-CA-DEX are significantly broader with PDIs of 2.37 and 20.6, respectively, and can be considered polydisperse. The PSD of H4-CA-DEX and H6-CA-DEX also showed evidence of particle aggregation manifested in a secondary population of particles > 500 nm (Figure 5.8). The colloidal solutions of H2-CA-DEX, H4-CA-DEX, H6-CA-DEX and H7-CA-DEX all remained stable over a period of at least a month prior to aggregation. The zeta potential (ζ) of these colloids is virtually $\zeta = 0$, which implies that colloidal stability is entirely dependent on the steric stabilization provided by dextran. In the absence of electrostatic repulsion, aggregation will be driven by van der Waals forces and the amount of dextran grafted to the poly(MMA-*co*-EGDMA) may not be sufficient to mask these attractions. However, aggregation of the H6-CA-DEX colloids was found to be reversible and the nanoparticles could be re-suspended by gentle sonication. The H7-CA-DEX colloids could only partially be re-suspended.

The characteristics of the PSD (in terms of the width and the average particle sizes) can be correlated to the properties of the hyperbranched core polymer. The average molecular weight and therewith the hydrodynamic radius (R_H, see Table 5.2) of the core governs the average particle size. H2-CA-DEX, H4-CA-DEX and H6-CA-DEX all have M_w values around $30 \cdot 10^3$ $g \cdot mol^{-1}$ and a $R_H \sim 2.2$ nm. The similarity in the molecular weight properties is also apparent in the corresponding particle sizes of these polymers (see Table 5.3). H7-CA-DEX has a significantly higher M_w (and R_H), which results in colloids significantly larger in size. The D and g' of the hyperbranched core seem to affect the PDI of the final nanoparticle. H7-CA-DEX has a $D = 12.9$ and a $g' = 0.03$ (see Table 5.2) which indicate a very broad MWD and a high degree of branching. The corresponding PSD of this copolymer is also very broad with a PDI = 20.6. The direct correlation between the molecular weight properties of the hyperbranched core and the

final PSD of the colloidal nanoparticles provides a useful tool to tune the colloidal properties of the final core cross-linked micelle analogues.

TEM microscopy provides an insight in the colloidal structure of the poly(MMA-*co*-EGDMA)-*b*-DEX nanoparticles. Figure 5.9 presents the images of a 1 wt% suspension of H6-CA-DEX in distilled deionized water at two different magnifications.

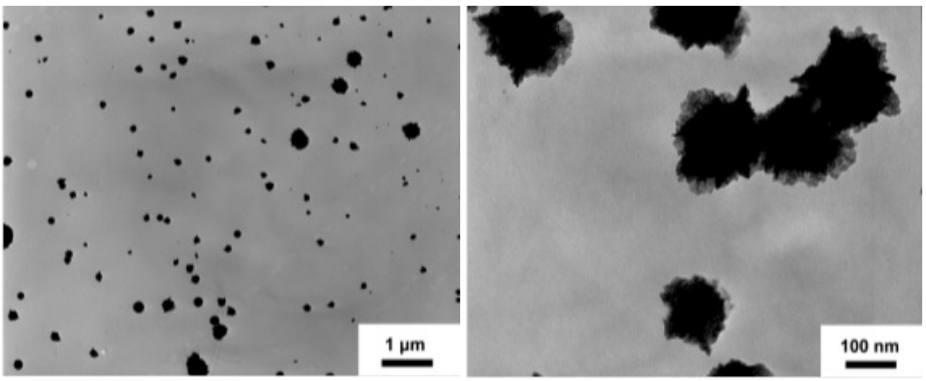


Figure 5.9. TEM photographs showing core-shell-like structure of H6-CA-DEX (1 wt.%) in distilled deionized water. Scale bars are 1 µm (left) and 100 nm (right).

It is apparent from Figure 5.9 that the final PSD is relatively broad and that a number of large colloids (~ 500 nm) can be observed. These large colloids were also observed in the DLS measurements (Figure 5.8) and can likely be attributed to the formation of aggregates. At higher magnification, the images reveal an irregularly shaped dark coloured core surrounded by a lighter grey coloured corona. The dark core surrounded by a light grey corona confirms that there is indeed a core-shell structure present: the dark core corresponds to the poly(MMA-*co*-EGDMA) hyperbranched core and the grey to the collapsed dextran corona. The particles in this micrograph have diameters of approximately 100 to 150 nm, significantly larger than the averages measured by DLS (Table 5.3). For the particles that were imaged, this is predominantly the case. The discrepancy between TEM and DLS could be accounted to the use of inaccurate refractive indices

in the DLS measurements or artefacts during TEM analysis (e.g. the relatively small number of micrographs taken).

The synthesized poly(MMA-*co*-EGDMA)-*b*-DEX copolymers synthesized here proved to yield colloidally stable nanoparticles that resemble the architecture of a core cross-linked micelle. The polymers can be readily obtained from a three-step synthetic approach using commercially available reagents and very low amounts of COBF. However, the relatively broad PSDs require optimization before these materials can be compared to block-copolymer micelles. The facile synthesis of these micelle-like materials has provided a basis for future work that will focus on narrowing the PSD and improving the colloidal stability of these nanoparticles.

5.4 Conclusions

In summary, a novel approach towards the synthesis of amphiphilic core cross-linked micelle analogues comprising a synthetic hyperbranched core and a polysaccharide shell is reported. The poly(MMA-*co*-EGDMA)-*b*-DEX copolymers are synthesized from a three-step synthetic approach: (*i*) CCTP to produce hyperbranched poly(MMA-*co*-EGDMA) with pendant double bonds, (*ii*) thiol addition with CA to yield an amine functionality and (*iii*) reductive amination with dextran.

The colloidal properties of the poly(MMA-*co*-EGDMA)-*b*-DEX polymers in water were found to be governed by the molecular weight properties of the hyperbranched core. The MWD and degree branching determine the average particle size and the width of the PSD. The amount of cobalt(II) and cross-linker used in the first step (synthesis of the core) determine the degree of functionality and therewith the amount of dextran that can be grafted onto the core. Since the polymerization conditions during CCTP govern the functionality and the MWD characteristics of the hyperbranched core, this synthetic step is crucial in controlling the colloidal properties of the

poly(MMA-*co*-EGDMA)-*b*-DEX polymers.

Once suspended in water, these polymers resemble the structural orientation of core cross-linked micelles. In this regard, the robust polymerization techniques and quantitative chemical modifications used here can be used create a diverse array of micelle-like structures circumventing the need for elaborate chemistry and/or self-assembly steps. Despite involving three synthetic steps, the approach described here provides an accessible route towards tailoring the properties of soft materials.

5.5 Acknowledgements

The authors would like to thank Dr. Xiaohu Yan for conducting TEM imaging. The authors would like to acknowledge the financial support provided by the Government of Ontario through the Ontario Research Chair Program (Cunningham). The authors would also like to acknowledge the Ministry of Research and Innovation for their support through the Early Researcher Award Program (Champagne) and Queen's University for their support by providing a Queen's Graduate Award and facilitating a G.E. Ted Courtnage Graduate Award in Engineering (Krasznai). The authors would like to thank the Courtnage family for establishing this award. A postdoctoral fellowship provided by the Ontario Ministry or Research and Innovation is also acknowledged for their financial support (Smeets).

5.6 References

(1) K. Matyjaszewski, *Science*, **2011**, 333, 1104–1105.
(2) L. Yang and P. Alexandridis, *Curr. Opin. Colloid Interface Sci.*, **2000**, 5, 132-143.
(3) M. Malmsten, *Softmatter*, **2006**, 2, 760–769.
(4) L Y. Qiu and Y.H. Bae, *Pharm. Res*., **2006**, 23, 1–30.
(5) D. Ruth, *Curr. Opin. Biotech*., **2011**, 22, 492–501.
(6) M. Talelli and W. E. Hennink, *Nanomedicine*, **2011**, 6, 1245-1255.
(7) S. J. Holder and N. A. J. M. Sommerdrijk, *Polym. Chem.*, **2011**, 2, 1018-1028.
(8) M. Yokoyama, *Expert. Opin. Drug Deliv.*, **2010**, 7, 145-158.

(9) C. J. Rijcken, C. J. Snel, R. M. Schiffelers, C. F. van Nostrum and W.E. Hennink, *Biomaterials*, **2007**, 28, 5581–5593.
(10) J. M. J. Fréchet, M. Henmi, I. Gitsov, S. Aoshima, M. R. Leduc, and R. B. Grubbs, *Science*, **1995**, 269, 1080–1083.
(11) D. Konkolewicz, M. J. Monteiro, and S. Perrier, *Macromolecules*, **2011**, 43, 5949-5955.
(12) B. Voit, *J. Polym. Sci. Part A: Polym. Chem.*, **2000**, 38, 2505–2525.
(13) H. Magnusson, E. Malmström, A. Hult, and M. Johansson, *Polymer*, **2002**, 43, 301–306.
(14) D. S. Thompson, L. J. Markoski, J. S. Moore, I. Sendijarevic, A. Lee, and A. J. McHugh, *Macromolecules*, **2000**, 33, 6412–6415.
(15) N. O'Brien, A. McKee, D. C. Sherrington, A. T. Slark, and A. Titterton, *Polymer*, **2000**, 41, 6027–6031.
(16) A. T. Slark, D. C. Sherrington, A. Titterton and I. K. Martin, *J. Mater. Chem.*, **2004**, 13, 2711-2720.
(17) S. V. Kurmaz, M. L. Bubnova, E. O. Perepelitsina and V. P. Roshchupkin, *Polym. Sci. Series A*, **2003**, 45, 201-208.
(18) N. M. B. Smeets, M. W. Freeman and T. F. L. McKenna, *Macromolecules*, **2011**, 44, 6701-6710.
(19) P. Chambon, L. Chen, S. Furzeland, D. Atkins, J. V. M. Weaver and D. J. Adams, *Polym. Chem.*, **2011**, 2, 941–949.
(20) M. Luzon, C. Boyer, C. Peinado, T. Corrales, M. Whittaker, L. Tao, and T. P. Davis, *J. Polym. Sci. Part A: Polym. Chem.*, **2010**, 48, 2783–2792.
(21) P. Besenius, S. Slavin, F. Vilela and D. C. Sherrington, *React, Funct, Polym.*, **2008**, 68, 1524–1533.
(22) A. A. Gridnev and S. D. Ittel, *Chem. Rev.*, **2001**, 101, 3611–3660.
(23) J. P. A. Heuts, G. E. Roberts, and J. D. Biasutti, *Aust. J. Chem.*, **2002**, 55, 381–398.
(24) J. P. A. Heuts and N. M. B. Smeets, *Polym. Chem.*, **2011**, 2**,** 2407-2423.
(25) T. P. Davis, D. M. Haddleton and S. N. Richards, *J. Macromol. Sci. Part C: Polym. Rev.*, **1994**, 34, 243–324.
(26) T. P. Davis, D. Kukulj, D. M. Haddleton and D. R. Maloney, *Trends Polym. Sci.*, **1995**, 3, 365–373.
(27) L. V. Karmilova, G. V. Ponomarev, B. R. Smirnov and I. M. Belgovskii, *Russ. Chem. Rev.*, **1984**, 53, 132-139.
(28) S. Camerlynck, P. A. G. Cormack, D. C. Sherrington and G. Saunders, *J. Macromol. Sci: Part B*, **2005**, 44, 881–895.
(29) S. V. Kurmaz, M. L. Bubnova, E. O. Perepelitsin and G. A. Estrina, *E-Polymer*, 2004.
(30) C. R. Becer, R. Hoogenboom and U. S. Schubert, *Angew. Chem. Int. Ed.*, **2009**, 48, 4900–4908.
(31) B. D. Mather, K. Viswanathan, K. M. Miller, and T. E. Long, *Prog. Polym. Sci*, **2006**, 31, 487–531.
(32) C. E. Hoyle, A. B. Lowe and C. N. Bowman, *Chem. Soc. Rev.*, **2010**, 39, 1355–1387.
(33) R. A. Evans, *Aus. J. Chem.*, **2007**, 60, 384-395.
(34) K. A. McEwan and D. M. Haddleton, *Polym. Chem.*, **2011**, 2, 1992–1999.
(35) J. P. Menzel, D. M. Haddleton and E. Khoshdel, *PMSE Prepr.*, **2010**, 102, 18–19.

(36) A. H. Soeriyadi, G. -Z. Li, S. Slavin, M. W. Jones, C. M. Amos, C. R. Becer, M. R. Whittaker, D. M. Haddleton, C. Boyer and T. P. Davis, *Polym. Chem.*, **2011**, 2, 815-822.
(37) L. Nurmi, J. Lindqvist, R. Randev, J. Syrett and D.M. Haddleton, *Chem. Commun.*, **2009**, 2727-2729.
(38) A. van der Ende, T. Croce, S. Hamilton, V. Sathiyakumar and E. Harth, *Softmatter*, **2009**, 5, 1417–1425.
(39) B. G. Davis and M. A. Robinson, *Curr. Opin. Drug. Disc. Dev.*, **2002**, 5, 279-288.
(40) B. Voit and D. Appelhans, *Macromol. Chem. Phys.*, **2010**, 211, 727–735.
(41) N. Röckendorf and T. K. Lindhorst, *Topics in Current Chemistry.*, **2001**, 217, 201–238.
(42) A. Pfaff, V. S. Shinde, Y. Lu, A. Wittemann, M. Ballauff, and A. H. E. Müller, *Macromol. Biosci.*, **2010**, 11, 199–210.
(43) R. K. Roy and S. Ramakrishnan, *Macromolecules*, **2011**, 44, 8398-8406.
(44) A. van der Ende, T. Croce, S. Hamilton, V. Sathiyakumar and E. Harth, *Softmatter*, **2009**, 5, 1417–1425.
(45) K. Knop, R. Hoogenboom, D. Fischer, and U. S. Schubert, *Angew. Chem. Int. Ed.*, **2010**, 49, 6288–6308.
(46) Z. Liu, Y. Jiao, Y. Wang, C. Zhou, and Z. Zhang, *Adv. Drug. Deliv. Rev.*, **2008**, *60*, 1650–1662.
(47) Y. C. Lee and R. T. Lee, *Acc. Chem. Res.*, **1995**, 28, 321–327.
(48) Y. Miura, *J. Polym. Sci. Part A: Polym. Chem.*, **2007**, 45, 5031–5036.
(49) A. Pfaff, A. Schallon, T. M. Ruhland, A. P. Majewski, H. Schmalz, R. Freitag and A. H. E. Müller, *Biomacromolecules*, **2011**, 12, 3805-3811.
(50) L. Zhang, J. Bernard, T. P. Davis, C. Barner-Kowollik and M. H. Stenzel, *Macromol. Rapid Commun.*, **2008**, 29, 123–129.
(51) A. Bakac, M. E. Brynildson and J. H. Espenson, *Inorg. Chem.*, **1986**, 25, 4108–4114.
(52) S. Beuermann, D. A. Paquet, J. H. McMinn and R. A. Hutchinson, *Macromolecules*, **1996**, 29, 4206–4215.
(53) R. A. Hutchinson, J. H. McMinn, D. A. Paquet, S. Beuermann and C. Jackson, *Ind. Eng. Chem. Res.*, **1997**, 36, 1103–1113.
(54) H. Galina and X. Zhu, Hyperbranched Polymers, John Wiley & Sons, Inc., Hoboken, NJ, USA, **2011**, 301–316
(55) B. H. Zimm and W. H. Stockmayer, *J. Chem. Phys.*, **1949**, 17, 1301–1314.
(56) A. Pillay Narrainen and P. A. Lovell, *Polymer*, **2010**, 51, 6115-6122.

Chapter 6

Conclusions and Recommendations for Future Work

6.1 Conclusions

Naturally produced polymers, such as those found within lignocellulosic biomass, are abundant, renewable, versatile, and inexpensive; however, these materials are relatively under utilized as high-value commercial polymers given recent developments in the field of polymer chemistry. Moreover, the variable composition of lignocellulosic biomass and the laborious nature of determining the composition of this natural material (see Chapter 2) further complicates the development of continuous processes devoted to processing this renewable material into high value copolymers. Recent developments in characterization, polymerization, and post-polymerization modification techniques have shown tremendous potential for producing polysaccharide-containing amphiphilic copolymers. This thesis described the synthesis and characterization of polymers containing synthetic and natural units.

The rapid compositional analysis of lignocellulosic biomass was made possible by building a multivariate regression model using calibration mixtures containing varying quantities of cellulose, hemicellulose, and lignin (see Chapter 3). These multivariate techniques could provide a rapid and effective alternative to the laborious wet-chemical methodologies that have traditionally been involved in the compositional analysis of lignocellulosic biomass. However, the calibration mixtures as well as the data pretreatment regime need to be carefully chosen to build an effective model.

In an effort to produce amphiphilic polysaccharide-containing copolymers, a series of polymer chemistries were explored, including ATRP, Cu(0)-mediated SET-LRP, and cobalt(II)

mediated CCTP. In addition, thiol addition and reductive amination chemistries were used as post-polymerization modifications following CCTP.

Homogenous SET-LRP was attempted from the reducing-end of dextran to give the amphiphilic block copolymer dextran-*b*-poly(methyl methacrylate) (see Chapter 4). The first step in the synthesis of this block copolymer was to generate a functional initiator that could be coupled, *via* reductive amination, to the reducing end of dextran. This functional initiator containing a highly reactive hydrazide functionality and was new to published literature. The functional initiator, 2-bromo-2-methylpropane hydrazide, was designed to allow for homogenous SET-LRP to be conducted from the chain-end of dextran or other polysaccharides containing a reducing-end. However, stability issues arising from the highly reactive hydrazide and alkyl bromide functionalities on the initiator unfortunately prevented the synthesis of the desired block copolymer. Despite these stability issues, the hydrazide functionality was shown to be highly reactive towards the reducing-end of dextran, which could be exploited in the future to improve the efficiency and yields of the reductive amination step.

Heterogeneous ATRP was conducted from the surface of Whatman No. 1 cellulosic filter paper to give the graft copolymer cellulose-*g*-poly(methyl methacrylate) (see Appendix 1). The cellulosic filter paper containing grafted poly(methyl methacrylate) showed hydrophobic properties, as determined by contact angle measurements. These types of hybrid materials containing synthetic polymers grafted from cellulose could potentially be used as membrane materials for wastewater treatment.

Cobalt(II) mediated CCTP was conducted using commercially available monomers to create a hyperbranched synthetic copolymer that was reminiscent of desirable polymer architectures like dendrimers or micelles (see Chapter 5). Dextran was introduced to the surface

of the hyperbranched synthetic copolymer to give an amphiphilic structure that immediately formed a colloidal suspension in water similar to other more intractable polymer architectures. This contribution provides an engineering approach towards the synthesis of novel colloidal structures that could be used for the targeted delivery of poorly water-soluble therapeutics.

This thesis also provides a basis for future work in this field, as will be outlined in the following recommendations section.

6.2 Recommendations

A number of recommendations for future work were generated from the conclusions of each original research chapter. The following section provides a summary of these recommendations categorized by the corresponding chapter.

Chapter 3 involved the development of an expedient and inexpensive multivariate regression model based on infrared spectra of lignocellulosic biomass. Recommendations for future work on the material presented in Chapter 3 include:

1. Evaluate the effect of particle size on the predictive performance of the multivariate model
2. Use model compounds for the calibration mixtures that are more representative of the species with unknown composition (e.g., use wood hemicellulose when attempting to predict the composition of wood samples or grass hemicellulose when attempting to predict the composition of grass samples)
3. Attempt to build an on-line process whereby compositional data is used to control the parameters of unit operations, such as chemical or physical pretreatment

Chapter 4 described the synthesis of a novel hydrazide-containing LRP initiator for the production of amphiphilic block copolymers. Despite the unsuccessful attempts at LRP from the

initiated reducing-end of dextran, a number of recommendations were made that might allow for this approach to be successful in the future. Recommendations for future work on the material presented in Chapter 4 include:

1. Produce a stable LRP initiator system that is highly reactive towards aldehydes (e.g., more reactive than an amine, but more stable than the hydrazide functionality explored in this work)
2. Evaluate techniques that increase the reactivity of the hemiacetal moiety at the reducing-end of the polysaccharide (e.g., microwave or more industrial relevant techniques)
3. The effect of mutarotation at the reducing-end facilitated by microwave irradiation as postulated by Pagnotta et al.[1] and Verma et al.[2] should be further explored as a means to improve reducing-end reactivity. Moreover, chemical methods, such as the use of alumina catalysis in dimethyl sulfoxide as described by Dunstan and Pincock[3], could provide a more industrially relevant alternative to microwave irradiation for the activation of reducing-ends.
4. Use a grafting-to approach as noted in Chapter 4 (Scheme 4.5) to produce amphiphilic copolymers (i.e., modify living polymer to present an amine-terminus with cysteamine, dissolve this polymer with the desired polysaccharide in a homogenous solvent and allow the reaction to proceed *via* reductive amination)
5. Synthesize a hydrazide-containing -ene or thiol that can be selectively coupled to the reducing-end of a polysaccharide to form a hydrazide and then couple a synthetic polymer *via* thiol addition
6. Produce a range of different molecular weight polysaccharides (e.g., alginate) *via* free-radical degradation to facilitate self-assembly after coupling with synthetic living polymer
7. With respect to SI-ATRP (Appendix 1), evaluate the use of the graft-modified filter paper in wastewater applications as e.g., a flocculating agent or a selective membrane

Chapter 5 described the synthesis and characterization of novel polysaccharide-stabilized micelle analogues. Recommendations for future work on the material presented in Chapter 5 include:

1. Produce a hyperbranched core with degradable units (e.g., using disulfide linkages that can be reduced using glutathione that is naturally produced *in vivo*)
2. Synthesize a responsive hyperbranched core copolymer to trigger the release of therapeutics (e.g., pH, temperature, infrared, and/or ultrasound responsive hyperbranched core)
3. Use amine-modified hyperbranched copolymers as a low-viscosity alternative to carbon dioxide switchable polymers developed by the Cunningham and Jessop groups

6.3 References

(1) Pagnotta, M.; Pooley, C. L. F.; Gurland, B.; Choi, M. *J. Phys. Org. Chem.* **1993**, 6, 407–411.
(2) Verma, M. S.; Gu, F. X. *Carbohydrate Polymers,* **2012**, DOI: 10.1016/j.carbpol.2011.11.025.
(3) Dunstan, T. D. J.; Pincock, R. E. *J. Org. Chem* **1985**, 50, 863–866.

Appendix 1

Surface Initiated Atom Transfer Radical Polymerization of Methyl Methacrylate from Cellulosic Filter Paper

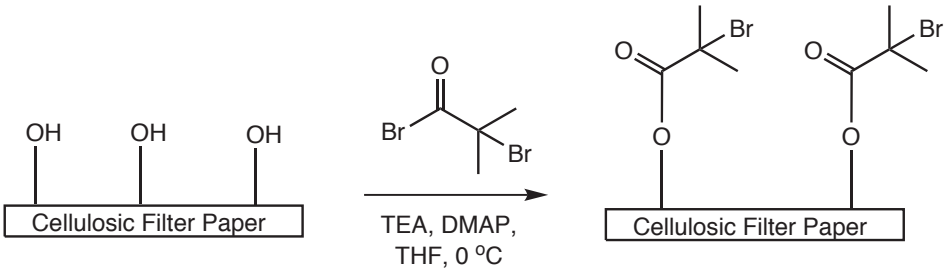
Introduction

Applying polymers to the surface of wood, metals, plastics, or glass can greatly improve the interfacial behaviour of these materials. Typically, surface modification involves coating the material with polymer films that dry to give the desired behaviour, such as corrosion resistance, lubrication, or antibacterial properties. Alternatively, polymer brushes can be grown from the surface of these materials by covalently functionalizing the surface chemistry.[1,2] For example, controlled/living radical polymerization techniques, such as Atom Transfer Radical Polymerization (ATRP)[3], can be used to "grow" well-defined polymers from the surface of materials.

Surface-initiated Atom Transfer Radical Polymerization (SI-ATRP) is a heterogeneous polymerization technique whereby alkyl halide functionalized surfaces serve as initiating species in conventional ATRP. A variety of surfaces and polymers are amenable to SI-ATRP, which results in hybrid materials that can be non-fouling or antibacterial[4,5,6] or may have cell-targeting capabilities[7] or glucose sensing moieties[8]. Growing polymer brushes from the surface of cellulosic substrates has also been demonstrated with a variety of polymerization chemistries, monomers, and initiating species.[9,10,11,12]

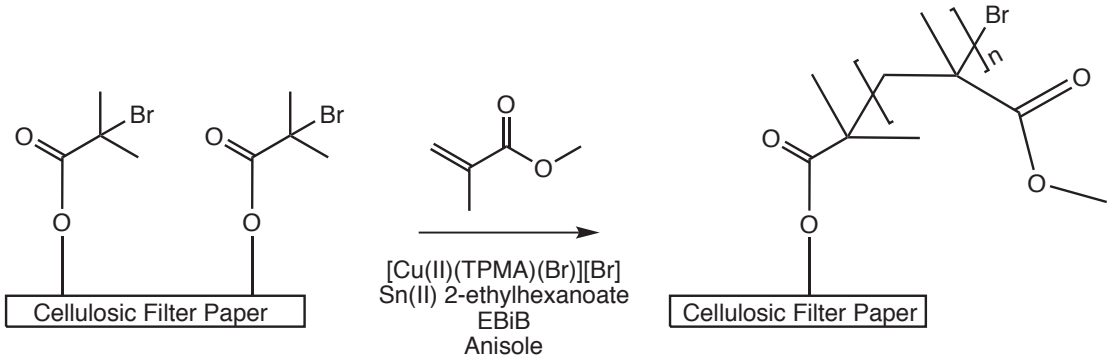
The following is a short investigation into the SI-ATRP of MMA from surface-initiated Whatman No. 1 filter paper using tris(2-pyridylmethyl)amine (TPMA) as the ligand and free ethyl α-bromoisobutryate (EBiB) to monitor the polymerization *via* size exclusion

chromatography (SEC). Sn(II) 2-ethylhexanoate was used as a reducing agent to continuously produce the activating [Cu(I)(TPMA)(Br)] species from the deactivating [Cu(II)(TPMA)(Br)][Br] species as per the Activators ReGenerated by Electron Transfer (ARGET) protocol.[13] First, the surface hydroxyl groups of the cellulosic filter paper were initiated with BriB, as shown in Scheme A1.



Scheme A1. Surface initiation of cellulosic filter paper using BriB for subsequent SI-ATRP.

Following surface initiation using BriB, SI-ATRP was conducted using MMA, [Cu(II)(TPMA)(Br)][Br] as the mediating species, Sn(II) 2-ethylhexanoate as the reducing agent, and anisole as the polymerization solvent, as shown in Scheme A2.



Scheme A2. SI-ATRP of MMA from the surface of cellulosic filter paper following surface initiation with BriB.

Solution-borne PMMA (i.e., initiated by EBiB) was used to monitor the progression of molecular weight as a proxy for surface-bound PMMA (i.e., initiated by surface-bound alkyl halide). A study by Hansson et al.[10] using a cleavable initiator showed that the chains grown from the surface of an initiated material have similar molecular weights to that of free initiator in solution. Thus, it was assumed that the molecular weight data determined *via* SEC in tetrahydrofuran (THF) of the solution-borne polymers were representative of those grown from the surface of the cellulosic substrate.

Materials

The following materials were used as received: α-bromoisobutyryl bromide (BriB, Aldrich, 98%), ethyl α-bromoisobutyrate (EBiB, Aldrich, 98%), tin(II) 2-ethylhexanoate (95%, Aldrich), anisole (Sigma-Aldrich, 99%), tetrahydrofuran (THF, Aldrich, 99%), methyl methacrylate (MMA, Aldrich, 99%), 4-(dimethylamino) pyridine (DMAP, Aldrich, 99%), triethylamine (TEA, Sigma-Aldrich, 99%), anhydrous ethanol (EtOH, Commercial Alcohols, anhydrous). Whatman No. 1 filter paper was washed with methanol and acetone and dried at 85 °C for 30 minutes before use. The copper(II) tris(2-pyridylmethyl)amine (TPMA) complex [Cu(II)(TPMA)Br][Br] was synthesized according to the methods of Britovsek et al..[14]

Methods

Surface Initiation of the Cellulosic Filter Paper

A small piece of previously dried filter paper (0.0529 g) was added to a 100 mL three-neck round bottom flask containing a magnetic stir bar. DMAP (0.16 g, 1.31 mmol) was then added followed by THF (20 mL) and TEA (1.1 mL, 7.89 mmol). The flask was then sealed and purged with nitrogen with magnetic stirring in an ice bath. BriB (1 mL, 1.87 g, 8.13 mmol) was

then diluted with THF (5 mL) and added to the flask *via* a degassed syringe. The reaction was left overnight under nitrogen. In the morning, the surface-initiated filter paper was removed from the flask and washed for 5 minutes with THF in a sonication bath. This washing procedure was then repeated with anhydrous ethanol. The washed filter paper was then dried overnight in a vacuum oven at 50 °C.

Surface-initiated Atom Transfer Radical Polymerization (SI-ATRP)

In a 100 mL three-neck round bottom flask containing a magnetic stir bar, the following materials were added in order: [Cu(II)(TPMA)(Br)][Br] (0.009 g, 0.0175 mmol, 0.1 eq), anisole (10 g), MMA (35 g, 350 mmol, 2000 eq.), and surface-initiated filter paper (0.0064 g). The flask was then purged with nitrogen for approximately 1 hour and heated to 70 °C in an oil bath. After the desired temperature was reached, EBiB (0.0341 g, 0.175 mmol, 1 eq.) was diluted with anisole (2.5 g) and added, *via* a degassed syringe, to the polymerization vessel. After 10 minutes, Sn(II) 2-ethylhexanoate (0.0708 g, 0.175 mmol, 1 eq.) was added to the system to reduce the deactivator [Cu(II)(TPMA)(Br)][Br] complex to the activator [Cu(I)(TPMA)(Br)] complex as per the ARGET protocol. The polymerization was allowed to proceed under these conditions for approximately 20 hours with periodic sampling for conversion estimation and SEC analysis. The polymerization mixture gelled during the overnight period. The filter paper was recovered and placed in THF and sonicated for 1 hour in a sonication bath. The THF was then discarded and fresh THF added, allowing the filter paper to soak overnight to remove PMMA that was not covalently linked to the surface.

Characterization

Size Exclusion Chromatography (SEC)

Polymer samples were dried, weighed, and dissolved in approximately 2 mL of THF to give a target concentration of about 20 $mg \cdot mL^{-1}$. After complete dissolution, the samples were filtered through 0.2 μm nylon syringe filters into auto-sampler vials for Size Exclusion Chromatography (SEC).

SEC was performed on free PMMA in solution using a Waters 2690 separation module and a model 410 differential refractometer. Five Waters Styragel HR columns (HR5.0, HR4.0, HR3.0, HR1.0, and HR0.5) in series were used at a constant temperature of 40 °C. Distilled THF was used as eluent at a flow rate of 1.0 mL min^{-1}. The LSEC system was calibrated using narrow molecular weight poly(styrene) standards ranging from 374 to 400 × 10^3 g mol^{-1}. The Mark-Houwink-Kuhn-Sakurada parameters used for the poly(styrene) standards were K = 1.14 × 10^{-4} $dL \cdot g^{-1}$ and a = 0.716.[15] Similarly, for poly(methyl methacrylate), K = 9.44 × 10^{-5} $dL \cdot g^{-1}$ and a = 0.719.[16]

Attenuated Total Reflectance Fourier Transform Infrared (ATR FT-IR) Spectroscopy

Fourier Transform Infrared (FT-IR) spectroscopy was performed on a Thermo Scientific Nicolet 6700 instrument using an Attenuated Total Reflectance (ATR) accessory equipped with a diamond crystal. A total of 64 scans were co-added per spectrum with a background spectrum of air taken between successive samples. Constant pressure was applied to the sample on the diamond crystal using a hand-tightened anvil.

Contact angle analysis

Contact angles were measured using a VCA Optima (AST Products Inc.) visual contact angle device with distilled deionized water (Millipore Synergy). The left and right contact angles were averaged over three samples for both the surface initiated filter paper and the PMMA grafted filter paper.

Results and Discussion

The molecular weight distribution (MWD) of the SI-ATRP was monitored using free initiator in solution (see comment in Introduction section regarding use of free polymer for SEC of SI-ATRP reactions). Figure A1 shows the progression of the MWD with respect to conversion for the SI-ATRP of MMA.

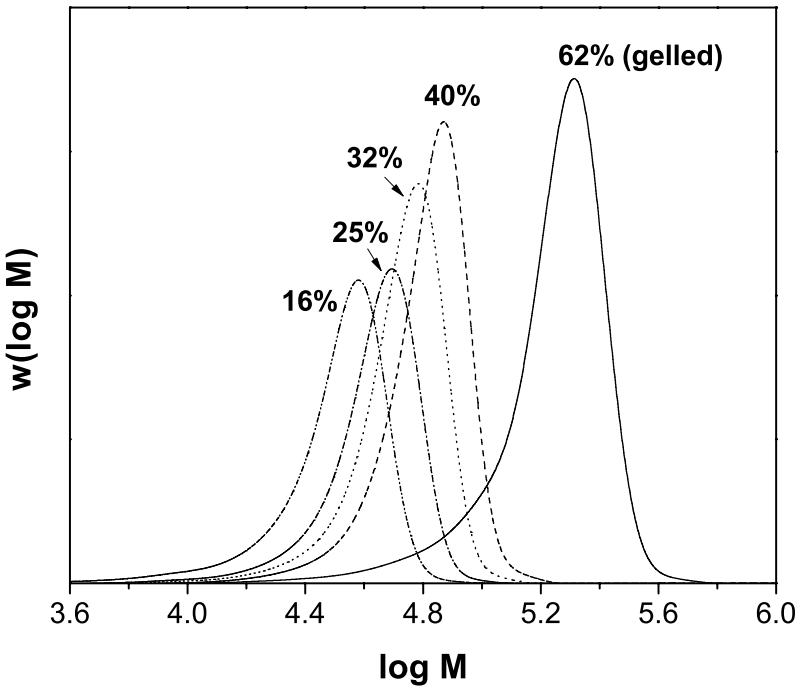


Figure A1. Progression of the MWD for the SI-ATRP if MMA from the surface of cellulosic filter paper. The MWD was determined by monitoring free polymer in solution (i.e., polymer initiated by EBiB).

The MWD was observed shift towards higher molecular weights as the reaction progressed. This increase in MWD with respect to conversion can be expected for controlled/living radical polymerizations. The MWDs were also narrow with dispersities (M_w / M_n) under 1.3 for the duration of the polymerization prior to gelation. Upon gelation, the dispersity reached a maximum of 1.5. The number average molecular weight (M_n) and dispersity (M_w / M_n) as a function of conversion are summarized in Figure A2.

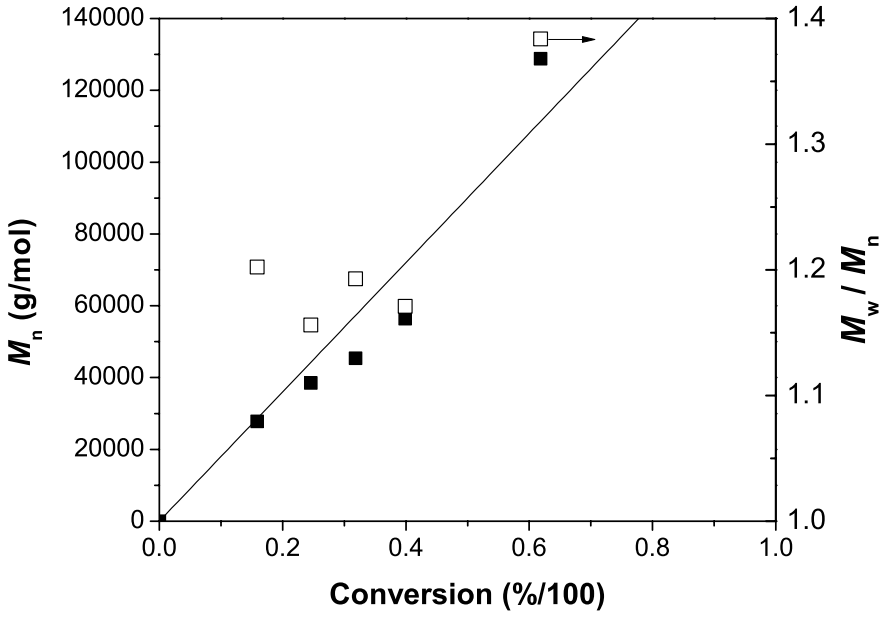


Figure A2. Profiles of M_n and dispersity (M_w / M_n) as a function of conversion for the SI-ATRP of MMA from surface initiated cellulosic filter paper. The line represents the theoretical M_n.

A linear increase of M_n as a function of conversion can be observed up until the point of gelation. The linear increase in M_n as a function of conversion is indicative of a controlled polymerization. Likewise, linear first-order kinetics were observed, as shown in Figure A3, indicating that this polymerization was indeed controlled/living.[17]

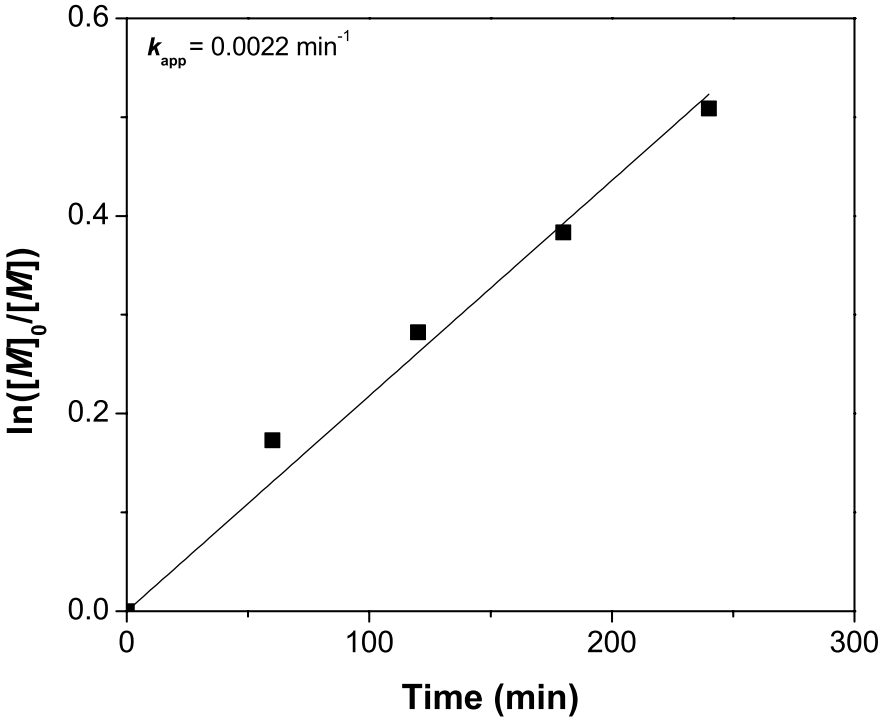


Figure A3. Semi-logarithmic plot of ln([*M*]$_0$/[*M*]) vs. time showing linear kinetics indicative of a controlled/living radical polymerization. A k_{app} of 0.0022 min^{-1} was calculated from the slope of this plot.

A linear first-order kinetic plot was observed for this SI-ATRP, giving an apparent rate coefficient of propagation (k_{app}) of approximately 0.0022 min^{-1}. The last data point (i.e., gelled polymer) was omitted since it could not be assumed that the radical concentration was constant upon gelation.

SEC data alone could not verify the successful grafting of PMMA chains from the surface of the initiated cellulosic filter paper. High temperatures and exotic solvents, such as ionic liquids or DMAc/LiBr, are required for the dissolution of cellulose; furthermore, the viscosity of these cellulose solutions is prohibitively high for most conventional SEC apparatus. Thus, the surface chemistry was analyzed qualitatively using Fourier Transform Infrared (FT-IR)

spectrometer equipped with a diamond Attenuated Total Reflectance (ATR) accessory. This ATR FT-IR device probes the surface chemistry of the material using the infrared spectrum and requires little to no sample preparation. The ATR FT-IR spectra of both virgin Whatman No. 1 cellulosic filter paper and the PMMA-grafted cellulosic filter paper are shown in Figure A4.

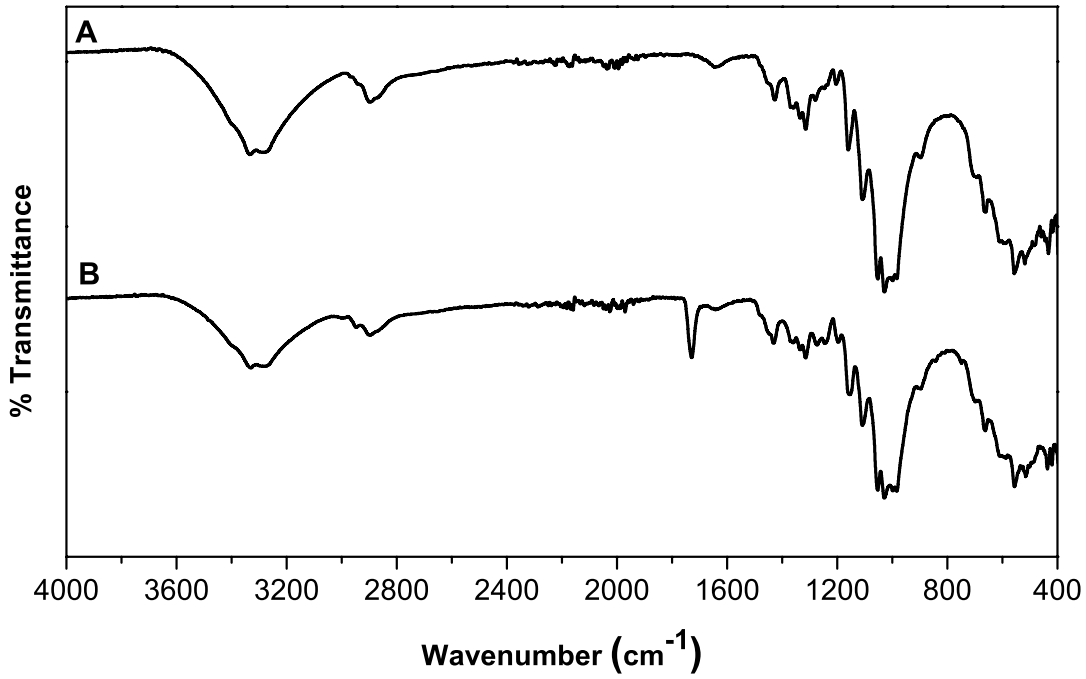


Figure A4. ATR FT-IR spectrum of virgin Whatman No. 1 cellulosic filter paper (A) and Whatman No. 1 filter paper with PMMA grafted from the surface *via* SI-ATRP (B). The pMMA surface grafts can be observed as a strong absorption peak at approximately 1750 cm^{-1} due to the C=O streching mode of the PMMA carbonyl.

The spectra of the virgin filter paper and the PMMA-grafted filter paper show most of the same characteristic vibration modes, except for a strong absorbance peak around 1750 cm^{-1} in spectrum B corresponding to the C=O stretching of the PMMA carbonyl on the surface of the PMMA-grafted filter paper. Because the PMMA-grafted filter paper was rigorously washed following SI-

ATRP, it was assumed that this C=O stretching mode was primarily due to surface-grafted PMMA chains.

The increase in hydrophobicity of this PMMA-grafted filter paper was investigated using contact angle measurements with distilled deionized water. The contact angle of water was observed to increase for the surface-initiated filter paper and further for the PMMA-grafted filter paper, as evidenced in the photographs shown in Figure A5.

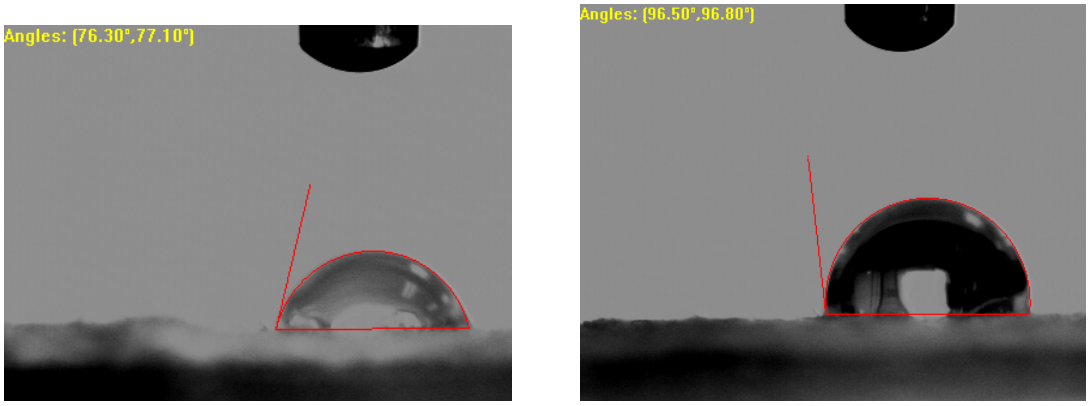


Figure A5. Contact angle of distilled deionized water with surface-initiated filter paper (left) and PMMA filter paper (right).

Virgin filter paper was observed to fully wet (and wick) the droplet of water immediately upon contact with the surface. However, upon introducing the alkyl bromide functionality (Scheme A1), a considerable increase in contact angle of 75 ± 4° was observed. Furthermore, the PMMA-grafted filter paper (Scheme A2) gave a contact angle with water of 98 ± 3°, which was significantly higher than that of the surface-initiated filter paper with alkyl bromide moieties. These contact angle results are summarized in Table A1.

Table A1. Summary of contact angles of distilled deionized water with filter paper, surface-initiated (SI) filter paper, and PMMA filter paper. Three sets of left and right contact angles were averaged for each entry.

Sample	Average Contact Angle (°)
Filter paper	Wetting
SI filter paper	75 ± 4
PMMA filter paper	98 ± 3

It is clear that the SI-ATRP of PMMA from the surface of Whatman No. 1 cellulosic filter paper produces a material with vastly different surface properties. Furthermore, because SI-ATRP is a controlled/living radical polymerization technique, this surface chemistry can be tuned *via* targeting different M_n values; by chain-extension to give block copolymers; or by substitution reactions at the terminal -Br group to give responsive functionalities.

Conclusions

The versatility and relative ease of SI-ATRP was demonstrated by grafting PMMA from the surface of Whatman No. 1 cellulosic filter paper. The polymerization proceeded *via* ARGET in a controlled/living manner to give a paper surface with PMMA characteristics. The surface-grafted PMMA endowed the filter paper with hydrophobic character – a sharp juxtaposition to its original hydrophilic behaviour.

Acknowledgements

The authors would like to acknowledge Nicky Chan for his help conducting these experiments and for synthesizing the [Cu(II)(TPMA)(Br)][Br] complex. The authors would like to thank Jeffery Wood and Dr. Aristedes Docoslis for their help with contact angle measurements. The authors would also like to acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC) for financial support through the Discovery Grant Program

(Champagne, Cunningham). The authors would like to acknowledge the Ministry of Research and Innovation for their support through the Early Researcher Award Program (Champagne) and Queen's University for their support by providing a Queen's Graduate Award and facilitating a G.E. Ted Courtnage Graduate Award in Engineering (Krasznai). The authors would like to thank the Courtnage family for establishing this award.

References

(1) S. Edmondson, V. L. Osborne, and W. T. S. Huck, *Chem. Soc. Rev.*, **2004**, 33, 14.
(2) B. Zhao and W. J. Brittain, *Progress in Polymer Science*, **2000**, 25, 677–710.
(3) J.-S. Wang and K. Matyjaszewski, *Macromolecules*, **1995**, 28, 7572–7573.
(4) H. Ma, J. Hyun, P. Stiller, and A. Chilkoti, *Adv. Mater.*, **2004**, 16, 338–341.
(5) X. Fan, L. Lin, and P. B. Messersmith, *Biomacromolecules*, **2006**, 7, 2443–2448.
(6) Z. Cheng, X. Zhu, Z. L. Shi, K. G. Neoh, and E. T. Kang, *Ind. Eng. Chem. Res*, **2005**, 44, 7098–7104.
(7) Hu, K. G. Neoh, L. Cen, and E.-T. Kang, *Biomacromolecules*, **2006**, 7, 809–816.
(8) F. J. Xu, Q. J. Cai, Y. L. Li, E. T. Kang, and K. G. Neoh, *Biomacromolecules*, **2005**, 6, 1012–1020.
(9) A. Carlmark and E. Malmström, *Journal of the American Chemical Society*, **2002**, 124, 900–901.
(10) S. Hansson, P. Antoni, H. Bergenudd, and E. Malmström, *Polym. Chem.*, **2011**, 2, 556.
(11) S. Hansson, T. Tischer, A. S. Goldmann, A. Carlmark, C. Barner-Kowollik, and E. Malmström, *Polym. Chem.*, **2012**.
(12) E. Malmström and A. Carlmark, *Polym. Chem.*, **2011**, DOI: 10.1039/c1py00445j.
(13) K. Matyjaszewski, W. Jakubowski, K. Min, W. Tang, J. Huang, W. A. Braunecker, and N. V. Tsarevsky, *Proceedings of the National Academy of Sciences*, **2006**, 103, 15309–15314.
(14) G. J. P. Britovsek, J. England, and A. J. P. White, *Inorg. Chem.*, **2005**, 44, 8125–8134.
(15) S. Beuermann, D. A. Paquet, J. H. McMinn, and R. A. Hutchinson, *Macromolecules*, **1996**, 29, 4206–4215.
(16) T. Gruendling, T. Junkers, M. Guilhaus, and C. Barner-Kowollik, *Macromol. Chem. Phys.*, **2010**, 211, 520–528.
(17) Tsarevsky, N. V.; Matyjaszewski, K. *J. Polym. Sci. A Polym. Chem.* **2006**, 44, 5098–5112.