Engineering Mixed Solvent Environments for Acid-Catalyzed Biomass Conversion Reactions

By

Theodore W. Walker

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

Doctor of Philosophy (Chemical Engineering)

at the

UNIVERSITY OF WISCONSIN–MADISON 2019

Date of final oral examination: 12/02/2019

The dissertation is approved by the following members of the Final Oral Committee:
George W. Huber, Professor, Chemical and Biological Engineering
James A. Dumesic, Professor, Chemical and Biological Engineering
Reid C. Van Lehn, Assistant Professor, Chemical and Biological Engineering
Victor M. Zavala, Professor, Chemical and Biological Engineering
Brian G. Fox, Professor, Biochemistry

Engineering Mixed Solvent Environments for Acid-Catalyzed Biomass Conversion Reactions

Theodore W. Walker
Under the supervision of Prof. James A. Dumesic and Prof. George W. Huber at the University of Wisconsin–Madison

Abstract

It has been shown that mixtures of water with polar aprotic cosolvents (*mixed solvent environments*) can be used to control the rates and selectivities of acid-catalyzed biomass conversion reactions. A quantitative understanding of these solvent effects would therefore enable a powerful measure of control over biomass conversion processes, but this framework is currently lacking. This dissertation combines reaction kinetics studies, spectroscopic methods and molecular dynamics investigations to explore the fundamental aspects underlying the reactivity of biomass-derived oxygenates in mixed solvent environments, and the solvent-mediated decomposition of real biomasses over acid catalysts.

Chapters 3 and 4 focus on the production of levoglucosenone (LGO) from cellulose in mixed solvent environments containing dilute sulfuric acid. In Chapter 3, we show that water mediates the interconversion of LGO and its isomer, 5-hydroxymethylfurfural (HMF). Following this insight, we demonstrate in Chapter 4 how the composition water/tetrahydrofuran mixtures can be modulated to control the selectivity of LGO and HMF as coproducts from cellulose conversion.

Chapters 5, 6 and 7 combine reaction kinetics studies and molecular dynamics (MD) investigations to probe the fundamental bases whereby mixed solvent environments control the rates and selectivities of acid-catalyzed biomass conversion reactions. In Chapter 5, we show that the rates of acid-catalyzed reactions of biomass-derived oxygenates generalize with the MD-observable properties of water-enriched local solvent domains that nucleate in the immediate

vicinity of solvated reactant molecules. In Chapter 6, we demonstrate how the selectivity of acid-catalyzed biomass conversion reactions can be partially understood in terms of the thermodynamic relationships between solvated reactant and product molecules in mixed solvent environments. In Chapter 7, we distill these insights into a model-predictive framework that allows for the rational design of mixed solvent environments for biomass conversion processes using computationally efficient screening methods and minimal experimentation.

In Chapter 8, we use solid-state nuclear magnetic resonance (NMR) spectroscopy to explore how mixed solvents containing acid catalysts alter the properties cellulosic structures in real biomass. We show that these NMR-observable properties predict the extent to which hydrolytic enzymes are able to depolymerize the residual cellulose into glucose. This dissertation is concluded with a discussion of future directions.

Acknowledgements

It has been one of the foremost privileges of my life to pursue graduate studies in chemical engineering at the University of Wisconsin – Madison. The chance to dedicate years of one's life to the pursuit of fundamental scientific knowledge, and to acquisition of the necessary skills to that end, is a transformative one that is afforded to precious few, though many more must certainly be eager for, and deserving of the same. Accordingly, it is with humility and gratitude that I dedicate this thesis, the culmination of my serendipitous and privileged years here in Wisconsin, to those whom have been critical to the efforts underlying it.

First and foremost, I thank my advisors, Professors George Huber and James Dumesic, for their mentorship, patience and generosity over the years. In academic research, one's ideas are a precious resource, and sharing one's ideas honestly and completely is an act of trust and confidence. The advisor-student relationship is therefore built entirely upon trust: the advisor imparts their scientific world view to the student, trusting that the student will represent it faithfully to the community at large; the student imbibes this world view, trusting that it will empower them to solve challenging, original problems going forward. I thank my advisors for entrusting their own scientific worldviews to me, and I promise to represent them faithfully, to the best of my abilities. I thank Prof. George Huber for teaching me to prioritize research questions in the context of real and timely societal needs, and to avoid oversimplifying inherently complex scientific and engineering problems to the point where key insights are overlooked. I thank Prof. James Dumesic for teaching, practical motivations aside, to approach research with the singular goal of elucidating the fundamental, scientific details underlying the behaviors being studied, recognizing that these insights constitute the true, lasting value of our work. In particular, I thank Prof. James Dumesic for training me to temper my use of "Ted-sims," and to engage meaningfully with the theoretical

aspects of my research through collaborations, always aspiring to exemplify the qualities of a "whole-assed experimentalist," as opposed to a "half-assed theorist who also does experiments."

I thank my committee for taking the time to evaluate my candidacy for a Ph.D. in Chemical Engineering. Standing before an assembled panel of accomplished experts to lay one's ideas and scientific abilities bare is an opportunity for growth and reflection that I do not take likely. I thank my undergraduate advisors from the University of Illinois at Chicago, in particular Professors Randall Meyer, Vivek Sharma, Lew Wedgewood, and Chancellor Michael Amiridis, for first instilling me with an interest in academic research, and for investing so selflessly in preparing me for a graduate career in Chemical Engineering. I am truly indebted.

I thank my lab mates in the Huber and Dumesic groups for their friendship, guidance, and many, many insightful discussions over the years. In particular, I thank those who suffered and tolerated me daily, especially Dr. Siddarth Krishna, Dr. Nat Eagan, Dr. Dan McClelland, Dr. Insoo Ro, Peter Galebach, Mark Lindsay, Dr. Kevin Barrnett, Dr. Joe Chada, Dr. Pranav Kumbharkar, Dr. Ali Hussain Motagamwala, Dr. Maddie Ball, Anthony Anderson and Dr. Zach Brentzel. I especially thank Dr. Max Mellmer, who's graduate work inspired me, and formed the basis for much of my own research. I hope you approve of how I have built upon your ideas. I thank my undergraduate research assistants, especially Liam Witteman and Nathan Frelka, who taught me so much, and who accomplished so much in their time in the Huber and Dumesic Labs, despite me. I'll miss all of you, and I hope that we stay in touch. I wish you all the best, and I am sure you will all do very well.

I thank my collaborators and co-authors for their contributions to the exciting research we've carried out tougher, both published and unpublished. In particular, I'd like to thank Dr. Siddarth Krishna, Mark Lindsay, Dr. Kefeng Huang, Alex K. Chew, Prof. Reid C. Van Lehn,

Nathaniel Kuch, Dr. Kirk Van Der Meulen, Prof. Brian Fox, Dr. Catherine Clewett, Dr. Heike Hoffstetter and Dr. Charlie Fry, all of whose contributions appear extensively in this dissertation. I look at what we have learned together and, comparing it to what very little I would have been able to establish on my own, I know that my most precious resource going forward will not be state-of-the-art instruments, nor generous research grants, but relationships. I hope my interactions with you all over the past four years have not spoiled my chance to partner with you in research in the future!

I thank the academic staff in the UW-Madison Department of Chemical Engineering, especially Judy Lewison, Dr. James Miller, Steven Schumacher, Dr. Eric Codner, Mary Heimbecker, Michelle Kirch, Lynn Robinson, Kathy Heinzen, Teresa Neisius, Oscar Garcia-Romero, Susann Eva Ely, Kate Fanis, Russ Poyner and Janet Frederick. We students often come to you at our worst: frustrated, overwhelmed and helpless. In the midst of all that, I doubt it is obvious to you that your hard work, creativity and generosity does not go unnoticed by us students. As such, I'd like to take this opportunity to say that we do; we do notice. Thank you.

I'd like to take a step back from the department, and thank those who's friendships and shared experiences have shaped my life so profoundly: my fellow warriors in A Co., 2nd Ranger Battalion, and the Afghan National Army. To Sgt. Evans, Staff Sgt. Newville, Sgt. Mayes, Staff Sgt. Harpool, Sgt. First Class Mills, Litzenburger, Tommy Guns, Pedro, Kaiser, Kirkland, Meatwad, Silkwood, Marci, Ammons, Staff Sgt. Purdy, Sgt. Campos, Sgt. Martinez, Sgt. Baker, Sgt. "Paul T." Baker, Staff Sgt. Anderson, Doc Hormer, Doc Morgan, Doc Bonk, Dang-ole Jerimiah Johnson, Capt. McCarthy, Major Keller, First Sgt. Schwartz, Sgt. Reid, Staff Sgt. Reid, Sgt. Johnson… I meant what I said to you all when I left: I don't think I'll ever do anything more important that what we did together. I didn't leave because I didn't believe in what we were doing.

I left because I would have slowed you down, and I'm less in the way elsewhere. I'll never forget you guys. I'll never forget what we did, what we lost, and what we learned together. I know you don't like being thanked for what you do, but that's tough. The rest of the world will never know the half of it, so I'm saying it every chance I get: thank you for what you do. Robo says thank you. Mahmood, Saeed… if I knew how to even begin looking for you, I would. I hope you are still alive. You are some of the bravest men I've known. I hope that you get the peace and the country you are fighting for. I hope we meet again. Joel Clarkson…. I think about you every day. I think about the last time I saw you. I think about how happy you were. No matter what I do in life, I'll never be half as good as what you are right now. The first chapter of this thesis is for the Airborne Ranger in the Sky. Rangers Lead the Way.

I thank my mother and my father, Jill and Grant Walker, for having raised me, and for their continued, tireless presence in my life. I know I did not make things easy on you, but you never gave up on me. Looking back on my life so far, I see that Lena, Kaline and I were always the most important thing in the world to you. It showed in the sacrifices you made, in the love you gave, and in the way you tirelessly and mercilessly endeavored to improve yourselves so that you could provide for us and be there for us. I promise to do the same for my own family. I thank my sisters for growing up with me, and for generally being awesome. I love and admire you both. To my parents and siblings, the fifth chapter of this thesis is for you.

I thank my boys, Alexander and Ira Xavier. No matter what accomplishments are attributed to me in my life, and no matter what small laurels or trifling accolades are laid upon me, if any, I want you two to know that the only reason I do *anything*, is for you. I want the world to be a better place for you to inherit, and I want your lives to be filled with love, purpose, joy and wonder. I'd do anything for that. I love you so much that I don't recognize myself, and I'm so grateful to be

your dad I don't always know what to do with it; it's awesome and overwhelming and utterly disarming. All I know is that I will never stop trying to be better for you. Please don't ever lose your silliness, despite what I've said while trying to write this thesis from the living-room/bedroom/playroom/laundry-room/office/dinning-room of our tiny graduate student apartment! The eighth chapter of this thesis is for you both.

Lastly, and most of all, I want to thank my Wife, Julia. I wouldn't be here without you. I love you in every way that I know how. You have taught me *what* love is, and for that I am yours forever. You have saved my life more times than I can count. You've given your heart to me, then endured me, forgiven me, and waited so patiently and bravely for me to become the kind of man who is deserving of it. I am not, and I have so far to go, but I promise that I will never stop trying to deserve it. You believed in me when I had nothing, and for that I promise that whatever I have is yours, and yours alone for the rest of my life. You gave me our children, and you sacrificed so much to do so. I promise there is nothing I will not sacrifice for you. You've brought magic into my life and opened my eyes to new wonders. I promise to make new adventures for us, and to open your eyes to new wonders. You are everything to me. You astound and fascinate me. You guide me and inspire me. I can't take my eyes off of you. I promise I never will. I love you baby. The final chapter of this thesis is for you.

The research presented in this thesis was supported in part by: the Department of Energy, Office of Energy Efficiency and Renewable Energy (EERE); by the U.S. Department of Energy, Office of Basic Energy Sciences; by the DOE Great Lakes Bioenergy Research Center, through a cooperative agreement between the Board of Regents of the University of Wisconsin and the U.S. Department of Energy; and by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison with funding from the Wisconsin Alumni Research Foundation. The

computational work in this thesis used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation. This work also used the computing resources and assistance of the UW-Madison Center for High Throughput Computing (CHTC) in the Department of Computer Sciences. The CHTC is supported by UW-Madison, the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation, the Wisconsin Institutes for Discovery, and the National Science Foundation, and is an active member of the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science.

On Wisconsin.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1. Lignocellulosic Biomass as a Renewable Carbon Resource

The use of fossil resources as feedstocks for fuels and chemicals production is an unsustainable strategy to meet global energy and material demands due to the finite nature of natural reserves and their associations with the deleterious effects of climate change. Accordingly, there exists a timely and urgent need to displace global dependence on petroleum, coal, and other non-renewable feedstocks. In this context, lignocellulosic biomass, the non-edible portion of woody plant matter, represents one of the only realistic, near-term sources of organic carbon.(*1, 2*)

The major representative structures in biomass are cellulose (semi-crystalline arrays of linear glucose-based polymers, ~40-50 wt% of dry biomass)(*3*); hemicellulose (amorphous hexosan- and pentosan-based copolymers, ~25-30 wt%)(*4*); and lignin (complex, cross-linked phenolic oligomers by ~15-25 wt%).(*5*) These structures are displayed schematically in Figure 1.1.1., along with important, representative products and chemical intermediates that can be produced therefrom. Furanics afforded from acid-catalyzed reactions of hemicellulose have garnered interest as replacements for polymer precursors and fuel additives currently derived from petroleum.(*6*) Lignin represents the only non-petroleum-based, naturally abundant and renewable source of aromatics;(*7*) cellulose is a non-food-competitive source of glucose monomers, whose well-defined structures makes them a promising source of organic carbon for the large-scale production of food additives, chiral synthons(*8*) and, in principle, even liquid fuels.(*9*)

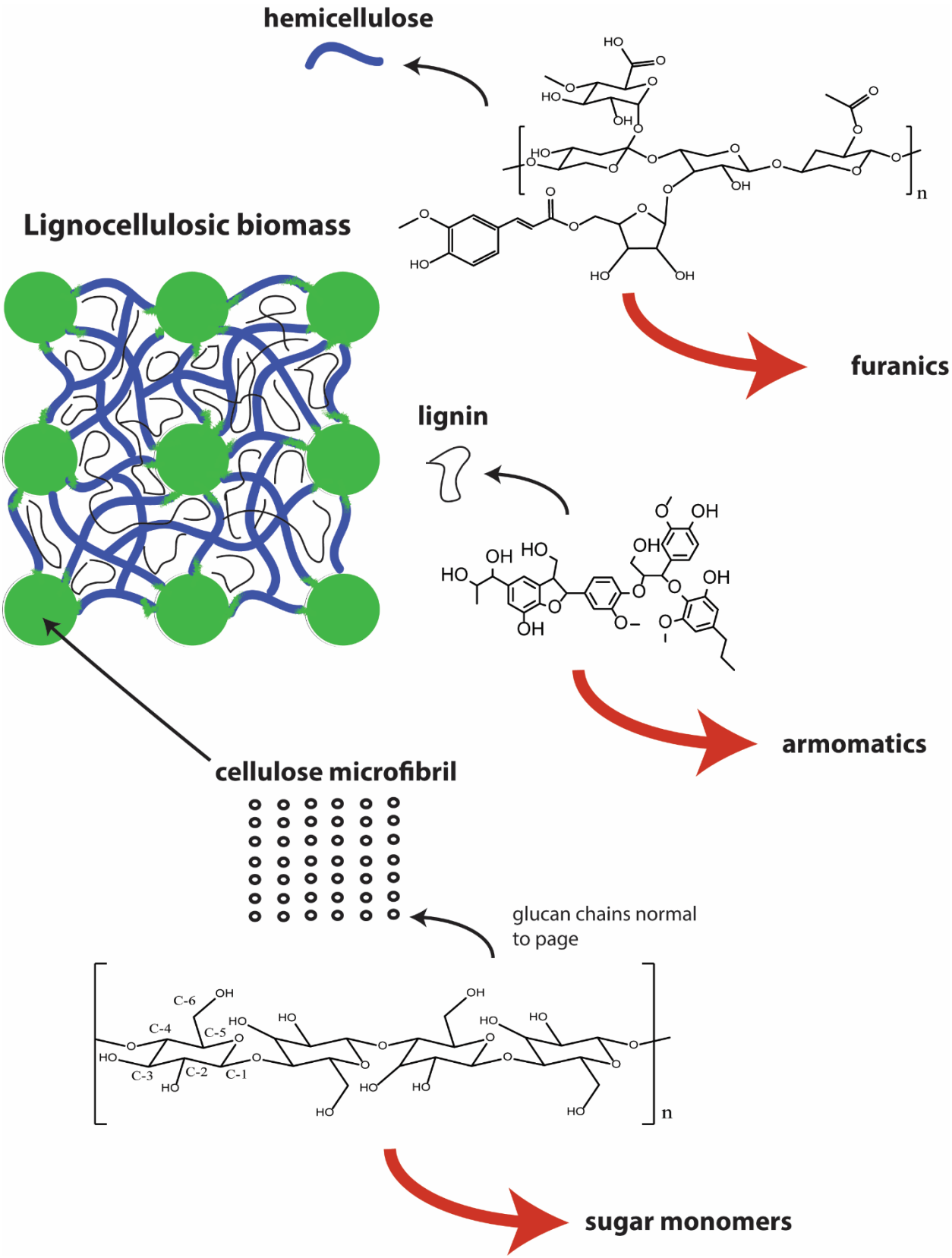


Figure 1.1.1. Representative structures in lignocellulosic biomass and exemplary product class that can be derived therefrom.

Accordingly, the past ten years has seen significant interest and investment in new technologies to convert biomass into renewable fuels and chemicals. However, the representative body of industrial efforts to this end has faced significant challenges, with several startup companies having commercialized biomass conversion technologies but having struggled to reach commercial scale; in fact, many of these companies have filed for bankruptcy.(*10, 11*) These outcomes are attributable to the recalcitrant nature of the feedstock, and to financial and engineering challenges associated with bring new and proven chemical process technologies to market; a process typically characterized by large capital costs and decade-long time frames.(*11*)

Nevertheless, several emerging catalytic technologies have either entered the market place, or are currently demonstrating their technologies in fully integrated pilot plants. Commercial and near commercial technologies for non-cellulosic-ethanol-based biomass conversion technologies to-date include, among others: small scale production of renewable jet and diesel from landfill gases *via* Fischer-Tropsch synthesis;(*12*) catalytic conversion of carbohydrates into gasoline and aromatics;(*13*) hydropyrolysis of biomass into gasoline and diesel,(*14*) and; catalytic conversion of wood into aromatics.(*15*)

To be successful, however, biomass conversion technologies must ultimately be able to compete economically with petroleum technologies, which are already operating at large commercial scales and have been practiced for decades. By any measure, a critical metric of the economic potential of a process is its efficiency with respect to the yield of products from raw materials, and this metric of performance is directly related to atom efficiency of the underlying chemical reactions. Therefore, while promising biomass conversion technologies continue to demonstrate progress towards commercialization, there remains an important and timely need to

better elucidate the detailed chemistries underlying catalytic reactions of real biomass-derived materials, and to use these insights to drive the development of more selective and atom-economical processes for converting biomass feedstocks into fungible chemical and fuels.

1.2. The Role of Mixed Solvent Environments in Acid-Catalyzed Biomass Conversion Reactions

In contrast to petroleum, biomass is a solid, non-volatile material that contains up to 50 wt% oxygen, which must be removed *via* chemical transformations to afford higher-value products or fuels that are compatible with current infrastructure.(*16*) A promising strategy to this end is to decompose solid biomass into its constituent cellulose, hemicellulose and lignin fractions using solvents, and to remove the oxygen from the resulting chemical residues in the form of water using acid catalysts.(*17*) However, the partially deoxygenated products derived from such processes, which involve multiple different reactions occurring both in series and in parallel, are as chemically diverse as biomass itself, and are therefore difficult to control with respect to selectivity towards specific products.

An important variable to control the reactivity and selectivity for these acid-catalyzed processes is the solvent composition, which affects reaction rates,(*18, 19*) product selectivities,(*20, 21*) the stability of desired products.(*22-24*) Prior studies have shown that aqueous mixtures containing polar aprotic cosolvents (*i.e.*, mixed-solvent environments) are of particular interest in this context; a minimum amount of water is often required to facilitate the solvation of biomass-derived materials, while the cosolvent can enhance reaction performance.(*25, 26*) Table 1.2.1 displays some of these results. For example, the rate of xylose dehydration to furfural in aqueous mixtures of γ-valerolactone increases 30-fold compared to the same reaction in pure water, while the formation of undesired humins *via* degradation of the reactant and product is suppressed,(*19*) improving the furfural selectivity

by over 50%. In another example, mixtures of water with γ-valerolactone are able to solubilize all fractions of lignocellulosic biomass, affording high yields of water-soluble carbohydrates from raw biomass.(*27*) Also, Rinadli and coworkers demonstrated that semi-aqueous cosolvent mixtures lead to superior yields of soluble products from lignin hydrogenolysis.(*28, 29*) The forgoing examples demonstrate how a fundamental knowledge of solvent effects in biomass conversion reactions should empower the rational design of *tunable liquid phase processes* for the selective conversion of feedstocks with *diverse chemical compositions*.(*30*)

Table 1.2.1. Exemplary reports detailing the use of mixed solvent environments to convert biomass-derived feedstocks into platform chemicals.

Feedstock (wt% solids)	Products	Solvent components	Temperature / ºC	catalyst	Yields / carbon %	ref
Corn stover, (1-7 wt%)	Xylose and glucose monomers and oligomers	GVL/water,	160-220	Sulfuric acid	70-90	(*27*)
Xylose (1 wt%)	Furfural	GVL/water, Dioxane/water, THF/water	145-175	Sulfuric acid	75	(*19*)
Lignin, bio oil (2 wt%)	arenes	IPA/water	160	Raney Ni, zeolites	70-80	(*28*)
Poplar wood (5 wt%)	Phenolic monomers	EtOH/water, MeOH/water	200	Pd/C, H_3PO_4	20-50	(*31*)
Fructose (10-50 wt%)	HMF	DMSO/water	90-180	HCl, H_3PO_4, ion exchange resins	25-72	(*20*)
Xylose (10 wt%)	furfural	Methyl-isobutyl-ketone/water	150	HCl, sulfuric acid, $AlCl_3$, others	45-90	(*32*)
Glucose, fructose, sucrose	HMF, levunilic acid	GVL/water	130	Sulfuric acid	50-75	(*33*)

(1-5 wt%) fructose (25 wt%)	HMF	Acetone/water	120	Sulfuric acid, HCl, others	85	(*34*)
White birch	Cellulose fibers, lignin residues, furfural	GVL/water	100-225 (multi-step process)	Sulfuric acid	~80%	(*35*)

1.3. Fundamentals of Solvation Effects in Acid-Catalyzed Biomass Conversion Reactions: State-of-the-Art and Challenges

Catalysis was once considered a "black art" in that the mechanistic details underlying many important catalytic processes, including some practiced at large scales, were virtually unknown.(*36*) However, owing to the development of advanced spectroscopic and computational methods, it is now possible to resolve molecular-level details regarding how elementary reaction sequences proceed on a catalyst's surface,(*37, 38*) and to produce accurate, *a priori* estimates of the corresponding reaction barriers.(*39, 40*) These molecular details are often combined into quantitative structure-property relationships that empower the rational design of new catalytic materials, tailored for specific applications.(*41-43*) Accordingly, catalyst research and design has assumed a key role in addressing challenges across the full spectrum of societal needs.(*44*)

In contrast, there exist no well-established theoretical tools to predict catalytic performance in arbitrary solvent systems. The reason for this is two-fold: solvent-solute interactions are difficult to characterize experimentally;(*45*) and the details of how these interactions translate to changes in reaction barriers are not well understood.(*46*) In general, solvent effects can be understood at three different length scales: bulk effects (on the order of meters) that can be accounted for in terms of tabulated quantities such as activity coefficients;(*47*) molecular-scale effects involving semi-structured solvent-solute adducts (on the order of nanometers); and atomistic interactions between

solvent molecules and activated complexes (on the order of Angstroms). Molecular scale events involve the transient formation of characteristic structures of solvent molecules about a substrate that can induce gradients in the chemical properties between the bulk fluid and the solute.(*48*) Atomistic events involve steric and electronic interactions between solvent molecules and the actual bond-breaking or –forming events that constitute elementary reaction steps.(*49*)

Many attempts have been made to explain solvent effects at these different length by theoretical treatment of experimental results. These studies have taken one of three typical forms, as summarized in Figure 1.3.1: mean-field models based on the bulk properties of the solvent or acid catalyst;(*50-52*) *ab initio* molecular dynamics (MD) simulations;(*53, 54*) and DFT simulations.(*55, 56*)

Mean-field models are the most simplistic, and attempt to correlate kinetic effects to tabulated, bulk properties of the solvent (such as the dielectric constant). In some cases, these models have been fit to experimental data for select reactions with well-defined mechanisms.(*50*) However, the implicit assumptions that underlie their formulation mean that they are not capable of providing general insight.

MD studies can access simulation scales that provide detailed information about how solvent and solute molecules arrange themselves in least-energy configurations. For example, Vlachos and co-workers showed how the oxygen atoms of a fructose molecule in solutions of water and dimethyl sulfoxide (DMSO) are preferentially solvated via hydrogen bonds with neighboring water molecules. It was speculated that the initial proton-transfer may be promoted in this case, thus increasing the rate of HMF formation in DMSO/water mixtures relative to pure aqueous solvent, in agreement with experimental results.(*57*) However, due to their inability to

directly model key bond-forming events, these types of simulations are limited in the types of mechanistic insights that they can provide.(*53*)

DFT studies in liquid-phase reaction environments can yield detailed free energy landscapes and corresponding mechanistic insights as to the role of the solvent for simple reaction mechanisms.(*58*) However, in the often mechanistically complex transformations of larger molecules (such as carbohydrates), these detailed simulations seldom produce consistent results between research groups.(*56, 59*) In all three cases: mean-field, MD, and DFT studies; the common limiting factor that diminishes the applicability of theoretically derived results for liquid-phase reactions is the absence of directly comparable, intrinsic reactivity data for well-defined, model systems against which the predictive power of the simulations can be benchmarked.

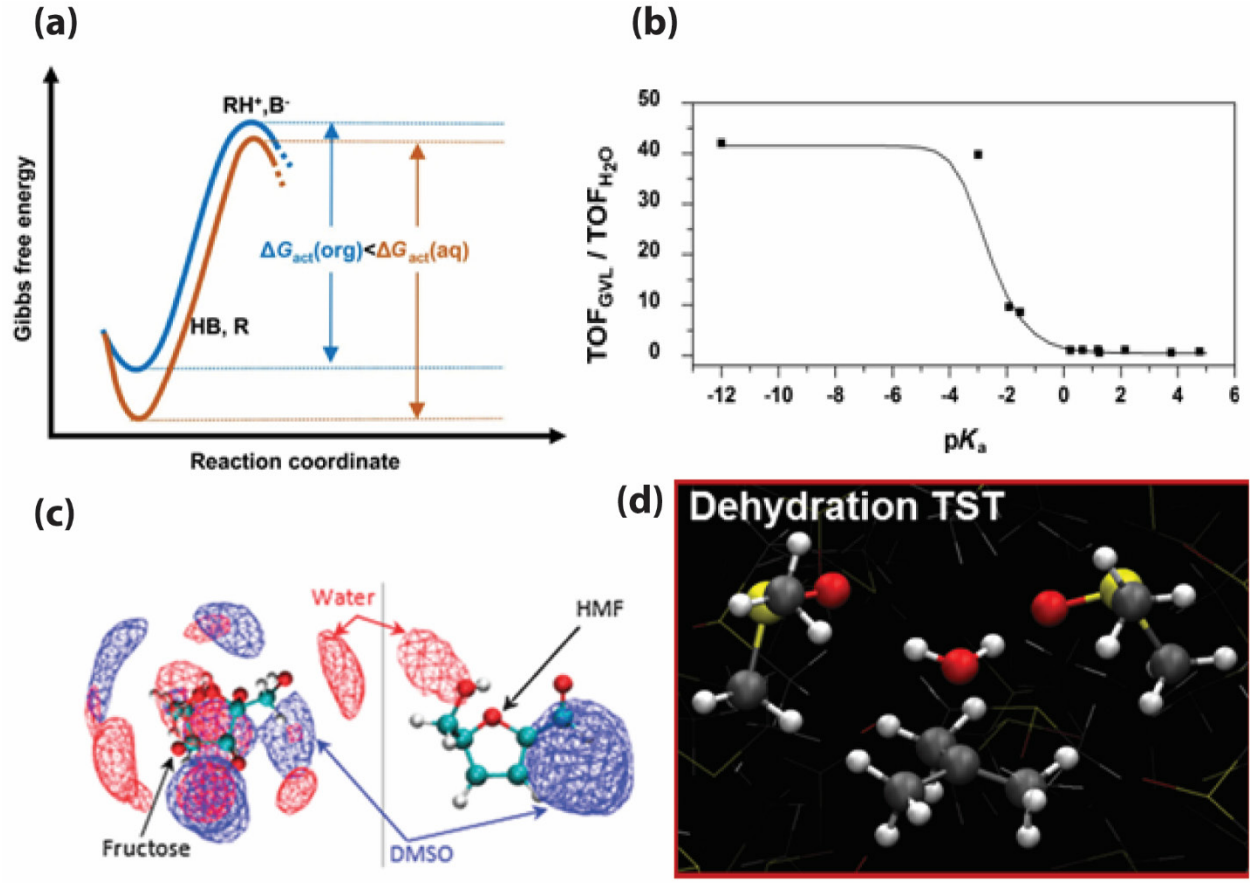


Figure 1.3.1. Exemplary frameworks for understanding solvation effects in acid-catalyzed reaction mechanisms across multiple length and time scales. Implicit consideration of (a) - the organic cosolvent's role in destabilizing the acidic proton, and (b) - the stability of conjugate base (pK_a) in controlling the per-proton turnover frequency (TOF) in the acid-catalyzed dehydration of xylose to furfural.(*19*) (c) - using classical molecular dynamics to explore the preferential solvation of the hydroxyl moieties of biomass-derived compounds by water and a dimethyl sulfoxide (DMSO) cosolvent and implications for the selective of fructose to 5-hydroxymethylfurfural in DMSO/water mixtures.(*57*) (d) - *ab initio* molecular dynamics simulations of *tert*-butanol dehydration (an analog for biomass-derived oxygenates) in DMSO/water mixtures.(*58*)

Our understanding of the role mixed solvents in mediating liquid-phase chemical processes can therefore be compared to the state of catalyst science and technology forty years ago, in that experimental efforts have identified optimal solvent compositions for key applications; while, in parallel, state-of-the-art quantum chemical methods and advanced spectroscopic techniques have produced important, molecular-level insights regarding the role of the solvent system in representative case studies. However, in contrast to gas-phase reactions over heterogeneous catalysts, a general framework to distill these insights into predictive structure-property relationships is lacking, and it is therefore not generally possible to anticipate how the composition of the liquid phase will affect the rates and selectivities of acid-catalyzed biomass conversion reactions in mixed solvents *a priori*. Therefore, designing mixed solvents for biomass conversion processes currently involves laborious, empirical screening of the continuous space of possible solvent compositions.

1.4. The Scope of this Dissertation

The aforementioned results demonstrate how first-principles based computational methods are able to decompose solvation effects into contributions by fundamental, molecular scale interactions, providing detailed insight as to the role of the solvent in controlling the reactivity of biomass-derived molecules in mixed solvent environments. Furthermore, recent advances in analytical techniques provide tools to better understand the chemistry and structure of native biomass and the reaction products derived therefrom. However, *ab intitio* quantum chemical investigations are limited in their ability to produce broadly-applicable insights that can be leveraged for solvent design purposes, and spectroscopic methods have had limited success in connecting the structure of native biomass to the factors that promote the selective, liquid-phase

conversion of biomass into valuable products. In this dissertation, we discuss some of our efforts to partially bridge these gaps by answering two fundamental questions:

(1) on what general bases do mixed solvent environments control the rates and selectivities of acid-catalyzed reactions of biomass-derived oxygenates, and how can these effects be predicted quantitatively using computationally efficient methods and a manageable number of experiments?
(2) How do mixed solvents mediate the decomposition of real biomass over acid catalysts, and what factors connect the structure and reactivity of the resulting materials as compared to native biomass?

In Chapter 2, we summarize the experimental, analytical and computational methods used to carry out the research summarized in the following chapters of this dissertation. Chapters 3 and 4 focus on the acid-catalyzed conversion of cellulose to levoglucosenone (LGO) and 5-hydroxymethyl furfural (HMF) in mixed solvent environments. Chapter 3 explores the kinetics and mechanistic details underlying water-mediated interconversion of LGO and HMF over acid catalysts. In Chapter 4, we demonstrate how the role of water in this mediating this interconversion, as well as the reactivity of key reaction intermediates can be leveraged to control the selectivity of GLO and HMF in the acid-catalyzed dehydration of cellulose in tetrahydrofuran (THF)/water mixtures. Importantly, when a small amount of water is added to the THF cosolvent, HMF replaces LGO as the major product of cellulose dehydration, but the combined yield of LGO and HMF as co-products from cellulose conversion is improved as compared to when wither species is targeted as a sole product. These results demonstrate how a quantitative knowledge of solvents effects in

acid-catalyzed biomass conversion reaction can afford a powerful measure of control over the reactivity of the underlying catalytic cascades.

In Chapters 5, 6 and 7 report combined reaction kinetics studies and classical MD investigations into the reactions of biomass-analogous model oxygenates in mixed solvent environments. In Chapter 5, we demonstrate how the reactivity of biomass-derived molecules in mixtures of water with THF, dioxane and GVL universally increase with decreasing water content of the solvent system and increasing extent of the reactants solvent-accessible surface area occupied by hydroxyl groups. We use the MD results to explain this behavior in terms of water-enriched or –deficient local solvent domains that nucleate in the vicinity of solvated reactant molecules. In Chapter 6, we use experimental reaction kinetics studies and MD-based perturbation free energy calculations to demonstrate how the selectivity of acid-catalyzed dehydration reactions in mixed solvent environments can be partially understood in terms of the thermodynamic relationships between the solvated reactant and product molecules as a function of solvent composition. In Chapter 8, we distill the forgoing insights into a generalized work flow to estimate solvent effects in acid-catalyzed biomass conversion reactions using minimal experiments and computational methods, partially enabling the rational design of mixed solvent environments for biomass conversion processes, and alleviating the laborious, experimentally-driven exercise that typically accompanies the development of new processes.

In Chapter 9, we use a combination of solid-state NMR, X-ray diffraction (XRD) and vibrational spectroscopy methods (FTIR) to characterize the effects of GVL-mediated biomass pretreatments on the cellulosic structures in real biomass, and compare these effects to the enzymatic digestibility of same. Based on the NMR, XRD and FTIR results, we suggest that the solvent pretreatments at low temperatures (90-140ºC) partially destroy the intimate linkages

between extracellular material (hemicellulose and lignin) and the solvent-accessible facets of cellulose microfibrils in the cell wall, while not altering the ultrastructure or characteristic dimensions of the semi-ordered cellulose domains. We show that behavior renders the residual cellulose better accessible to hydrolytic enzymes (cellulases), so that combined GVL-based solvent pretreatments and enzymatic hydrolysis of biomass affords improved yields of soluble sugars as compared to thermochemical or enzymatic methods alone.

This dissertation is concluded with future directions in Chapter 9, wherein we discuss how the insights presented in the previous sections can be extended to designing solvent systems for converting real biomass-derived materials (bio oils) into value added chemicals (arenes and aromatics), and to deconstructing complicated heteropolymer materials (multilayer plastic packaging films) into easily recyclable, isolated resins. Finally, we provide an outlook promising research directions into solvent effects in liquid-phase processes.

1.5 References

1. C.-H. Zhou, X. Xia, C.-X. Lin, D.-S. Tong, J. Beltramini, Catalytic conversion of lignocellulosic biomass to fine chemicals and fuels. *Chemical Society Reviews* **40**, 5588-5617 (2011).
2. M. J. Climent, A. Corma, S. Iborra, Conversion of biomass platform molecules into fuel additives and liquid hydrocarbon fuels. *Green Chemistry* **16**, 516-547 (2014).
3. A. L. Barnette *et al.*, Quantification of crystalline cellulose in lignocellulosic biomass using sum frequency generation (SFG) vibration spectroscopy and comparison with other analytical methods. *Carbohydrate polymers* **89**, 802-809 (2012).
4. A. Ebringerová, Z. Hromádková, T. Heinze, in *Polysaccharides i*. (Springer, 2005), pp. 1-67.
5. R. Vanholme, B. Demedts, K. Morreel, J. Ralph, W. Boerjan, Lignin biosynthesis and structure. *Plant physiology* **153**, 895-905 (2010).
6. A. S. Mamman *et al.*, Furfural: Hemicellulose/xylosederived biochemical. *Biofuels, Bioproducts and Biorefining: Innovation for a sustainable economy* **2**, 438-454 (2008).
7. Z. Strassberger, S. Tanase, G. Rothenberg, The pros and cons of lignin valorisation in an integrated biorefinery. *RSC Advances* **4**, 25310-25318 (2014).
8. F. Cao *et al.*, Dehydration of cellulose to levoglucosenone using polar aprotic solvents. *Energy & Environmental Science* **8**, 1808-1815 (2015).

9. E. L. Kunkes *et al.*, Catalytic conversion of biomass to monofunctional hydrocarbons and targeted liquid-fuel classes. *Science* **322**, 417-421 (2008).
10. B. E. Gaddy, V. Sivaram, T. B. Jones, L. Wayman, Venture capital and cleantech: The wrong model for energy innovation. *Energy Policy* **102**, 385-395 (2017).
11. L. Dixon *et al.*, Process and Outcome Evaluation of the Alternative and Renewable Fuel and Vehicle Technology Program. (2017).
12. M. M. Bomgardner, Seeking Better Biofuels. *Chemical & Engineering News* **93**, 8-10 (2015).
13. W.-C. Wang *et al.*, "Review of biojet fuel conversion technologies," (National Renewable Energy Lab.(NREL), Golden, CO (United States), 2016).
14. O. Winjobi *et al.*, Carbon Footprint Analysis of Gasoline and Diesel from Forest Residues and Algae using Integrated Hydropyrolysis and Hydroconversion Plus Fischer–Tropsch (IH2 Plus cool GTL). *ACS Sustainable Chemistry & Engineering* **6**, 10766-10777 (2018).
15. P. U. Karanjkar *et al.*, Production of aromatics by catalytic fast pyrolysis of cellulose in a bubbling fluidized bed reactor. *AIChE Journal* **60**, 1320-1335 (2014).
16. G. W. Huber, S. Iborra, A. Corma, Synthesis of transportation fuels from biomass: chemistry, catalysts, and engineering. *Chemical reviews* **106**, 4044-4098 (2006).
17. J. N. Chheda, G. W. Huber, J. A. Dumesic, Liquid-phase catalytic processing of biomass-derived oxygenated hydrocarbons to fuels and chemicals. *Angewandte Chemie International Edition* **46**, 7164-7183 (2007).
18. M. A. Mellmer, D. M. Alonso, J. S. Luterbacher, J. M. R. Gallo, J. A. Dumesic, Effects of γ-valerolactone in hydrolysis of lignocellulosic biomass to monosaccharides. *Green Chemistry* **16**, 4659-4662 (2014).
19. M. A. Mellmer *et al.*, Solvent effects in acid-catalyzed biomass conversion reactions. *Angewandte chemie international edition* **53**, 11872-11875 (2014).
20. Y. Román-Leshkov, J. N. Chheda, J. A. Dumesic, Phase modifiers promote efficient production of hydroxymethylfurfural from fructose. *Science* **312**, 1933-1937 (2006).
21. Z. Wei, Y. Li, D. Thushara, Y. Liu, Q. Ren, Novel dehydration of carbohydrates to 5-hydroxymethylfurfural catalyzed by Ir and Au chlorides in ionic liquids. *Journal of the Taiwan Institute of Chemical Engineers* **42**, 363-370 (2011).
22. M. A. Mellmer, J. M. R. Gallo, D. Martin Alonso, J. A. Dumesic, Selective production of levulinic acid from furfuryl alcohol in THF solvent systems over H-ZSM-5. *ACS Catalysis* **5**, 3354-3359 (2015).
23. Y. J. Pagan-Torres, T. Wang, J. M. R. Gallo, B. H. Shanks, J. A. Dumesic, Production of 5-hydroxymethylfurfural from glucose using a combination of Lewis and Brønsted acid catalysts in water in a biphasic reactor with an alkylphenol solvent. *Acs Catalysis* **2**, 930-934 (2012).
24. C. M. Cai, T. Zhang, R. Kumar, C. E. Wyman, THF co-solvent enhances hydrocarbon fuel precursor yields from lignocellulosic biomass. *Green Chemistry* **15**, 3140-3145 (2013).
25. B. Mostofian *et al.*, Local phase separation of co-solvents enhances pretreatment of biomass for bioenergy applications. *Journal of the American Chemical Society* **138**, 10869-10878 (2016).

26. T. Y. Nguyen, C. M. Cai, R. Kumar, C. E. Wyman, Co-solvent pretreatment reduces costly enzyme requirements for high sugar and ethanol yields from lignocellulosic biomass. *ChemSusChem* **8**, 1716-1725 (2015).
27. J. S. Luterbacher *et al.*, Nonenzymatic sugar production from biomass using biomass-derived γ-valerolactone. *Science* **343**, 277-280 (2014).
28. X. Wang, R. Rinaldi, A route for lignin and bio-oil conversion: dehydroxylation of phenols into arenes by catalytic tandem reactions. *Angewandte Chemie International Edition* **52**, 11499-11503 (2013).
29. X. Wang, R. Rinaldi, Solvent effects on the hydrogenolysis of diphenyl ether with Raney nickel and their implications for the conversion of lignin. *ChemSusChem* **5**, 1455-1466 (2012).
30. L. Shuai, J. Luterbacher, Organic solvent effects in biomass conversion reactions. *ChemSusChem* **9**, 133-155 (2016).
31. T. Renders *et al.*, Synergetic effects of alcohol/water mixing on the catalytic reductive fractionation of poplar wood. *ACS Sustainable Chemistry & Engineering* **4**, 6894-6904 (2016).
32. F. Tao, H. Song, L. Chou, Efficient process for the conversion of xylose to furfural with acidic ionic liquid. *Canadian Journal of Chemistry* **89**, 83-87 (2010).
33. L. Qi *et al.*, Catalytic conversion of fructose, glucose, and sucrose to 5-(hydroxymethyl) furfural and levulinic and formic acids in γ-valerolactone as a green solvent. *Acs catalysis* **4**, 1470-1477 (2014).
34. A. H. Motagamwala, K. Huang, C. T. Maravelias, J. A. Dumesic, Solvent system for effective near-term production of hydroxymethylfurfural (HMF) with potential for long-term process improvement. *Energy & Environmental Science*, (2019).
35. D. M. Alonso *et al.*, Increasing the revenue from lignocellulosic biomass: Maximizing feedstock utilization. *Science advances* **3**, e1603301 (2017).
36. T. H. Maugh, Catalysis: no longer a black art. *Science (Washington, DC);(United States)* **219**, (1983).
37. J. Dumesic, H. Topsøe, S. Khammouma, M. Boudart, Surface, catalytic and magnetic properties of small iron particles: II. Structure sensitivity of ammonia synthesis. *Journal of Catalysis* **37**, 503-512 (1975).
38. C. Ledesma, J. Yang, D. Chen, A. Holmen, Recent approaches in mechanistic and kinetic studies of catalytic reactions using SSITKA technique. *ACS Catalysis* **4**, 4527-4547 (2014).
39. L. Fan, T. Ziegler, Nonlocal density functional theory as a practical tool in calculations on transition states and activation energies. Applications to elementary reaction steps in organic chemistry. *Journal of the American Chemical Society* **114**, 10890-10897 (1992).
40. A. A. Gokhale, S. Kandoi, J. P. Greeley, M. Mavrikakis, J. A. Dumesic, Molecular-level descriptions of surface chemistry in kinetic models using density functional theory. *Chemical Engineering Science* **59**, 4679-4691 (2004).
41. J. K. Nørskov, T. Bligaard, J. Rossmeisl, C. H. Christensen, Towards the computational design of solid catalysts. *Nature chemistry* **1**, 37 (2009).
42. J. K. Nørskov *et al.*, Universality in heterogeneous catalysis. *Journal of catalysis* **209**, 275-278 (2002).
43. L. Zhang *et al.*, Platinum-based nanocages with subnanometer-thick walls and well-defined, controllable facets. *Science* **349**, 412-416 (2015).

44. G. Centi, S. Perathoner, Catalysis: role and challenges for a sustainable energy. *Topics in Catalysis* **52**, 948-961 (2009).
45. U. Halim *et al.*, A rational design of cosolvent exfoliation of layered materials by directly probing liquid–solid interaction. *Nature communications* **4**, 2213 (2013).
46. Y. Zhao, D. G. Truhlar, Hybrid meta density functional theory methods for thermochemistry, thermochemical kinetics, and noncovalent interactions: the MPW1B95 and MPWB1K models and comparative assessments for hydrogen bonding and van der Waals interactions. *The Journal of Physical Chemistry A* **108**, 6908-6918 (2004).
47. J. E. Leffler, E. Grunwald, *Rates and equilibria of organic reactions: as treated by statistical, thermodynamic and extrathermodynamic methods*. (Courier Corporation, 2013).
48. A. A. Hassanali, J. Cuny, V. Verdolino, M. Parrinello, Aqueous solutions: state of the art in ab initio molecular dynamics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **372**, 20120482 (2014).
49. J. V. Burda, A. Toro-Labbe, S. Gutierrez-Oliva, J. S. Murray, P. Politzer, Reaction force decomposition of activation barriers to elucidate solvent effects. *The Journal of Physical Chemistry A* **111**, 2455-2457 (2007).
50. D. Y. Murzin, Solvent effects in catalysis: implementation for modelling of kinetics. *Catalysis Science & Technology* **6**, 5700-5713 (2016).
51. S. Behtash, J. Lu, E. Walker, O. Mamun, A. Heyden, Solvent effects in the liquid phase hydrodeoxygenation of methyl propionate over a Pd (1 1 1) catalyst model. *Journal of catalysis* **333**, 171-183 (2016).
52. P. Gogoi, S. Hazarika, N. N. Dutta, P. G. Rao, Laccase catalysed conjugation of catechin with poly (allylamine): Solvent effect. *Chemical Engineering Journal* **155**, 810-815 (2009).
53. S. H. Mushrif, V. Vasudevan, C. B. Krishnamurthy, B. Venkatesh, Multiscale molecular modeling can be an effective tool to aid the development of biomass conversion technology: A perspective. *Chemical Engineering Science* **121**, 217-235 (2015).
54. S. Hamad, C. Moon, C. R. A. Catlow, A. T. Hulme, S. L. Price, Kinetic insights into the role of the solvent in the polymorphism of 5-fluorouracil from molecular dynamics simulations. *The Journal of Physical Chemistry B* **110**, 3323-3329 (2006).
55. S. Caratzoulas, D. G. Vlachos, Converting fructose to 5-hydroxymethylfurfural: a quantum mechanics/molecular mechanics study of the mechanism and energetics. *Carbohydrate Research* **346**, 664-672 (2011).
56. L. Yang, G. Tsilomelekis, S. Caratzoulas, D. G. Vlachos, Mechanism of Brønsted Acid-Catalyzed Glucose Dehydration. *ChemSusChem* **8**, 1334-1341 (2015).
57. S. H. Mushrif, S. Caratzoulas, D. G. Vlachos, Understanding solvent effects in the selective conversion of fructose to 5-hydroxymethyl-furfural: a molecular dynamics investigation. *Physical Chemistry Chemical Physics* **14**, 2637-2644 (2012).
58. M. A. Mellmer *et al.*, Solvent-enabled control of reactivity for liquid-phase reactions of biomass-derived compounds. *Nature Catalysis* **1**, 199-207 (2018).
59. X. Qian, Mechanisms and Energetics for Acid Catalyzed β-d-Glucose Conversion to 5-Hydroxymethylfurfurl. *The Journal of Physical Chemistry A* **115**, 11740-11748 (2011).

Chapter 2. Experimental and Computational Methods

2.1. Batch reaction studies (glass reactors)

Reactions kinetics studies in chapters 3, 5, 6 and 7 were carried out in closed, 10 mL thick-walled glass reactors (Alltech), and sealed with Teflon liners (Alltech) inserted into plastic caps.. In a typical experiment, an appropriate amount of reactant (*e.g.*, xylitol), acid catalyst, water and organic cosolvent (*e.g.*, *1,4*-dioxane) were charged to the reactors, which were then sealed with and placed in a 2500 mL crystallization dish filled with silicon oil at the appropriate temperature. Heating was effected by a stirring hot plate (Thermo Scientific). Reactors and the silicon oil heating medium were stirred at 500 rpm by a triangular and magnetic stir bars, respectively. Reactors were removed at times corresponding to the desired reaction time, and quenched in an ice bath at 273 *K*. The total liquid feed volume was 2 mL per 10 mL glass reactor in each experiment. Trifluoromethane sulfonic (triflic) acid ($pK_{a,\ H2O}$ = -14.7, $pK_{a,\ DMSO}$ = -14.3, $pK_{a,\ MeCN}$ = 0.7) was used as catalyst in all experiments, which has been shown to behave as a strong acid even in pure polar aprotic solvents.(*1, 2*) We thus assume complete dissociation of the acidic proton in all mixed solvent environments, allowing for normalization of the apparent rate constants on a per-proton basis.

For the cellulose pretreatment experiments discussed in Chapter 8, 6 grams of as-received 5-mm biomass chips (*e.g.*, P39 poplar) were loaded into a 60- mL pressure-sealed glass reactor equipped with a magnetic stir bar. 60 grams of an appropriate solvent system (*e.g.*, 80 wt% GVL with 20 wt% water) containing 100 mM sulfuric acid were then added, and the glass reactor was sealed with a screw-top cap equipped with a Teflon O-ring. The glass reactor was then submerged in a silicon oil bath at the desired temperature (*e.g.*, at 140ºC). The reactor contents were agitated

by rotating the stir bar at 500 rpm. The reactor was held at the desired temperature for a period of time (*e.g.*, one hour), and then quenched in an ice bath at $0^{o}C$ for 15 minutes. The reactor contents were then removed, and the liquid and solid fractions were separated by vacuum filtration with a fine-mesh cellulose filter paper, and a Buchner funnel connected to a house vacuum system at -15 in. Hg.

2.2. Batch reaction studies (metal autoclaves)

Reactions were carried out in a 100 mL autoclave (Parr Instrument Company, series 4560, Hastelloy (C-276)). The vessel and head were dried overnight at 70 ˚C to remove residual water prior to each reaction. The reagents, including cellulose (Avicel® PH-101, moisture content ca. 3 wt%), glucose (Sigma Aldrich, anhydrous), LGA (TCI, purity 99%), THF (Acros, anhydrous, 99.9%, stabilized with BHT), and sulfuric acid (Fisher Chemical, A300-500), were used as received. Reactants, solvent, and sulfuric acid were added to the autoclave sequentially, at which point the vessel was purged five times with helium (99.999%, Airgas) and charged to 3.4 MPa He. The vessel was then heated to the desired reaction temperature and pressurized to a final pressure of 1000 psi He. Zero time was defined as the point at which the heating was started, except where noted. The temperature and stirring were maintained at 700 rpm and controlled by a Parr 4848 Controller. Around 1 mL samples were taken periodically through a dip tube during the reaction. The reactor was re-pressurized with helium to 1000 psi after each sampling. The samples were immediately quenched in a dry ice bath and then filtered with a 0.22 μm syringe filter (Restek, PTFE (polytetrafluoroethylene)) prior to the analysis.

2.3. Analytical Methods

Reaction products in Chapters 3-8 were analyzed using high-performance liquid chromatographs equipped with differential refractometers and photo diode array detectors, or gas chromatographs equipped with flame ionization detectors. Details regarding data acquisition methods, equipment configurations, calibrations and sample prep are discussed in the corresponding chapters.

Attenuated total reflection–Fourier transform infrared spectroscopy (ATR–FTIR) results in Chapter 8 were collected using a Bruker Optics Vertex system with a diamond-germanium ATR single reflection crystal. Untreated and solvent-pretreated biomass samples were dried in a vacuum oven overnight to remove water content prior to analysis, and were pressed uniformly against the diamond surface using a spring-loaded anvil. Sample spectra were obtained in triplicates using an average of 128 scans over the range between 400 cm^{-1} and 4000 cm^{-1} with a spectral resolution of 2 cm^{-1}. Air was used as background.

Cellulose microfibrils scatter incident X-rays to produce a diffraction pattern consistent with a monoclinic unit cell.(*3*) Accordingly, X-ray diffractograms presented in Chapter 8 were collected for the native P39 and solvent-pretreated biomass samples, using a Bruker D8 Discovery diffractometer with a Cu K_a X-ray source operating at 1000 kV and 100 mA, with a 5-mm aperture and 600 sec exposure time. Biomass samples were analyzed without additional grinding to reduce particle size. Samples were self-adhered, when necessary, with vacuum grease. A separate sample of boron nitride was analyzed to assess the inherent line broadening of the instrument.

2.4. Data analysis

Rate constants in Chapters 3, 5, 6 and 7 were determined by solving nonlinear differentials equations describing the volume-averaged rate of reactant conversion or product formation per unit time using linear least squares regression of the reaction kinetics data in MATLAB using the nlinfit regression function (Levenberg–Marquardt non-linear least squares algorithm). Confidence intervals are reported at the 95% confidence level (nparci function). Reactions kinetics data in Chapters 4, 6 and 7 were also determined by the methods of initial rates in the kinetic regime at less ~20% conversion.

2.5. Nuclear Magnetic Resonance Spectroscopy (NMR)

2.5.1. Solution Phase NMR

Solution phase ^{1}H- and ^{13}C-NMR results discussed in Chapter 3 were collected using a Bruker Avance 500 MHz spectrometer equipped with a liquid nitrogen cooled, broad-band fluorine-observe (BBFO) cryogenic probe at room temperature. 10% D_2O was added to the solution to facilitate locking to the deuterium resonance. ^{1}H spectra were acquired using a recycle delay of 3.0 s with 30° ^{1}H excitation pulse lengths. ^{13}C spectra were acquired by using a recycle delay of 5.0 s with 30° ^{13}C excitation pulse lengths using a carrier frequency offset of 100 ppm.

2.5.2. Solid State Magic Angle Spinning (MAS) NMR

Solid samples were packed into 4-mm silicon nitride rotors and sealed with glass-filled Torlon caps. Each rotor was spun at 4 kHz in a Doty Scientific magic-angle spinning (MAS) probe for ^{13}C cross-polarization (CP)/MAS NMR at 125.76 MHz in a Bruker Avance 500 MHz spectrometer. The proton radiofrequency strength was $\gamma B_H/(2\pi) = 40$ kHz during cross-polarization

and >50 kHz during acquisition, unless noted. Cross-polarization contact time was set at 20 μs. ^{13}C spectra were referenced to the upfield adamantine peak at 28.7 ppm. Spectra were collected by averaging at total 2056 scans with a 10 second recycle delay between scans.

2.5.3. Proton Spin-Relaxation Edited (PSRE)/MAS NMR

Crystalline or semi-ordered materials, such as cellulose and certain elements of the lignin-hemicellulose matrix, produce relatively sharp, well-defined ^{13}C CP/MAS-NMR spectral features. However, native biomass also contains a preponderance of amorphous, ill-defined materials mainly corresponding to the lignin and hemicellulose fractions, and these materials produce broad, asymmetric features that distort the baseline. In order to facilitate quantitative analysis of the cellulose fraction by NMR, is necessary to remove the overlapping signals produced by these amorphous materials. To this end, we followed the method of proton spin-relaxation editing (PSRE) developed by Newman *et al*.(*4*) In this method, differences are exploited in the rotating-frame proton relaxation behavior between cellulose and the non-crystalline lignin and hemicellulose matrices during cross-polarization to separate the NMR signals corresponding to these two fractions. Briefly, MAS-NMR spectra are collected at different spin-lock mixing times (0.03, 6.0 and 10.0 ms were used here), and linear combinations of these spectra are taken, with weighted factors informed by quantitative knowledge of the transverse spin-lattice relaxation times ($T_{1,\rho}$*'s*) corresponding to the crystalline and amorphous material in the sample. The result is two PSRE sub spectra: one corresponding to crystalline cellulose, and the other to the non-order material in the cell wall.

2.5.4. High Resolution (HR)/MAS NMR

^{1}H HR/MAS NMR experiments were performed on a Bruker Avance spectrometer at 500 MHz using a 4 mm Doty Scientific MAS NMR probe. Solvent-pretreated biomass samples of about 70 wt% moisture content were sealed into 30 μL Kel-F HR/MAS rotor inserts (Bruker Biospin Inc.), with the MAS experiments performed at 4 kHz with a ±0.1 K temperature regulation. Dried, solvent-pretreated biomass samples were re-wetted by soaking in an amount of water corresponding to the moisture content of the never-dried samples (typically ~70 wt% water to 30 wt% solids) for 72 hours, and then loaded into 30 μL Kel-F HR/MAS rotor inserts. The sample temperatures under HR/MAS conditions (T_{MAS}) were calibrated using a neat ethylene glycol thermometer. The one-dimensional ^{1}H NMR spectra were obtained using a single pulse Bloch decay sequence with 16 scan averages and 10 seconds recycle delays. Transverse spin–spin relaxation times (T_2) were found using a rotor-synchronized CPMG (Carr–Purcell–Meiboom–Gill) pulse sequence with 16 scan averages, and inter-pulse delays varied between 4 and 20000 rotor cycles (corresponding to 0.002 to 10 seconds transverse spin-spin relaxation times). The ^{1}H NMR chemical shifts were referenced to the internal standard of neat water, δ = +4.7 ppm at 298 K with respect to the chemical shift of TMS, δ = 0 ppm.

2.6. Simulation Details

All classical MD simulations were performed using the Gromacs version 2016 software.(*5*) Reactant, products, and cosolvents were parameterized using the CGenFF/CHARMM36 force fields.(*6-8*) Water was modeled using the Single Point Charged/Extended (SPC/E) model.(*9*) A (6 nm)^3 simulation box was initialized with water and cosolvent molecules at the desired composition. The system was equilibrated in a *NPT* simulation for 5 ns at T = 300 K and P = 1

bar with a velocity-rescale thermostat and Berendsen barostat. Then, a single reactant molecule was added to the system and equilibrated with the same barostat and thermostat for 500 ps. 4 ns *NPT* productions were performed at the reaction temperature and P = 1 bar using a Nose-Hoover thermostat and Parrinello-Rahman barostat. The 4 ns trajectory was partitioned into two 2 ns trajectories. Each partition was used to generate voxel representations. For each simulation snapshot, a 3D histogram was generated with the system centered on the center-of-mass of the reactant. The histogram covered a $(4\ \text{nm})^3$ volume that was divided into a 20×20×20 grid of bins corresponding to $(0.2\ \text{nm})^3$ volume elements. For each bin, normalized occurances of reactant, cosolvent, and water were stored in three separate channels to obtain a 20×20×20×3 grid of voxels for a single MD configuration. These grid values were averaged using 2 ns of simulation data, equivalent to 200 MD configurations, to generate a single voxel representations that capture the spatial distribution of atoms within the system. Voxel representations were then inputted into a pre-trained 3D convolutional neural network called SolventNet, which we have developed in-house, which outputs the predicted kinetic solvent parameters. Each 2 ns partition was treated as two independent trials, where the average predicted kinetic solvent parameter is reported and the standard deviation of the predictions is the error.

Solvation free energies were computed from a series of stochastic dynamics simulations, described in Ref. (*10*). Starting from an equilibrated solvent system, a reactant or product is added to the system. The total potential of the system was defined as a function of Lennard-Jones (λ_{LJ}) and electrostatic (λ_{elec}) potential parameters, shown in Equation 7.6:

$$U(\lambda_{LJ}, \lambda_{elec}) = U_{M,solv}^{LJ}(\lambda_{LJ}) + U_{M,solv}^{elec}(\lambda_{elec}) + U_{M}^{bonded} + U_{M}^{nonbonded} + U_{solv}^{bonded} + U_{solv}^{nonbonded} \quad (1)$$

where $U_{M,solv}^{LJ}$ and $U_{M,solv}^{elec}$ are the LJ and electrostatic potentials between solute and solvent, U_{M}^{bonded} and $U_{M}^{nonbonded}$ are intramolecular bonded and non-bonded potentials of the solute, and U_{solv}^{bonded} and $U_{solv}^{nonbonded}$ are the bonded and non-bonded potentials between all solvent molecules [30]. λ_{LJ} and λ_{elec} vary between 0 and 1 and modify the magnitude of the associated interactions; values of 0 correspond to a species in the vapor phase with no interactions with the solvent environment, while values of 1 correspond to a species in the liquid phase with complete interactions with the solvent environment. Seventeen independent simulations were performed for each free energy: fourteen in which electrostatic potential parameters were turned off ($\lambda_{elec} = 0$) and the VDW potential parameters were varied ($\lambda_{LJ} = 0.00$, 0.00922, 0.04794, 0.11505, 0.260634, 0.31608, 0.43738, 0.56262, 0.68392, 0.79366, 0.88495, 0.95206, 0.99078, or 1.00), three in which VDW potential parameters were on ($\lambda_{LJ} = 1.00$) and electrostatic potential parameters were varied $\lambda_{elec} = 0.25$, 0.75, or 1.00. For each simulation, the system was energy minimized with the steepest descent algorithm and equilibrated for 100 ps at constant temperature followed by 2 ns at constant temperature and constant pressure using the Berendsen barostat. A 5 ns production simulation at constant reaction temperature and pressure (P = 1 bar) was then performed with the Parrinello-Rahman barostat. Energy differences computed between all pairs of windows were collected every 0.2 ps and solvation free energies were computed with the Multistate Bennett Acceptance Ratio (*11*) method using the python alchemical-analysis tool.(*12*)

All simulations were performed using a leapfrog integrator with a 2-fs timestep. Verlet lists were generated using a 1.2 nm neighbor list cutoff. Van der Waals interactions were modeled with a Lennard-Jones potential that was smoothly shifted to zero between 1.0 nm and 1.2 nm. Electrostatic interactions were calculated using the smooth Particle Mesh Ewald method with a short-range cutoff of 1.2 nm, grid spacing of 0.12 nm, and 4th order interpolation. Bonds were

constrained using the LINCS algorithm. All thermostats used a 1.0 ps time constant and all barostats used a 5.0 ps time constant with an isothermal compressibility of 5.0×10^{-5} bar^{-1}.

2.7. References

1. A. Trummal, L. Lipping, I. Kaljurand, I. A. Koppel, I. Leito, Acidity of Strong Acids in Water and Dimethyl Sulfoxide. *The Journal of Physical Chemistry A* **120**, 3663-3669 (2016).
2. E. Raamat *et al.*, Acidities of strong neutral Brønsted acids in different media. *Journal of Physical Organic Chemistry* **26**, 162-170 (2013).
3. J. Sugiyama, R. Vuong, H. Chanzy, Electron diffraction study on the two crystalline phases occurring in native cellulose from an algal cell wall. *Macromolecules* **24**, 4168-4175 (1991).
4. R. H. Newman, Estimation of the lateral dimensions of cellulose crystallites using 13C NMR signal strengths. *Solid State Nuclear Magnetic Resonance* **15**, 21-29 (1999).
5. S. Páll, M. J. Abraham, C. Kutzner, B. Hess, E. Lindahl. (Springer International Publishing, Cham, 2015), pp. 3-27.
6. R. B. Best *et al.*, Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain $\chi1$ and $\chi2$ Dihedral Angles. *Journal of Chemical Theory and Computation* **8**, 3257-3273 (2012).
7. K. Vanommeslaeghe *et al.*, CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* **31**, 671-690 (2010).
8. W. Yu, X. He, K. Vanommeslaeghe, A. D. MacKerell Jr., Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *Journal of Computational Chemistry* **33**, 2451-2468 (2012).
9. H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma, The missing term in effective pair potentials. *The Journal of Physical Chemistry* **91**, 6269-6271 (1987).
10. A. K. Chew, R. C. Van Lehn, Quantifying the Stability of the Hydronium Ion in Organic Solvents With Molecular Dynamics Simulations. *Frontiers in Chemistry* **7**, (2019).
11. M. R. Shirts, J. D. Chodera, Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics* **129**, 124105 (2008).
12. P. V. Klimovich, M. R. Shirts, D. L. Mobley, Guidelines for the analysis of free energy calculations. *Journal of Computer-Aided Molecular Design* **29**, 397-411 (2015).

Chapter 3. Kinetics of Levoglucosenone Isomerization[1]

3.1. Introduction

A key aspect of the utilization of renewable sources of carbon involves the production of platform molecules from lignocellulosic biomass and the upgrading of these molecules to valuable products.(*1*) Recent reports have shown that levoglucosenone (LGO) is a promising intermediate for the production of sustainable chemicals from biomass. For example, LGO can be hydrogenated to Cyrene (dihydrolevoglucosenone), a non-toxic solvent which has similar properties to more environmentally harmful solvents such as N-methylpyrrolidone.(*2*) DuPont has shown that further hydrogenation of LGO produces 1,6-hexanediol, a commodity chemical used in applications such as polyurethanes and coatings, in up to 70% yield.(*3*) LGO can also be converted to pharmaceutical compounds such as the antibiotic (+)-chloriolide.(*4*) While LGO production has previously been reported in low to moderate yields from the pyrolysis of cellulose,(*5, 6*) our group has shown that LGO can be produced from cellulose in 50% yield using polar aprotic solvents at mild conditions (210^{o}C, 20 mM H_2SO_4).(*7*) Cellulose is first depolymerized to form levoglucosan (LGA), which is then dehydrated to form LGO. Small quantities of water (e.g., 3 wt%) shift the selectivity to 5-hydroxymethylfurfural (HMF) rather than LGO.

The aforementioned shift in selectivity toward HMF could be due to the rehydration of LGA to glucose followed by glucose conversion to HMF. Another explanation for this selectivity

[1] This Chapter was adopted from: S. H. Krishna, T. W. Walker, J. A. Dumesic, G. W. Huber, Kinetics of levoglucosenone isomerization. *ChemSusChem* **10**, 129-138 (2017). T.W. Walker, S. Krishna and G. W. Huber conceived the work. S. Krishna designed the experiments and produced the data presented in Figure 3.3.11. TW Walker carried out all other experiments and developed the reaction kinetics model. T.W. Walker and S. Krishna wrote the manuscript. All authors edited manuscript.

shift is that LGO can be isomerized to HMF in the presence of water and an acid catalyst.(*5, 8*) However, the kinetics of LGO isomerization (Figure 3.1.1) remain unknown. An understanding of the factors which promote LGO isomerization could aid in developing strategies to minimize LGO degradation, thereby improving the yield of LGO from cellulose. Increasing the yield of LGO, in turn, would improve the economic viability of producing of high-value chemicals from lignocellulosic biomass. Furthermore, optimizing the isomerization of LGO could allow for the production of HMF or levulinic acid (LA) in high yield, which can in turn be converted to fuels(*9*) and chemicals.(*10*)

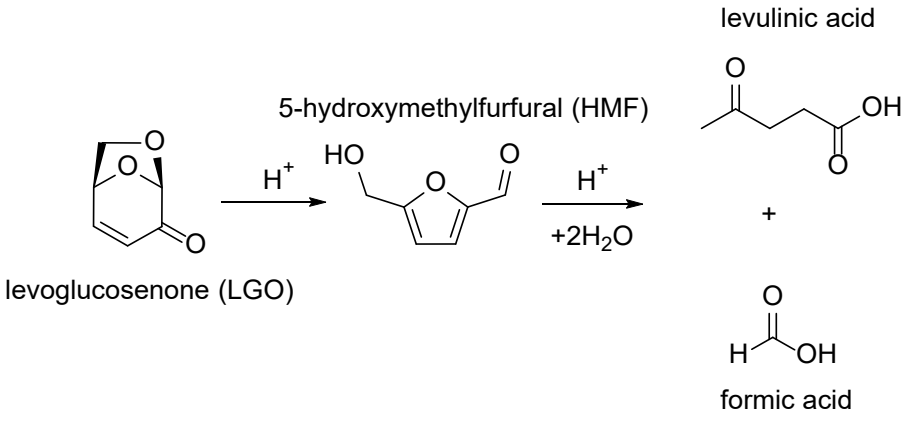


Figure 3.1.1. LGO isomerization to HMF, and subsequent hydration of HMF to form levulinic and formic acids.

The chemistry of LGO isomerization is also relevant to the ring rearrangement of sugars in general. LGO isomerization to HMF involves a transformation from a 6-membered ring to a 5-membered ring, analogous to the isomerization of glucose to fructose. Interestingly, LGO isomerization has been shown to proceed over a Brønsted acid catalyst,(*5, 8*) while glucose isomerization to fructose requires a Lewis acid catalyst.(*11-13*) An investigation of LGO isomerization would provide a basis for comparison to the glucose isomerization mechanism.

Herein, we investigate the isomerization of LGO to HMF in an aqueous solvent system with a sulfuric acid catalyst. We elucidate the reaction network and develop a reaction kinetics model to describe this system across a range of experimental conditions. We propose a mechanism for LGO isomerization analogous to glucose isomerization which is consistent with the experimental results. Our findings provide directions for designing catalytic systems to control the LGO isomerization reaction depending on which products are desired.

3.2. Experimental

Levoglucosenone (98%, Apollo Scientific), 5-hydroxymethylfurfural (99%, Sigma Aldrich), H_2SO_4 (96%, Sigma Aldrich), levulinic acid (99%, Sigma Aldrich), formic acid (98%, Sigma Aldrich), tetrahydrofuran (99.9%, Acros Organics), and HPLC grade water (Fisher Scientific) were used as received.

3.2.1 Reaction kinetics studies.

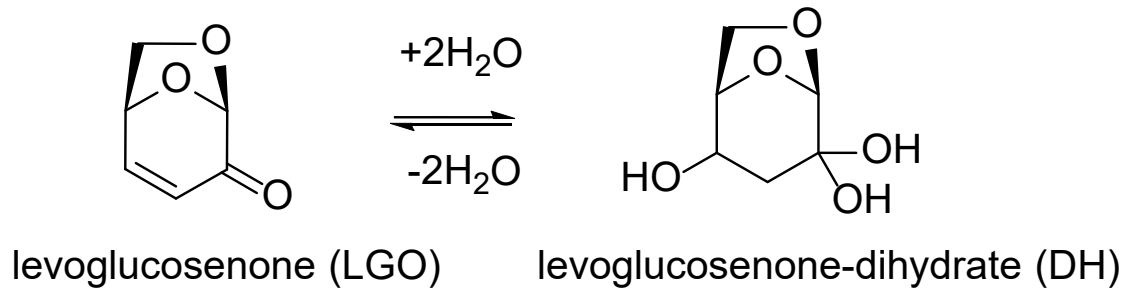
2 mL of solution containing the appropriate solute and sulfuric acid concentrations was added to 10 mL thick-walled glass reactors. The reactors were placed in an oil bath at the desired temperature and stirred at 500 rpm. Reactors were removed from the oil bath and quenched in an ice water bath at the appropriate time. Reaction products were filtered using a 0.2 μm polyethersulfone (PES) syringe filter prior to analysis.

3.2.2. Product analysis.

Products were quantified using a Shimadzu high performance liquid chromatograph (HPLC) with a BioRad Aminex 87H column with 5 mM H_2SO_4 mobile phase. The column was

operated at a flow rate of 0.6 mL/min and a temperature of 30ºC. A refractive index (RI) detector was used to quantify levulinic acid, formic acid, and the hydrated form of levoglucosenone. A photodiode array (UV) detector was used to quantify LGO and HMF. Because LGO and HMF elute at nearly the same retention time, LGO and HMF were quantified using two different UV wavelengths. HMF was quantified at 290 nm while LGO was quantified at 370 nm. The contribution of HMF to the LGO peak at 370 nm, and the contribution of LGO to the HMF peak at 290 nm, were taken into account by using the sensitivity of HMF at 370 nm and the sensitivity of LGO at 290 nm, respectively. This quantification method was validated using mixtures of LGO and HMF at known concentrations. Concentrations of all products were calculated based on calibration curves using standards.

When LGO was diluted in water, a single major product of hydration was observed by HPLC. ^{1}H and ^{13}C NMR were used to identify this compound as the doubly hydrated form of LGO (Figure 3.2.1). The LGO-dihydrate (DH) was prepared in 90% yield by diluting LGO in water and allowing the hydration to occur at room temperature for >100 h. NMR spectra were collected using a Bruker Avance 500 MHz spectrometer at room temperature. 10% D_2O was added to the solution to facilitate locking to the deuterium resonance. The ^{1}H spectrum of the doubly hydrated form of LGO is in agreement with ^{1}H chemical shift values for this compound in the literature.(*5*) The existence of a single identifiable hydration product allowed for an HPLC-RI calibration curve to be developed by tracking the disappearance of LGO as the hydration progressed at room temperature.



levoglucosenone (LGO) levoglucosenone-dihydrate (DH)

Figure 3.2.1. Formation of the LGO-dihydrate (DH) in water via hydration of the C=C double bond to form a hydroxyl group, and hydration of the ketone functionality to form a geminal diol.

3.2.3. Reaction kinetics modeling.

Reaction kinetics modeling was carried out using MATLAB (Version 2015b). The set of coupled ODEs describing the concentrations of each species over time were solved using the "ode23t" differential equation solver. The error between the experimental data and the model was minimized using the non-linear optimization function "nlinfit".

3.3. Results and Discussion

Figure 3.3.1 shows the results of LGO isomerization in water over a sulfuric acid catalyst. DH and HMF are the primary products of LGO conversion. The ratio of the concentrations of DH to LGO approaches a constant value of ~1.2, suggesting that these species may be in equilibrium. The concentrations of both LGO and DH continue to decrease over time, suggesting that either species can be converted to HMF, which is then hydrated to form LA and formic acid (FA), as reported in the literature.(*14-16*) The carbon balance decreases over the course of the reaction, indicating the formation of unaccounted degradation products (humins).

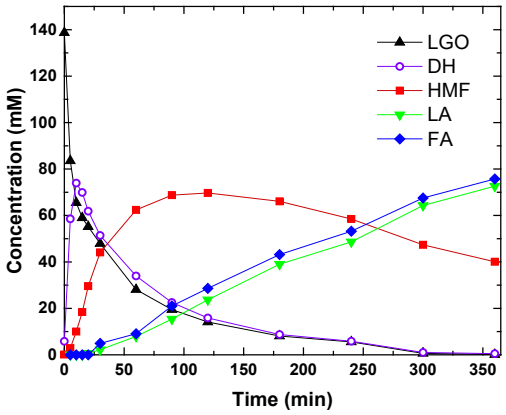


Figure 3.3.1. Acid-catalyzed transformation of LGO to HMF in water, and hydration of HMF to form LA and FA. *Reaction conditions: 150 mM LGO in 50 mM H_2SO_4 (aq); 125°C.* Solid lines are visual aids.

3.3.1. Equilibrium between LGO and DH.

The thermal hydration of LGO to DH, in the absence of an acid catalyst, was studied in water at 125°C (Figure 3.3.2). At these conditions, the concentrations of the two species remain at a constant ratio of ~1.2, similar to the ratio of DH to LGO during the acid-catalyzed reaction as shown in Figure 1. The >95% molar carbon balance maintained over the course of 4 hours indicates that LGO and DH exist in equilibrium at the reaction conditions of interest.

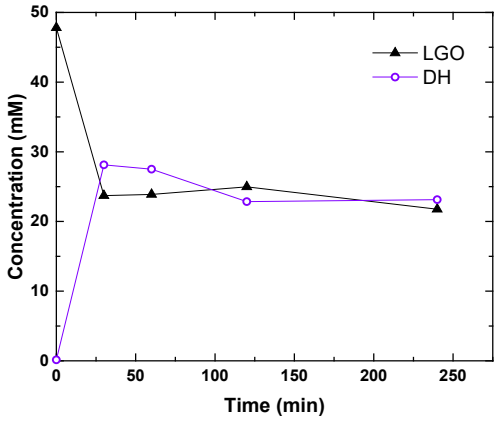


Figure 3.3.2. Thermal hydration of LGO in pure water. *Reaction conditions: 50 mM LGO, 125°C.* Solid lines are visual aids.

At room temperature, LGO is converted to DH in approximately 90% yield by diluting LGO in water to 75 mM and waiting for 100 h. Using this approach, a sample of DH in water was prepared and used as a feedstock for the acid-catalyzed reaction. As shown in Figure 3.3.3, DH is converted to LGO rapidly during the reaction. The trends in the concentrations of HMF and LA are similar to those observed in Figure 3.3.1. Finally, Figure 3.3.4 shows the ratio of the concentrations of DH to LGO versus time for each of the aforementioned experiments. In all three experiments, the ratio of the two species is seen to approach a near-constant value of ~1.2, which is maintained over the course of the reaction. These results provide further evidence that LGO and DH are in equilibrium at the reaction conditions.

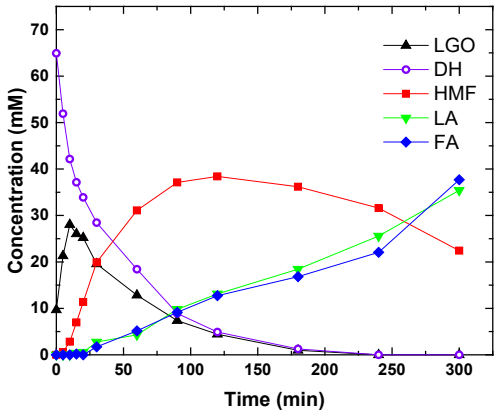


Figure 3.3.3. LGO isomerization in water using an LGO-dihydrate rich feed. *Reaction conditions: 65 mM DH plus 10 mM LGO; 50 mM H_2SO_4 (aq); 125°C.* Solid lines are visual aids.

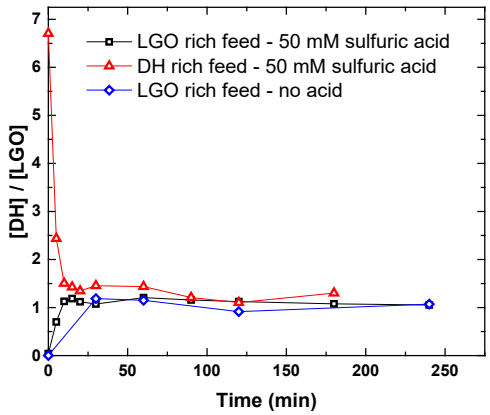
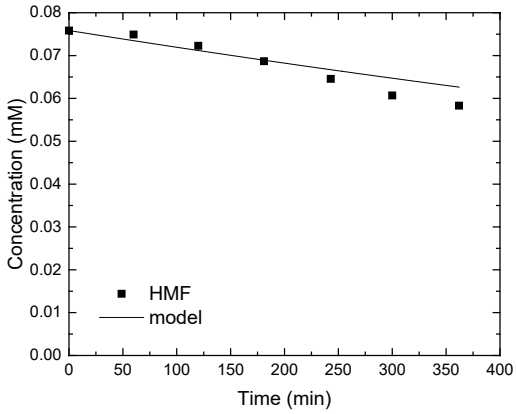


Figure 3.3.4. The ratio of DH to LGO vs time for various reaction conditions at 125°C, both with and without acid. Solid lines are visual aids.

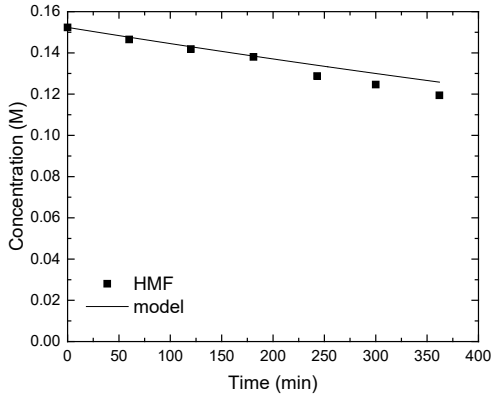
Because LGO and DH are in equilibrium at our experimental conditions, the rate of HMF formation can be modeled as being first order in the combined concentration of LGO and DH.

3.3.2. Thermal degradation reactions.

As shown in Figure 3.3.2, the pool of LGO and DH is thermally stable at the reaction conditions of interest. The thermal stabilities of HMF, LA, and FA were also studied by heating these species in water. At 125ºC, LA and FA did not degrade over the course of 4 hours. Upon heating in water, HMF degraded at a rate that is described by a first-order model, as can be seen in Figure 3.3.5. Girisuta, et al. studied the thermal stability of HMF in water.(*16*) In that study, HMF was heated at 98ºC for 2 hours in 1 molar solutions of FA or LA. Negligible HMF degradation was observed under those conditions, in agreement with our findings (Figure 3.3.5). At the higher temperatures and longer reaction times accessed in our experiments, however, thermal degradation of HMF occurred. Therefore, thermal HMF degradation was accounted for in our kinetic model by calculating the pre-exponential factor and activation energy for this reaction from the data in Figure 3.3.5.



a) 75 mM HMF; 125 °C



b) 150 mM HMF; 125 °C

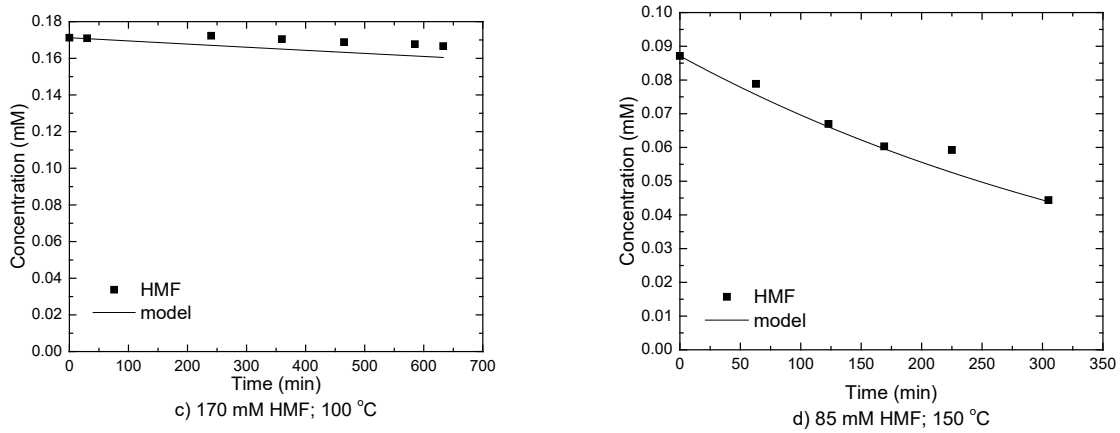


Figure 3.3.5. Thermal degradation of HMF in water at various initial concentrations and temperatures. Solid lines are model-predicted values.

3.3.3. Development of a reaction kinetics model.

Experiments in 100 mM H_2SO_4 (aq) revealed that LA and FA do not degrade appreciably at 125ºC. However, LGO and HMF undergo acid-catalyzed transformations at the same conditions (Scheme 1); therefore, it was not possible to independently measure the rates at which these species are converted. Quantification of these degradation routes was achieved by conducting a series of experiments at a temperature of 125ºC, wherein the initial concentrations of HMF, LGO and H_2SO_4 were independently varied. The results of these experiments are summarized in Figure 3.3.7, which shows the experimental data along with the model-predicted values. In all six experiments, the rate of carbon loss from identifiable products was seen to vary as a function of HMF concentration only, which precludes the possibility that degradation reactions involving LGO are occurring under these conditions. Furthermore, the rate of thermal HMF degradation alone was not sufficient to explain the loss of carbon at the reaction conditions, indicating that a

separate, acid-catalyzed degradation reaction of HMF was occurring at these conditions.

Based on the above findings, we propose a reaction network for the acid-catalyzed isomerization of LGO to HMF, and the subsequent reactions of HMF (Figure 3.3.6). LGO was modeled as existing in equilibrium with its dihydrate (DH). The rate of HMF production (r_1) was modeled as being proportional to the combined concentrations of LGO and DH, as discussed previously. HMF undergoes hydration to LA and FA (r_2), as well as thermal (r_4) and acid-catalyzed (r_3) reactions to unaccountable degradation products. The sulfuric acid catalyst was modeled as having a single fully dissociated proton.(*17*)

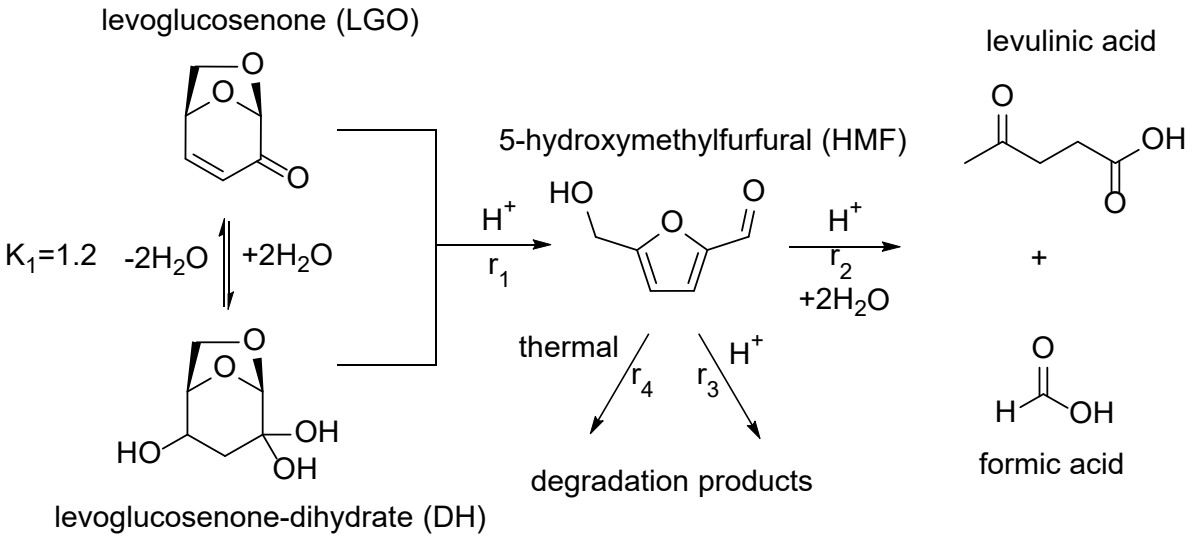


Figure 3.3.6. Proposed reaction network for the acid-catalyzed isomerization of LGO to HMF, and subsequent reactions of HMF in water.

Assuming that first-order kinetics are valid for all reactions in the network, the rates for each step in Scheme 2 are described by Equations 3.1-3.4:

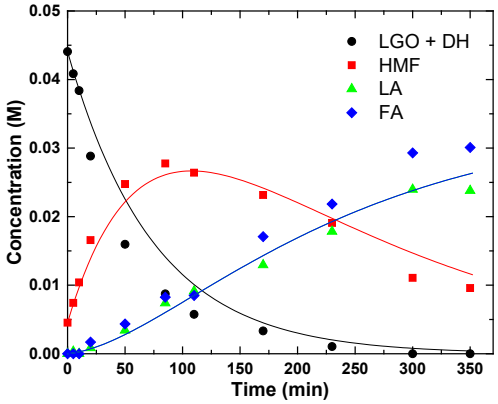
$$r_1 = k_1[\mathrm{LGO} + \mathrm{DH}][H^+] \qquad (3.1)$$

$$r_2 = k_2[\mathrm{HMF}][H^+] \qquad (3.2)$$

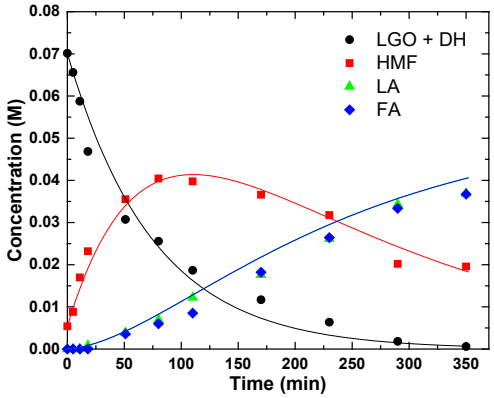
$$r_3 = k_3[\mathrm{HMF}][H^+] \qquad (3.3)$$

$$r_4 = k_4[\mathrm{HMF}] \qquad (3.4)$$

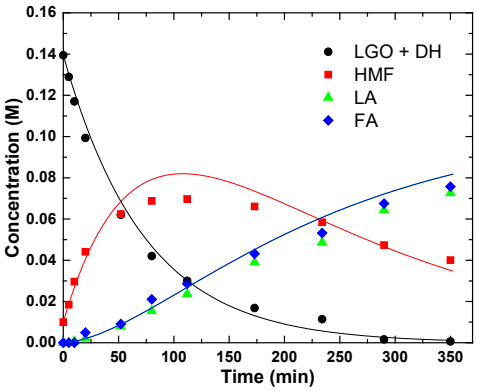
As shown in Figure 6, the reaction kinetics model describes the data across different reactant concentrations without the need to invoke fractional or higher order rate expressions. Table 3.3.1 shows the rate constants extracted by the model along with their associated 95% confidence intervals.



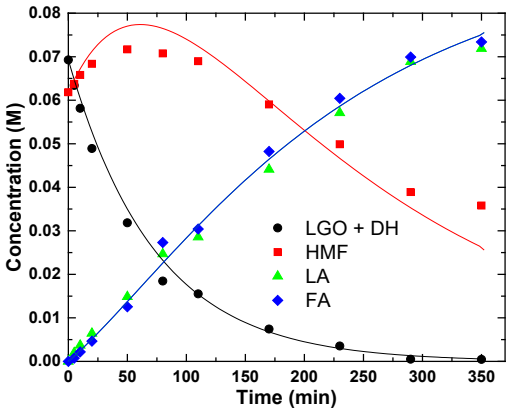
a) 50 mM LGO; 50 mM H_2SO_4; 125 °C



b) 75 mM LGO; 50 mM H_2SO_4; 125 °C



c) 150 mM LGO; 50 mM H_2SO_4; 125 °C



d) 75 mM LGO + 60 mM HMF; 50 mM H_2SO_4; 125 °C

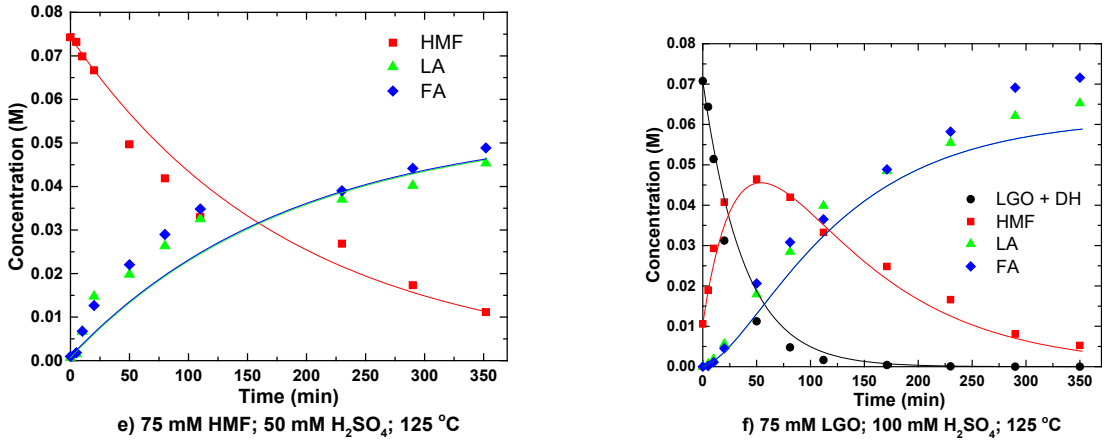


Fig. 3.3.7. Experimental and model-predicted concentration profiles for the transformation of (LGO + DH) to HMF, and subsequent reactions of HMF in water. Solid lines are model-predicted values.

Comparisons of the model-predicted versus experimental carbon balance for reactions of HMF and LGO in water are shown in Figures 3.3.8a and 3.3.8b, respectively. As seen in Figure 7, when LGO is used as a feed, the model under-predicts the carbon loss at early reaction times and accurately predicts the carbon loss at later reaction times. In the case where HMF is used as a feed, however, the model achieves good agreement with experiment across the entire span of the reaction. This behavior indicates that a small amount of LGO is initially converted to unidentified species which are ultimately converted to products. However, as seen in Figure 3.3.8b, the error introduced by this discrepancy is at most about 15%.

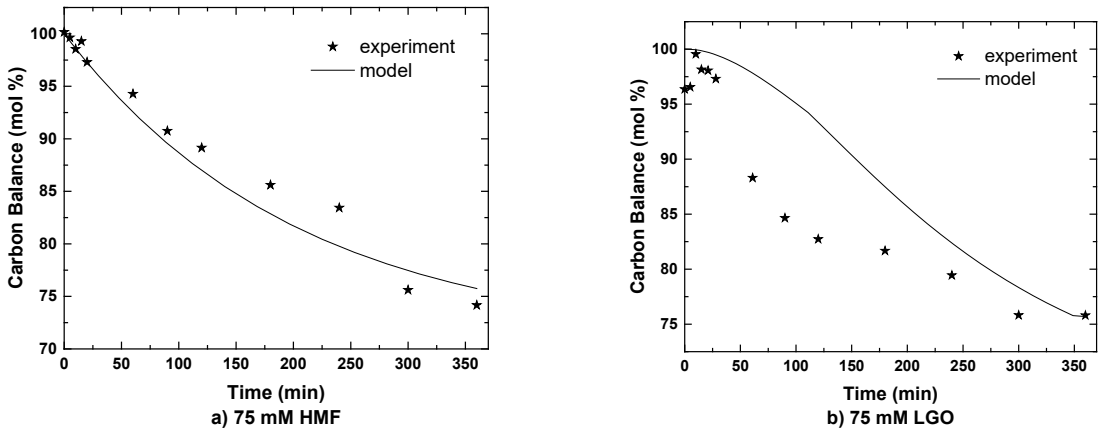


Fig. 3.3.8. Model-predicted and experimentally determined carbon balance versus time for reaction with (a) HMF and (b) LGO feedstocks. *Reaction conditions: 50 mM H_2SO_4 (aq); 125°C.* Solid lines are model-predicted values.

The temperature dependence of the reaction rates was measured by recasting the rate constants in Equations 3.1-3.4 in terms of the Arrhenius relationships below:

$$k_1 = k_1^{ref} \exp\left\{-\frac{E_{a1}}{R}\left(\frac{1}{T}-\frac{1}{T_R}\right)\right\} \qquad (3.5)$$

$$k_2 = k_2^{ref} \exp\left\{-\frac{E_{a2}}{R}\left(\frac{1}{T}-\frac{1}{T_R}\right)\right\} \qquad (3.6)$$

$$k_3 = k_3^{ref} \exp\left\{-\frac{E_{a3}}{R}\left(\frac{1}{T}-\frac{1}{T_R}\right)\right\} \qquad (3.7)$$

$$k_4 = k_4^{ref} \exp\left\{-\frac{E_{a4}}{R}\left(\frac{1}{T}-\frac{1}{T_R}\right)\right\} \qquad (3.8)$$

where k_i^{ref} is the rate constant for the i^{th} reaction measured at the reference temperature (T_R) of 398 K (125°C), E_{ai} is the activation energy for the i^{th} reaction, T is the reaction temperature in Kelvin, and R is the gas constant. Equations 3.5-3.8 were then defined in a separate model, which measured

the activation energies by optimizing their values to fit experimental data across four variable temperature experiments between 100-150ºC. The results of this process are shown in Figure 8, which shows agreement between model and experiment. Table 3.3.1 summarizes all of the reaction kinetics parameters measured in this study, along with their associated uncertainties.

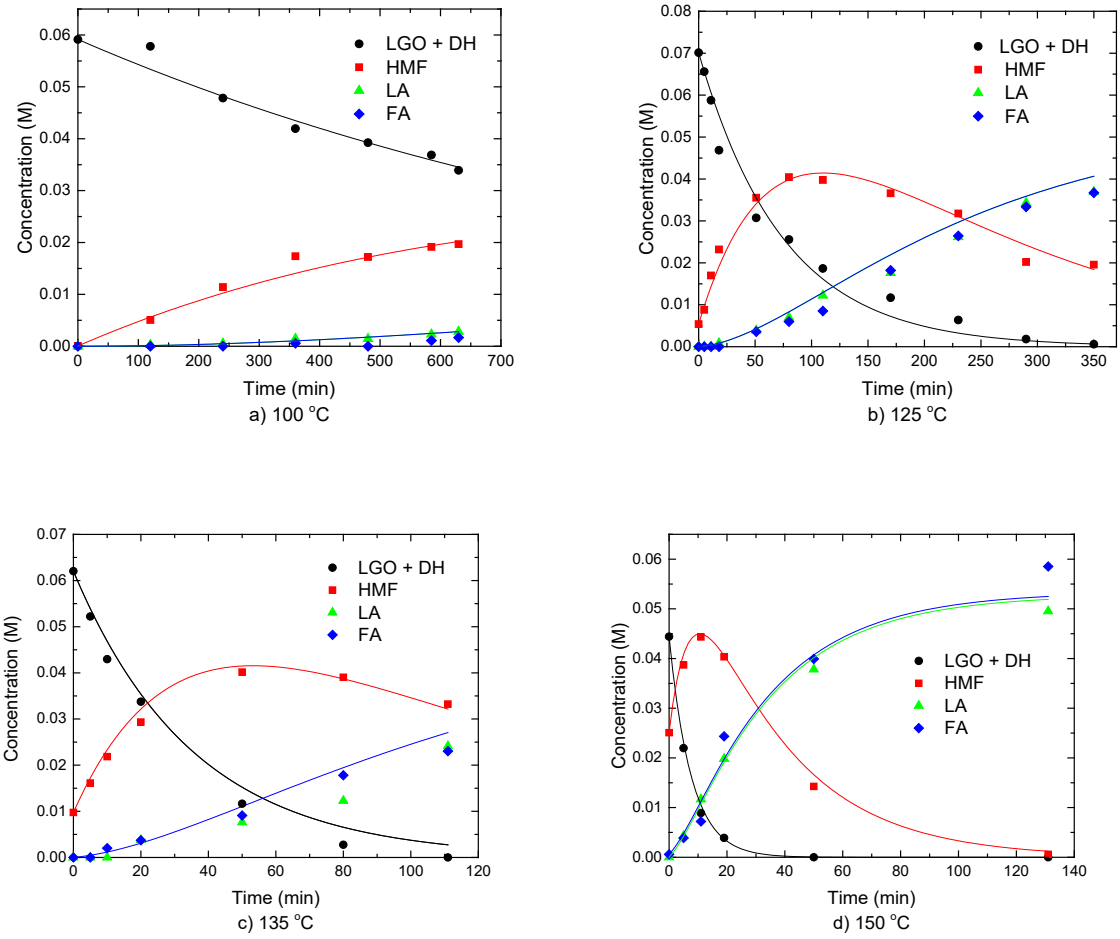


Fig. 3.3.9. Experimental and model-predicted concentration profiles for variable-temperature experiments. Solid lines are model-predicted values. *Reaction conditions: 75 mM LGO, 50 mM H_2SO_4 (aq).*

Table 3.3.1. Model-predicted reaction kinetics parameters (with associated 95% confidence intervals) for the acid-catalyzed isomerization of LGO to HMF, and subsequent reactions in water. Model obtained at the following range of conditions: 100-150^{o}C, 50-100 mM H_2SO_4, 50-150 mM LGO.

Rate Constants at Reference Temperature,

T_R = 125 ^{o}C

k_1^{ref}	0.274	+/-	0.012	L mol^{-1} min^{-1}
k_2^{ref}	0.077	+/-	0.002	L mol^{-1} min^{-1}
k_3^{ref}	0.019	+/-	0.003	L mol^{-1} min^{-1}
k_4^{ref}	0.00053	+/-	0.00007	min^{-1}

Activation Energies

E_{a1}	130.1	+/-	4.0	kJ mol^{-1}
E_{a2}	107.7	+/-	4.8	kJ mol^{-1}
E_{a3}	102.2	+/-	18.0	kJ mol^{-1}
E_{a4}	81.0	+/-	10.0	kJ mol^{-1}

As seen in Table 3.3.1, the activation energy of LGO isomerization (E_{a1}) to HMF is the highest value of all reactions, indicating that the HMF selectivity is maximized at high

temperatures and short reaction times. Using the model-predicted reaction kinetics parameters, it is possible to simulate the effects of temperature and reaction time on the maximum obtainable yield of HMF. The results of this analysis (Figure 3.3.10) indicate that a maximum HMF yield of ~76% could be achieved from LGO at a temperature of 220ºC in water. However, achieving this yield would require a reaction time of approximately 6 seconds. At a reaction time of 5 minutes, the yield of HMF at 165ºC and 50 mM H_2SO_4 is approximately 66%.

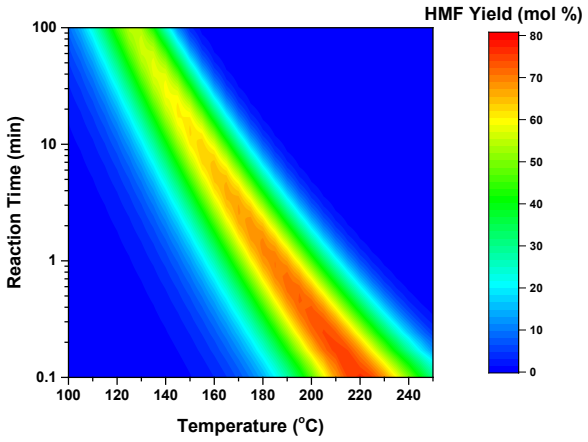


Fig. 3.3.10. Model-predicted yields of HMF from LGO in water versus temperature and reaction time. Acid strength is held fixed at 50 mM H_2SO_4.

While the kinetics of LGO isomerization to HMF have not been previously studied in the literature, our results for HMF conversion to LA and FA can be compared to literature reports summarized by Weingarten, et al.(*14*) The activation energy for HMF hydration to LA and FA is 95-110 kJ/mol based on the majority of literature reports,(*14, 16, 18*) in agreement with our measurement of 107.7 +/- 4.8 kJ/mol. Weingarten, et al. and Shen, et al. found that the activation

energy for HMF degradation to humins over an acid catalyst is approximately 145 kJ/mol,(*14, 19*) while Girisuta, et al. reported a value of 111 kJ/mol.(*18*) Jing, et al. reported an activation energy of 109 kJ/mol for the thermal degradation of HMF in water.(*20*) In this study, we measured an activation energy of 102.2 +/- 18.0 kJ/mol for catalytic degradation of HMF and an activation energy of 81.0 +/- 10.0 kJ/mol for thermal degradation of HMF. Our model-predicted activation energy for HMF degradation is on the lower end of the range compared to the literature, most likely because two routes of HMF degradation are incorporated. Both Weingarten, et al. and Girisuta, et al. found that the reaction orders for HMF conversion to LA and FA, as well as HMF degradation, are near unity,(*14, 16*) in agreement with our findings.

3.3.4. Mechanistic insights.

The role of water in the isomerization of LGO was investigated by studying the reaction in mixed solvent systems of water and tetrahydrofuran (THF). As shown in Figure 3.3.11, the yield of HMF plus LA decreases monotonically as the solvent system is changed from water to THF. In pure THF solvent, almost no HMF or LA is observed, and LGO is converted to unidentified products (most likely humins). These results suggest that water plays an important role in the LGO isomerization mechanism. These findings are consistent with our previous work on cellulose dehydration to LGO, in which we suggested that THF inhibits the conversion of LGO to HMF, and that LGO isomerization to HMF could contribute to the selectivity shift from LGO to HMF when water is added to the THF solvent system.(*7*)

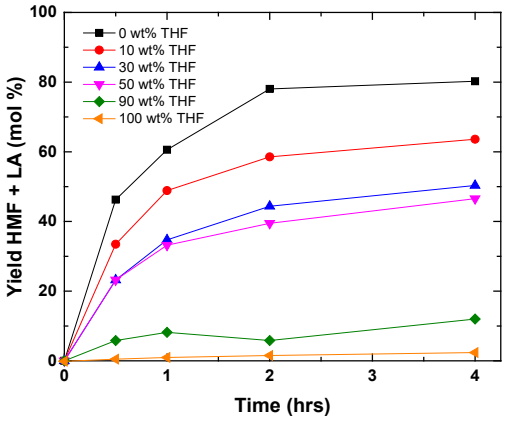


Figure 3.3.11. Yield of HMF plus LA versus time in different THF/water solvent systems. *Reaction conditions: 55 mM LGO; 50 mM H_2SO_4; 125°C.* Solid lines are visual aids.

The observation that LGO isomerization is facilitated by water is consistent with the intermediate proposed by Shafizadeh, et al. involving hydration of the anhydro-bridge prior to ring rearrangement to form HMF.(*5*) In agreement with this hypothesis, Kawamoto, et al. found that LGO conversion to furfural in a sulfolane solvent over a sulfuric acid catalyst was promoted by the presence of water.(*6*) The authors proposed a mechanism passing through the same hydrated intermediate as that posited by Shafizadeh.

Because the resulting hydrated intermediate resembles glucose, we interpret the subsequent ring rearrangement chemistry by analogy to glucose isomerization. Both the isomerization of glucose to fructose and the isomerization of LGO to HMF involve transformations from a 6-membered ring to a 5-membered ring. However, LGO isomerization proceeds over a Brønsted acid catalyst, while glucose isomerization requires a Lewis acid catalyst. In the case of glucose isomerization, the cyclic pyranose form of glucose exists in equilibrium with an acyclic aldehyde form in solution.(*13*) The Lewis acid facilitates a C_1-C_2 hydride shift, converting the aldose into a ketose, which can then ring-close to form the cyclic furanose form of fructose.(*11*) The fact that

LGO isomerization does not require a Lewis acid suggests that no hydride shift is required for this mechanism to proceed.

We have proposed a mechanism for LGO isomerization consistent with these observations, shown in 3.3.12. First, the anhydro-bridge of LGO is broken to form the hydrated intermediate, as proposed by Shafizadeh and Kawamoto. Next, this hydrated intermediate is ring-opened to an acyclic species containing both ketone and aldehyde functionalities. This species can then be ring-closed at the C_2 position to form the 5-membered ring of HMF. Because this species already contains a carbonyl group at the C_2 position, no hydride shift is required and therefore no Lewis acid is required.

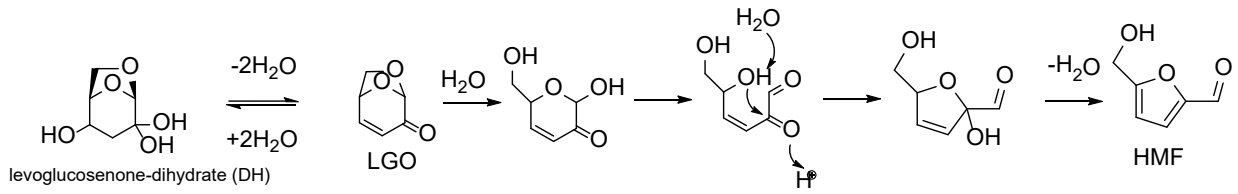


Figure 3.3.12. Proposed mechanism for LGO isomerization to HMF.

The acyclic intermediate shown in Scheme 4 resembles 3-deoxyglucosone, which is an acyclic species containing both aldehyde and ketone functionalities. Bols and co-workers showed that 3-deoxyglucosone can be converted to HMF in >90% yield over a sulfuric acid catalyst.(*21*) Because the acyclic intermediate in Scheme 4 is identical to 3-deoxyglucosone except for the C=C bond in LGO (which most likely does not participate in the ring rearrangement chemistry), it is plausible that this intermediate would react in a similar fashion as 3-deoxyglucosone.

The mechanism in Figure 3.3.12 is similar to a mechanism for glucose conversion to HMF, with LGO as a side-product, proposed by Bell and co-workers.(*22*) The mechanism by Bell and

co-workers involves glucose conversion to the same acyclic intermediate as the one shown in Figure 3.3.12, followed by parallel pathways for the production of HMF or LGO from this intermediate. This mechanism can be made consistent with Figure 3.3.12 if the reaction from the acyclic intermediate to LGO is reversible, allowing for ring-opening of LGO to form this intermediate followed by ring closure to form HMF. Note that the study by Bell and co-workers used a solid acid catalyst and ionic liquid solvent, while our study was conducted over a sulfuric acid catalyst in an aqueous solvent.

An alternative mechanism for production of HMF involves the intermediate formation of DH, as shown in Figure 3.3.13. In the mechanism passing through DH, the acyclic species would most likely contain a geminal diol functionality at the C_2 position. In the case of glucose isomerization to fructose, following the C_1-C_2 hydride shift, ring closure occurs via nucleophilic attack of the ketone at the C_2 position.(*11*) This step is not possible for the acyclic intermediate in Scheme 5 because no ketone group is present. However, dehydration of one of the alcohol groups of the geminal diol would lead to a carbocation which could facilitate ring closure.

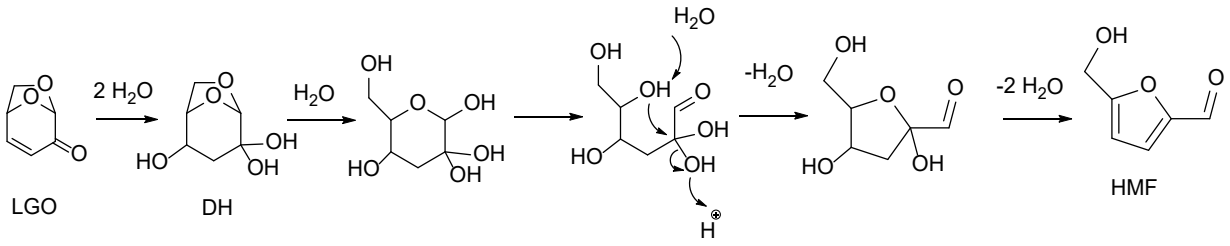
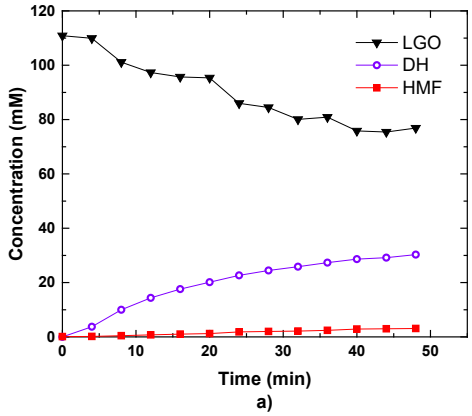


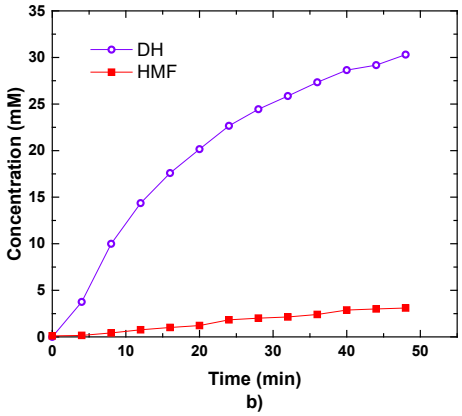
Figure 3.3.13. Alternative mechanism for LGO isomerization to HMF through the intermediate DH.

Due to the short timescales (<10 min) over which the equilibrium between LGO and DH develops in water, it is difficult to conclude experimentally whether DH is a spectator species

(Figure 3.3.12) and/or a key intermediate (Figure 3.3.13) for HMF production. To gain insight into which scheme is more valid, it is necessary to identify reaction conditions at which LGO and DH are far from equilibrium during early reaction times.

As shown in Figures 3.3.14a-c, carrying out the reaction in a 50 wt% mixture of THF and water and lowering the temperature to 100^{o}C shifts the equilibrium between DH and LGO towards LGO, and lengthens the time over which this equilibrium is achieved. The carbon balance is near 100% at all times during this experiment, and LA and FA are not observed. Therefore, HMF conversion can be ignored under these conditions. This system provides an opportunity to compare the initial rate of HMF production with the rates of LGO and DH consumption, when LGO and DH are not near equilibrium.





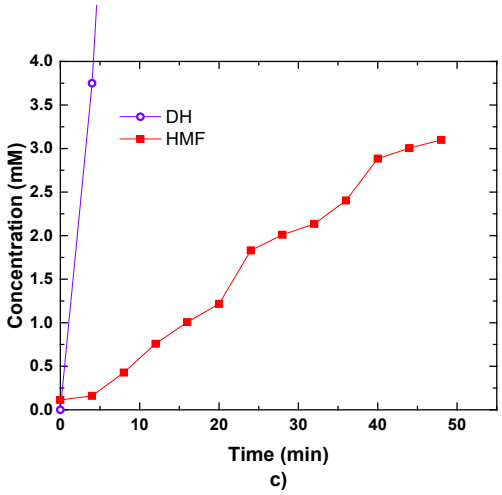


Figure 3.3.14. LGO isomerization to HMF in 50 wt% THF and water. Panels a-c show a single experiment with different scales to aid in visualizing the concentrations of LGO, DH, and HMF over time. *Reaction conditions: 110 mM LGO; 50 mM H_2SO_4; 100°C.* Solid lines are visual aids.

As shown in Figure 3.3.14b, the concentration of DH increases by an order of magnitude over the course of the experiment. If DH were the sole precursor for HMF production, then the rate of HMF production would scale linearly with the DH concentration (assuming that HMF production is first order in DH). However, as seen in Figure 3.3.14c, the rate of HMF production is nearly constant throughout the experiment. Therefore, HMF is not produced solely from DH. In contrast, the concentration of LGO does not change appreciably over the timescale of the experiment (Figure 3.3.14a), which is consistent with the near-constant rate of HMF production under the hypothesis that LGO is the precursor for HMF production. It is important to note, however, that this analysis does not preclude the possibility that HMF is produced from both LGO

and DH. We therefore conclude that LGO is isomerized directly to HMF, and also hydrated to DH, which may or may not be converted to HMF.

3.4. Conclusions

We have elucidated a reaction network for acid-catalyzed LGO isomerization in aqueous solvent systems and developed a reaction kinetics model which describes the experimental data. LGO is isomerized to HMF in the presence of a Brønsted acid catalyst, and HMF is further converted to LA and FA. LGO exists in equilibrium with a dihydrated species at reaction conditions. The main sources of carbon loss are thermal and catalytic degradation of HMF. Within the range of experimental conditions studied, the yield of HMF from LGO can be maximized at higher temperatures and shorter reaction times. The yields of HMF and LA decrease monotonically as the THF/water solvent ratio increases, suggesting that water plays a role in the LGO isomerization reaction. Using a mixed THF/water solvent system and lower temperature to slow the equilibration of LGO and DH, analysis of the initial rate of HMF production showed that HMF is produced from LGO and is not produced solely from DH. These results are consistent with a mechanism for LGO isomerization which proceeds via hydration of the anhydro–bridge, followed by ring rearrangement analogous to the isomerization of glucose to fructose. Our understanding of the factors which affect LGO isomerization to HMF could aid in designing catalytic processes to produce LGO from cellulose more efficiently, improving the economic viability of renewable chemicals production from lignocellulosic biomass.

3.5. Acknowledgements

This material is based upon work supported by the Department of Energy, Office of Energy Efficiency and Renewable Energy (EERE), under Award Number DE-EE0006878. S.H.K. acknowledges that this material is based upon work supported by the National Science Foundation under Grant No. DGE-1256259. We thank Quinn A. Rashke for useful discussions about the reaction mechanism. We thank the UW Department of Chemistry, especially Dr. Heike Hofstetter and Dr. Charles Fry, for assisting us in the use of Bruker Avance 500 MHz NMR Spectrometer. A generous gift from Paul J. Bender enabled this spectrometer to be purchased [2012].

3.6. References

1. J. N. Chheda, G. W. Huber, J. A. Dumesic, Liquid-phase catalytic processing of biomass-derived oxygenated hydrocarbons to fuels and chemicals. *Angewandte Chemie International Edition* **46**, 7164-7183 (2007).
2. J. Sherwood *et al.*, Dihydrolevoglucosenone (Cyrene) as a bio-based alternative for dipolar aprotic solvents. *Chemical Communications* **50**, 9650-9652 (2014).
3. A. M. Allgeier *et al.* (Google Patents, 2015).
4. M. Ostermeier, R. Schobert, Total Synthesis of (+)-Chloriolide. *The Journal of organic chemistry* **79**, 4038-4042 (2014).
5. F. Shafizadeh, R. H. Furneaux, T. T. Stevenson, Some reactions of levoglucosenone. *Carbohydrate Research* **71**, 169-191 (1979).
6. H. Kawamoto, S. Saito, W. Hatanaka, S. Saka, Catalytic pyrolysis of cellulose in sulfolane with some acidic catalysts. *Journal of wood science* **53**, 127-133 (2007).
7. F. Cao *et al.*, Dehydration of cellulose to levoglucosenone using polar aprotic solvents. *Energy & Environmental Science* **8**, 1808-1815 (2015).
8. L. Qi *et al.*, Catalytic conversion of fructose, glucose, and sucrose to 5-(hydroxymethyl) furfural and levulinic and formic acids in γ-valerolactone as a green solvent. *Acs catalysis* **4**, 1470-1477 (2014).
9. D. J. Braden, C. A. Henao, J. Heltzel, C. C. Maravelias, J. A. Dumesic, Production of liquid hydrocarbon fuels by catalytic conversion of biomass-derived levulinic acid. *Green Chemistry* **13**, 1755-1765 (2011).
10. T. Buntara *et al.*, Caprolactam from renewable resources: catalytic conversion of 5-hydroxymethylfurfural into caprolactone. *Angewandte Chemie International Edition* **50**, 7083-7087 (2011).

11. Y. Román-Leshkov, M. Moliner, J. A. Labinger, M. E. Davis, Mechanism of glucose isomerization using a solid Lewis acid catalyst in water. *Angewandte Chemie International Edition* **49**, 8954-8957 (2010).
12. R. Bermejo-Deval *et al.*, Metalloenzyme-like catalyzed isomerizations of sugars by Lewis acid zeolites. *Proceedings of the National Academy of Sciences* **109**, 9727-9732 (2012).
13. I. Delidovich, R. Palkovits, Catalytic Isomerization of Biomass-Derived Aldoses: A Review. *ChemSusChem* **9**, 547-561 (2016).
14. R. Weingarten, J. Cho, R. Xing, W. C. Conner Jr, G. W. Huber, Kinetics and reaction engineering of levulinic acid production from aqueous glucose solutions. *ChemSusChem* **5**, 1280-1290 (2012).
15. V. Choudhary *et al.*, Insights into the interplay of Lewis and Brønsted acid catalysts in glucose and fructose conversion to 5-(hydroxymethyl) furfural and levulinic acid in aqueous media. *Journal of the American Chemical Society* **135**, 3997-4006 (2013).
16. B. Girisuta, L. Janssen, H. Heeres, A kinetic study on the decomposition of 5-hydroxymethylfurfural into levulinic acid. *Green Chemistry* **8**, 701-709 (2006).
17. D. Knopf, B. Luo, U. Krieger, T. Koop, Thermodynamic dissociation constant of the bisulfate ion from Raman and ion interaction modeling studies of aqueous sulfuric acid at low temperatures. *The Journal of Physical Chemistry A* **107**, 4322-4332 (2003).
18. B. Girisuta, L. Janssen, H. Heeres, Kinetic study on the acid-catalyzed hydrolysis of cellulose to levulinic acid. *Industrial & engineering chemistry research* **46**, 1696-1708 (2007).
19. J. Shen, C. E. Wyman, Hydrochloric acid-catalyzed levulinic acid formation from cellulose: data and kinetic model to maximize yields. *AIChE Journal* **58**, 236-246 (2012).
20. J. Qi, L. Xiuyang, Kinetics of non-catalyzed decomposition of glucose in high-temperature liquid water. *Chinese Journal of Chemical Engineering* **16**, 890-894 (2008).
21. H. Jadhav, C. M. Pedersen, T. Sølling, M. Bols, 3-Deoxy-glucosone is an Intermediate in the Formation of Furfurals from D-Glucose. *ChemSusChem* **4**, 1049-1051 (2011).
22. M. Chidambaram, A. T. Bell, A two-step approach for the catalytic conversion of glucose to 2, 5-dimethylfuran in ionic liquids. *Green Chemistry* **12**, 1253-1262 (2010).

Chapter 4. Production of Levogluconesnone and 5-hydroxymethylfurfural in mixed solvent environments[2]

4.1. Introduction

In the previous chapter, we showed how water mediated the interconversion of the isomers LGO and HMF though an acid-catalyzed hydration-tautomerization-dehydration sequence. We suggested that this behavior might, in part, explain prior obversions that small amounts of water added to a THF solvent alter the reactivity of cellulose over sulfuric acid ,so that HMF replaces LGO as the major dehydration product in THF/water mixtures.(*1*) Following these insights, in this Chapter we explore the role of cellulose loading and water content in THF-H_2O mixtures on the reaction pathway underlying the production of LGO and HMF in detail. We also investigate the effect of different polar aprotic solvents and mineral acids on the production of HMF and/or LGO from cellulose. In addition, techno-economic analyses are carried out to demonstrate optimal process conditions for LGO and HMF co-production from cellulose, as expressed by a minimization in overall production costs.

4.2. Experimental

4.2.1. Dehydration reactions

Reactions were carried out in a 100 mL autoclave (Parr Instrument Company, series 4560, Hastelloy (C-276)). The vessel and head were dried overnight at 70 ˚C to remove residual water

[2] This Chapter was adapted from: J. He M. Liu, K. Huang, T. W. Walker, C. T. Maravelias, J. A. Dumesic, and G. W. Huber., "Production of levoglucosenone and 5-hydroxymethylfurfural from cellulose in polar aprotic solvent–water mixtures." *Green Chemistry* **19**, 3642-3653 (2017). J. He and G. W. Huber conceived the work. M. Liu and J. He carried out the experiments. T. W .Walker and J. He analyzed the reaction kinetics data and interpreted the kinetic solvent effects. K. Huang carried out the technoeconomic analyses. J. He and T. W. Walker wrote the manuscript. All authors assisted in editing the manuscript.

prior to each reaction. The reagents, including cellulose (Avicel® PH-101, moisture content ca. 3 wt%), glucose (Sigma Aldrich, anhydrous), LGA (TCI, purity 99%), THF (Acros, anhydrous, 99.9%, stabilized with BHT), and sulfuric acid (Fisher Chemical, A300-500), were used as received. Reactants, solvent, and sulfuric acid were added to the autoclave sequentially, at which point the vessel was purged five times with helium (99.999%, Airgas) and charged to 3.4 MPa He. The vessel was then heated to the desired reaction temperature and pressurized to a final pressure of 1000 psi He. Zero time was defined as the point at which the heating was started, except where noted. The temperature and stirring were maintained at 700 rpm and controlled by a Parr 4848 Controller. Around 1 mL samples were taken periodically through a dip tube during the reaction. The reactor was re-pressurized with helium to 1000 psi after each sampling. The samples were immediately quenched in a dry ice bath and then filtered with a 0.22 μm syringe filter (Restek, PTFE (polytetrafluoroethylene)) prior to the analysis.

4.2.2. Analytical methods

LGO (Standards, Apollo Scientific, purity 98%) and HMF (Standards, Sigma Aldrich, purity >99%) were analyzed using a high-performance liquid chromatograph (HPLC; Shimadzu, LC-20AT) equipped with UV (UV-vis; SPD-20AV) and RI (RID-10A) detectors. Separation was achieved using a Biorad Aminex HPX-87H column at 30 ˚C with 5 mM H_2SO_4 as the mobile phase, flowing at a rate of 0.6 $mL \cdot min^{-1}$. min^{-1}. The injection volume was 1 μL in each analysis.

All carbon yields were calculated as follows, where detectable products are glucose, LGA, LGO, HMF, levulinic acid, formic acid, and furfural:

$$Yield_i(\%) = 100 \times \frac{moles\ of\ carbon\ of\ product\ i}{Initial\ moles\ of\ carbon\ in\ feed} \quad (4.1)$$

overall yield $\Sigma\, Yield_i(\%) = 100 \times \frac{\Sigma(moles\ of\ carbon\ of\ product\ i)}{Initial\ moles\ of\ carbon\ in\ feed}$ (4.2)

4.3. Results and discussions

Shown in Figure 4.3.1 is the proposed reaction pathway for LGO and HMF production from cellulose in mixtures of water and polar aprotic co-solvents that we will discuss in this publication. All reactions occur in the presence of a Brønsted-acid catalyst. Cellulose initially undergoes anhydrous depolymerization to form LGA in the absence of water, or the presence of dilute H_2O (<5 wt.%) in polar aprotic solvents. It is noted that LGA has also been reported as the initial product of cellulose pyrolysis.(*2, 3*) LGA can be dehydrated to afford LGO, which oligomerizes to dimers/oligomers,(*4*) or polymerizes to form humins. Glucose, an HMF precursor, is the major secondary product formed from LGA hydration.(*5*) LGA can also undergo direct dehydration to form HMF,(*6*) furfural,(*7*) dimers, oligomers and humins.(*8*) Glucose can dehydrate to HMF or form humins.(*9*) HMF can undergo rehydration with water over an acid catalyst to produce levulinic acid (LA) and formic acid (FA). Furfural is a minor product formed from glucose (<24% furfural yield at 120 ˚C)(*10*) and/or LGA (<2% furfural yield at 170 ˚C),(*8*) which could also afford humins.(*6*) As will be shown in this paper, in THF with low concentrations of water (<5 wt.%) LGO isomerization is only a minor pathway for HMF formation.(*11*)

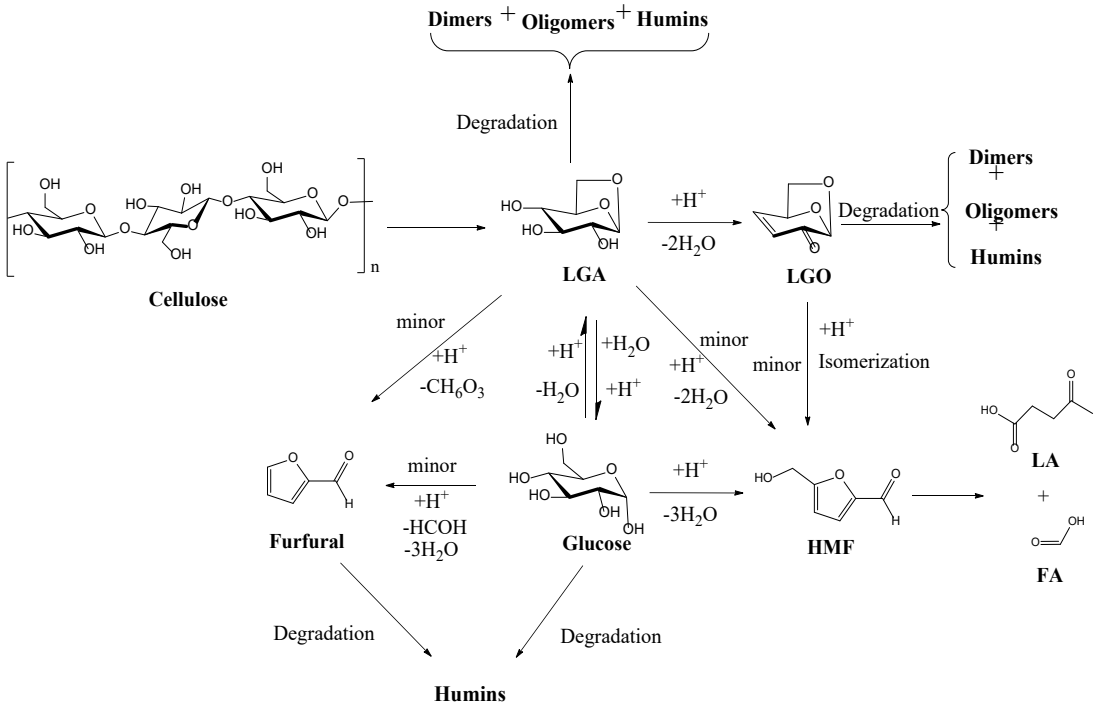


Figure 4.3.1. Proposed reaction pathway for LGO and HMF production in the presence/absence of low concentration of water. (the products from LGA)

4.3.1. Polar aprotic solvent effect on cellulose dehydration

The yields of all detectable products from cellulose conversion at 210 ˚C in different polar aprotic solvents are shown in Table 4.3.1. The product yields listed in Table 4.3.1 are at the reaction time when the maximum of LGO and HMF yields were obtained. Few detectable products were formed in DMSO, which may be attributable to the higher pK_a of H_2SO_4 in this solvent, and to the fact that DMSO is a considerably stronger base than water, so that fewer catalytic protons are available to initiate the acid-catalyzed reaction sequence which affords LGO, HMF and their precursors.(*12, 13*) Also of note, however, is the higher polarity and dipole moment of DMSO compared to the other common organic solvents,(*14*) as the reactivity of solvated molecules has in some cases been shown to generalize with these properties.(*15*) Another solvent with a relatively

high dipole moment (5.3 D), gamma-Valerolactone (GVL),(*16*) also led to a low LGO yield (15.5%) and HMF yield (11.9%). Noting these behaviors, we tested a series of solvents with relatively lower polarities (0.164-0.224), such as THF (0.207, 1.7 D),(*17, 18*) diglyme (0.224, 1.92 D),(*19*) tetraglyme (0.224, 1.92 D)(*20*) and 1,4-dioxane (0.164, 0.45 D),(*21*) which afforded higher yields of LGO (37.5-42.3%) and HMF (10.2-11.5%). In contrast to DMSO, the fact that appreciable fractions of identifiable products are formed in these solvents indicates that availability of catalytic entities in solution (acid protons) is not a relevant factor. It is noted that cyclopentyl methyl ether (CPME) afforded lower yields LGO (18%) and HMF (4%). With this exception in mind, however, it is generally the case that solvents characterized by moderate polarities/dipole moments afford greater yields of LGO and HMF. As such, we selected the subset of polar aprotic solvents in this category (THF, diglyme, tetraglyme and dioxane) for further study.

In diglyme (42.3%), tetraglyme (38.3%) and 1,4-dioxane (39.7%), slightly higher yields of LGO were obtained compared to THF (37.5%). However, the LGO yields decreased in diglyme, tetraglyeme and 1,4-dioxane (to 18.6% in 155 min, 16.9% in 102 min and 8.0% in 107 min, respectively). In contrast, the yield of LGO in THF decreased to only 35.2% in 158 min. Thus, the rate of LGO degradation was lower in THF than in diglyme, tetraglyme and 1,4-dioxane. Accordingly, for the remainder of this report, we focus on THF as a polar aprotic solvent for the selective production of LGO and HMF.

Table 4.3.1. Maximum LGO/HMF product yields for cellulose conversion in various solvents.

Solvent	Dipole Moment / *D*	Dielectric constant	Reaction time* (min.)	Yield (%)						
				LGO	HMF	LGA	Glucose	Furfural	LA	FA
THF	1.70	0.207	48	39.5	10.7	4.7	0.4	3.5	3.6	0.4
Diglyme	1.92	0.224	45	42.3	10.4	0.9	0.2	4.8	2.5	0.9
Tetraglyme	1.92	0.224	42	38.3	10.2	0	0	6.3	0.4	0.5
GVL	5.3		30	15.5	11.9	1.8	0.3	7.5	-	0.4
CPME	1.27		40	18.0	4.1	0	0	5.4	0.5	0.9
DMSO	4.10	0.444	50	0	0	3.3	0	0	0	0
1,4-dioxane[g]	0.45	0.164	47	39.7	11.5	0	0	5.2	2.0	0.4

Reaction conditions: Cellulose (1 wt.%), Solvent (60 mL), H_2SO_4 (64 μL Conc., 20 mM), 1000 psi He, 210 ˚C, 700 rpm. *

4.3.2. Acid effect on cellulose dehydration

We examined the effect of various acids including H_2SO_4, H_3PO_4, HCOOH and HCl on cellulose dehydration at 210 ˚C with 20 mM acid in THF as shown in Table 4.3.2. Our experiments show that an acid catalyst is necessary to produce LGA, LGO and HMF as confirmed by blank tests (without acid) with cellulose in THF, which resulted in negligible cellulose conversion (no detectable products) at 170 ˚C after 6h. The LGO/HMF yields decreased in the following order H_2SO_4 (37.5%/10.7%)>H_3PO_4 (10.4%/10%)>HCOOH (0%/1.1%). The LGO/HMF yields generally correlate with the proton donating ability of the acid (as expressed by their pKa's in H_2O at STP), which decreased in the following order H_2SO_4 (pKa = -3) > H_3PO_4 (pKa = 2.12) >

HCOOH (pKa = 3.75). The total yields of LGO, HMF and LGA are comparable (40.7-54.9%) with HCl, H_3PO_4 and H_2SO_4 (Table 4.3.2). The HMF yield (19.6%) in HCl is two times higher than in H_2SO_4 (10.7%) or H_3PO_4 (10%). This result may be explained by the known role of the chloride anion in stabilizing key transition states in liquid-phase carbohydrate conversions, in particular with regards to xylose dehydration to yield furfural.(*22*)

Table 4.3.2. Effect of acid on cellulose dehydration in THF.

Acid	**Reaction Time*** **(min.)**	**Yield (wt.%)**								
		LGO	**HMF**	**LGA**	**Glucose**	**Furfural**	**LA**	**FA**	**LGO and HMF**	**LGO, HMF and LGA**
H_2SO_4[a]	48	37.5	10.7	4.7	0.4	3.5	3.6	0.4	48.1	54.9
H_3PO_4[b]	160	10.4	10	20.5	0.3	0	1.0	0.3	20.4	40.9
HCOOH[c]	154	0	1.1	0.5	0	0.1	0	0	1.1	1.6
HCl[d]	170	5.3	19.6	15.8	1.9	3.4	0	0	24.9	40.7

Reaction conditions: Cellulose (1 wt.%), THF (60 mL), acid (20 mM), 1000 psi He, 210 ˚C, 700 rpm. * The reaction time for highest yield of LGO and HMF.

4.3.3. Role of water on the reaction pathway

Figure 4.3.2 shows the reaction time courses for cellulose and LGA conversion in pure THF, and in THF with 1 wt% H_2O at 170 ˚C with 7.5 mM H_2SO_4. Cellulose initially undergoes anhydrous depolymerization to form LGA. The LGA yield increases to a maximum of 23.4% at 40 min in pure THF and then gradually decreases to 11.4% in 90 min as shown in Fig 4.3.2**a**. LGO and HMF yields increase during the entire 90 minutes time course, suggesting that they are both formed from LGA. There is only a small amount of glucose (< 0.3%) observed during the first 90

mins. A control experiment for LGO conversion was performed in pure THF at 170 ˚C with 7.5 mM H_2SO_4. Only 0.5 % HMF yield was obtained with 25% LGO conversion. Thus, formation of HMF from LGO conversion is only a minor reaction pathway in pure THF.

According to Figure 4.3.2b, the primary product from cellulose conversion is still LGA in THF/H_2O (99/1) mixture. The LGA goes through a maximum of 26.5% yield at 45 mins and then decreases to 12.1% yield at 90 mins. However, the LGO yield is lower and the HMF yield is higher in the THF/H_2O (99/1) mixture than in the pure THF. Glucose (5.2% yield in 30 mins) yield is also higher in this solvent mixture. The HMF yield from cellulose increased from 11.4% in 90 mins in pure THF to 25.3% in the THF/H_2O (99/1) solvent. The LGA conversion time course in pure THF at 170 ˚C with 7.5 mM H_2SO_4 is shown in Figure 4.3.2c. LGA is converted to both LGO (31.5% yield in 90 mins) and HMF (5.5% yield in 90 mins.). The LGA conversion time course in THF/H_2O (99/1) mixture is shown in Figure 4.3.2d. The 25% LGO yield at 90 min obtained in the THF/H_2O (99/1) mixture is lower than the LGO yield (30.6% in 90 mins.) in pure THF. The much higher HMF yield in the presence of low concentrations of water (1 wt.%) is attributed to the dehydration of glucose formed from LGA hydration. Both soluble and insobluble humins are observed for reactions of both cellulose and LGA in both solvents systems.(*8*)

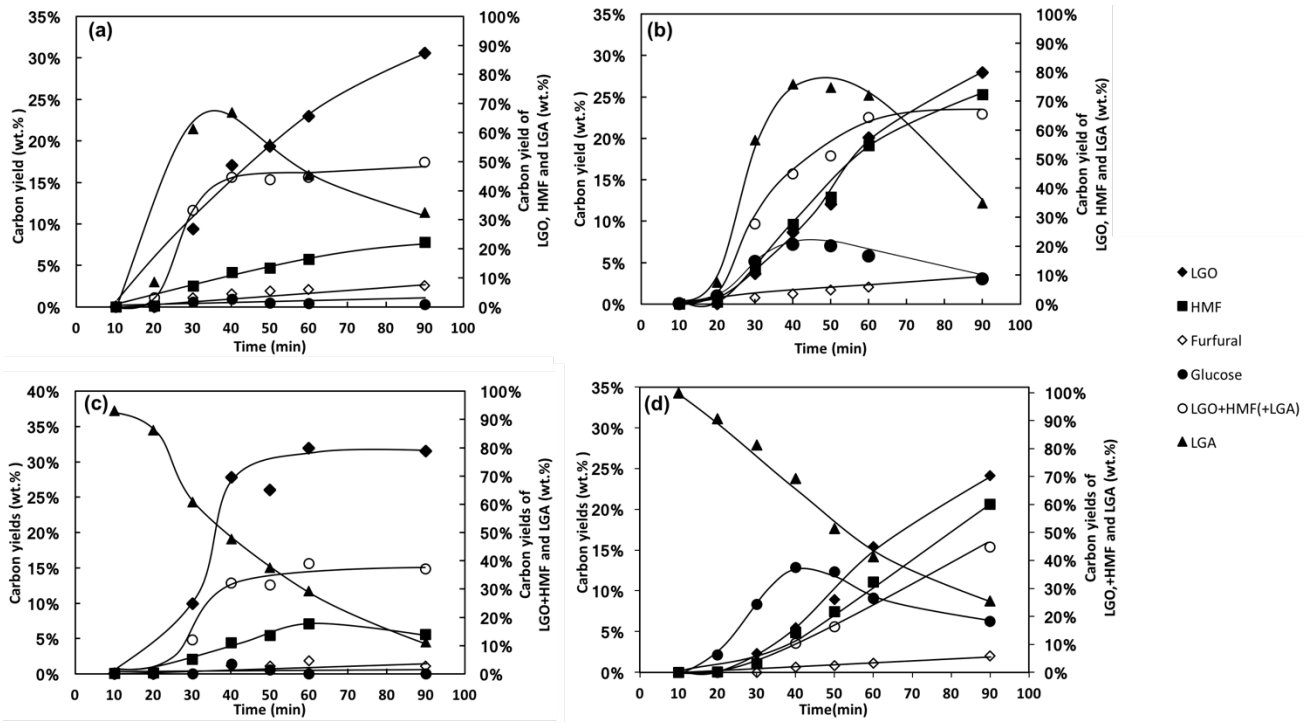


Figure 4.3.2. Influence of water on the conversion of (a) cellulose in pure THF (b) cellulose with 1 wt.% H_2O in THF (c) LGA in pure THF (d) LGA with 1 wt.% H_2O in THF. Reaction conditions: Cellulose (1 wt.%, 0.53 g) or LGA (0.4 wt.%, 0.2143 g), THF (60 mL), H_2SO_4 (24 µL Conc., 7.5 mM), 1000 psi He, 170 ˚C, 700 rpm.

Figure 4.3.3 shows the stability of LGO in THF/H_2O mixtures with and without H_2SO_4. Around 33.6% of the LGO degrades in 60 mins with pure THF without an acid. Only small amounts of HMF were observed in the HPLC during LGO conversion suggesting that most of the LGO was converted into low volatility humins. HMF was detected only when water was added to the THF in yields from 1.3% with 12.5% LGO conversion to 2.0% with 30.1% LGO conversion. The addition of 24 μL H_2SO_4, decreased the LGO degradation to 10.9% in 60 mins. This agrees with the work of Shafizadeh *et al.* who found that addition reactions of the C=C bond in LGO, which led to dimers, oligomers and humins, were catalyzed by bases (in the absence of acid).(*4*) By increasing initial water content from 1 wt.% to 5 wt.% in the presence of H_2SO_4 the amount of

LGO that degrades increases from 12.5% to 30.1% in 60 mins., respectively. The higher the initial water concentration the faster the LGO degradation. In summary, acids stabilize LGO in THF, while water could accelerate the degradation of LGO.

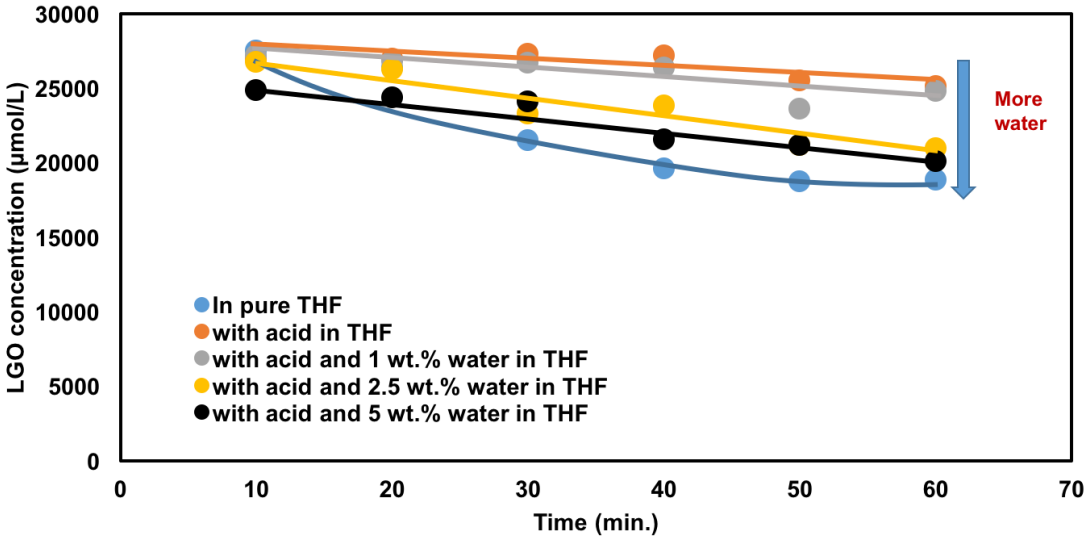


Figure 4.3.3. Influence of acid and water content on LGO degradation in THF. Reaction conditions: LGO (0.4 wt.%, 0.2143 g), THF (60 mL), H_2SO_4 (24 µL Conc., 7.5 mM), 1000 psi He, 170 ˚C, 700 rpm.

4.3.4. Effects of water content and cellulose loading on LGO and HMF production

Figure 4.3.4 shows the effect of the water content in the THF/water solvent and cellulose loading on the dehydration of cellulose in THF at 210 ˚C. The data in Figure 4.3.4 are at the reaction time where the highest yields of LGO and HMF are obtained for each solvent composition. The LGO yield decreases and the HMF yield increases (or goes through a maximum) with increasing water content in the solvent. The yields of furfural, levulinic acid and formic acid

typically decrease (or first goes through a maximum then decreases) with increasing water content of the solvent.

The yield of furfural increases from 5.1% with pure THF to the maximum of 5.7% with 1 wt.% water content and then gradually decreases to 3.3% with 10 wt.% water content. The yields of LA and FA are quite constant at 2.1% and 0.7% regardless of the water content of the reaction media. The sum of yields of LGO, HMF and LGA reaches the maximum of 65% in the presence of 1 or 2.5% water content.

Increasing the cellulose loading from 1 to 10 wt% causes a decrease in the LGO and HMF yields. Because the higher concentrations of cellulose reactant will generate higher concentrations of oxygenate intermediates, which readily participate in oligmerization/polymerization reactions to form soluble/insoluble humins. By adding 10 wt.% H_2O into pure THF, the HMF yields of 3 wt.%, 5 wt.% and 10 wt.% cellulose loadings will gradually increase from 13.2%, 12.4% and 10.0% to 41.6%, 31.1% and 18.5%, respectively. On the other hand, the LGO yields of 3 wt.%, 5 wt.% and 10 wt.% cellulose loadings will gradually decrease from 32.3%, 22.5% and 11.3% to 1.4%, 1.9% and 0.2%, respectively. In the presence of 1 wt.% or 2.5 wt.% H_2O, a maximum yield of LGO, HMF and LGA of 50.2%, 38%, 22.5% was achieved for 3 wt.%, 5 wt.% or 10 wt.% cellulose loadings, respectively. The furfural yield and FA yield is always around 5% and 1-2% for all cellulose loadings regardless of water content. In general, HMF yield will increase with increasing H_2O concentration from 0% to 10%, while LGO yield will decrease. A maximum sum yield of LGO, HMF and LGA could be achieved in the presence of 1 wt.% or 2.5 wt.% H_2O.

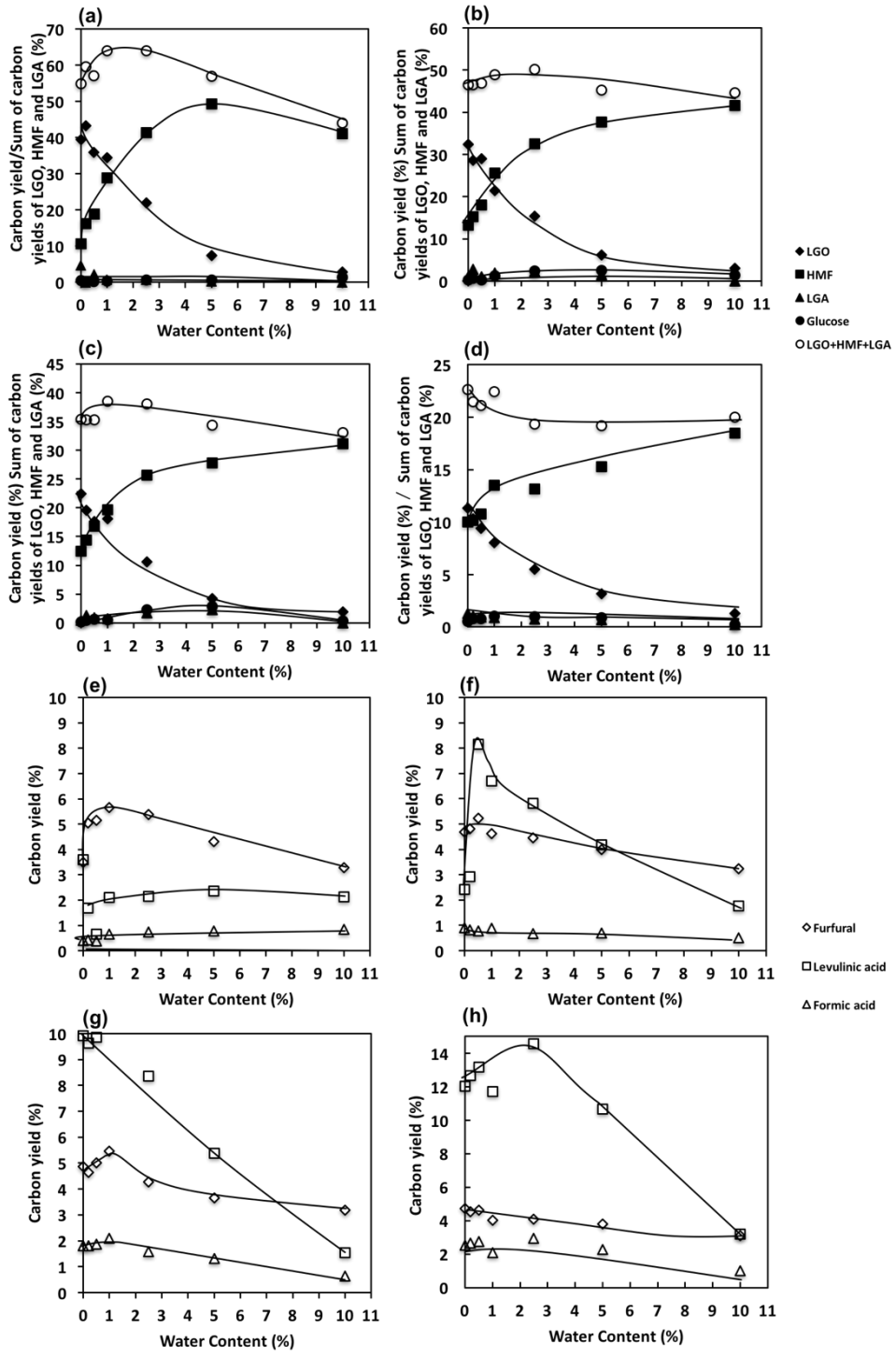


Figure 4.3.4. Influence of water content and cellulose loading on product yields for cellulose conversion. Reaction conditions: THF (60 mL), 210 ˚C, 1000 psi H_2, 700 rpm. Cellulose loading: (a, e) 1 wt.% (b, f) 3 wt.% (c, g) 5 wt.% (d, h) 10 wt.%.

Figure 4.3.5 shows the LGO and HMF yield as a function of time for cellulose conversion in a mixture of THF/H_2O (99/1). As shown in this figure LGO and HMF are formed at the same time demonstrating that HMF is not a secondary product from LGO.

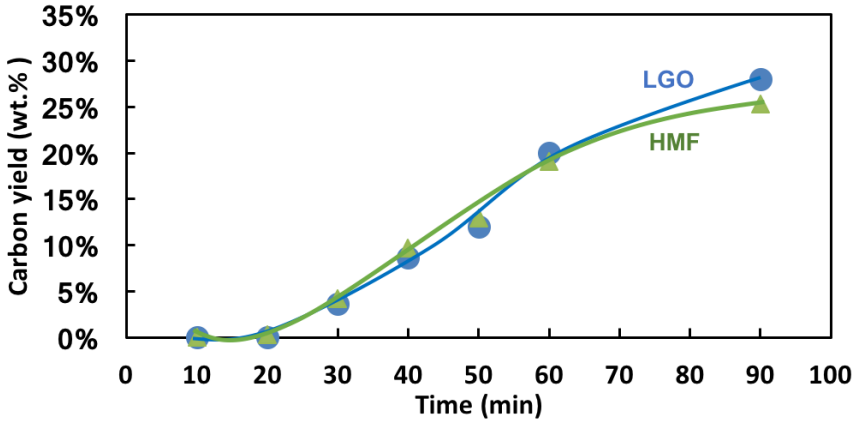


Figure 4.3.5. LGO and HMF yield during cellulose conversion in the presence of 1 wt.% H_2O. Reaction conditions: Cellulose (1 wt.%, 0.53 g) or LGA (0.4 wt.%, 0.2143 g), THF (60 mL), H_2SO_4 (24 µL Conc., 7.5 mM), H_2O (1 wt.%, 0.53 g), 1000 psi He, 170 ˚C, 700 rpm.

Figure 4.3.6 shows the conversion of LGO in a mixture of THF/H_2O (99/1). There is only 5% yield HMF obtained with 40% conversion of LGO further demonstrating that LGO isomerization to HMF is not a major reaction pathway for HMF formation when high amounts of polar aprotic solvents are used.

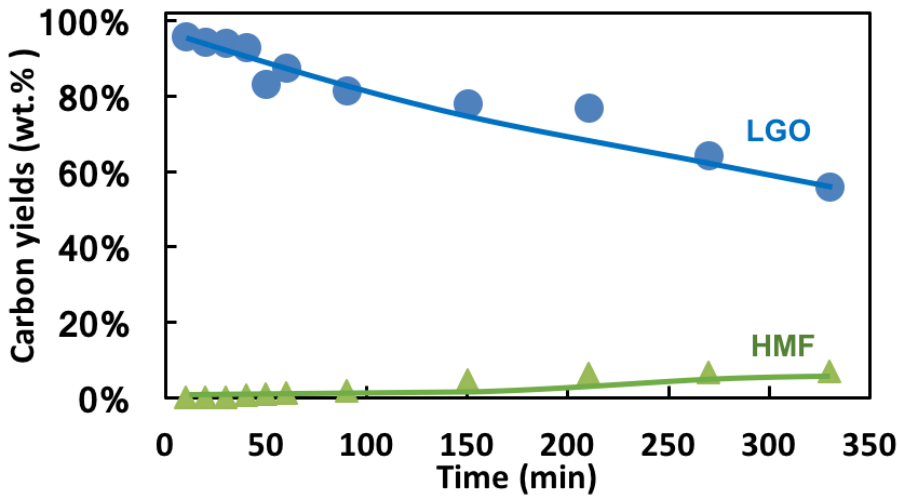


Figure 4.3.6. Conversion of LGO in the presence of 1 wt.% H_2O in THF. Reaction conditions: LGO (0.4 wt.%, 0.2143 g), THF (60 mL), H_2SO_4 (24 µL Conc., 7.5 mM), H_2O (1 wt.%, 0.53 g), 1000 psi He, 170 ˚C, 700 rpm.

4.3.5. Kinetic solvent effects in the THF/water system

The marked shift in LGO/HMF selectivity as a function of water content is attributable to either: a) changes in the glucose/LGA equilibrium; or b) changes in the rate of LGO isomerization to HMF (a water-mediated reaction). Having demonstrated that the latter is a negligible effect in the presence of dilute aqueous co-solvent, we now examine the role of water in the equilibrium-limited interconversion of glucose and LGA (precursors to HMF and LGO, respectively). As evidenced by Figure 4.3.2, the equilibrium between glucose and LGA develops quickly (<30 minutes) and stays constant throughout the reaction. So, as the amount of water in the solvent system increases, the relative amounts of LGO and HMF precursors formed from cellulose change proportionally. Table 4.3.3 compares the initial rates of LGO and HMF formation in THF/water mixtures to the equilibrium ratio of glucose to LGA in solution, which is constant throughout the reaction. If the change in HMF/LGO selectivity was only a function of the relative amounts of their chemical precursors in solution, then the initial rates of HMF and LGO production should

scale accordingly. As shown in Table 4.3.3, however, a 37-fold increase in the ratio of glucose to LGA corresponds to a 57-fold increase in the ratio of the formation rates of HMF and LGO. In other words, the rate of HMF formation increases with water content to a greater extent than can be explained by the amount of glucose in solution. This indicates that glucose not only becomes more abundant than LGA as water is added to the solvent system, but it becomes more reactive than LGA as well. Also of note is the 2-fold increase in the net rate of reaction (as expressed by the sum of LGO and HMF formation) that is effected by the addition of 5 wt% to the solvent system.

Table 4.3.3. Initial rates of HMF and LGO formation and equilibrium ratio of glucose to LGA concentrations as a function of water content.

wt% H_2O	$\left[C_{glucose}/C_{LGA}\right]_{eq}$	$r_{initial}^{HMF}$ (mmol L^{-1} hr^{-1})	$r_{initial}^{LGO}$ (mmol L^{-1} hr^{-1})	$r_{initial}^{HMF}/r_{initial}^{LGO}$	$r_{initial}^{HMF} + r_{initial}^{LGO}$
0	0.03	2.74	14.16	0.19	16.90
1	0.25	10.27	12.85	0.80	23.12
2.5	0.66	18.66	5.38	3.20	24.04
5	1.11	29.08	2.89	10.07	31.97

Reaction conditions: Cellulose (1 wt%, 0.53 g), THF (60 mL), H_2SO_4 (24 µL Conc., 7.5 mM), 1000 psi He, 170 ˚C, 700 rpm.

We note that glucose contains a greater number of hydroxyl groups than LGA, and these two species' interactions with water vs. THF molecules would therefore be characterized by different energetics. Vlachos and co-workers demonstrated how preferential solvation of hydroxyl

substituents by water molecules in 50/50 DMSO/water mixtures gives rise to an increased activity in the acid-catalyzed dehydration of fructose to yield HMF.(*23*) The details of how these types of solvent-solute interactions give rise to differences in the reactivity of glucose and LGA in THF/water mixtures are beyond the scope of this study, though they will certainly become the subject of concerted future research efforts. Furthermore, the fact that such dilute amounts of water in the THF solvent system induces large, solute-specific changes in reaction rates (and therefore selectivities) is of general significance from design standpoint, as will be demonstrated in the techno-economic analyses to follow.

4.3.6. Techno-economic analyses for LGO and HMF production from cellulose

We developed a conceptual process model for LGO and HMF production from cellulose follow five steps. Data for the proposed strategy come directly from the experimental work in this publication. First, a process flow diagram of the strategy was developed, as shown in Figure 4.3.7. The corresponding process models were developed using Aspen Plus Process Simulator (V8.8 Aspen Technology). The cellulose feed (Stream 1) is mixed with fresh makeup H_2SO_4 (Stream 2) and THF solvent (Stream 12) from THF accumulator (T-1) and supplied to the dehydration reactor (R-1), in which cellulose is converted to LGO, HMF, LGA, furfural, formaldehyde, LA, FA, and humin (the remainder of unconverted cellulose) at 170 °C and 1000 psi. According to the experimental results, H_2SO_4 is loaded at 1/30 mass ratio of cellulose, and the yields of dehydration products from cellulose depends on the cellulose loading (wt% in THF) and water content (wt% in THF and cellulose) at the reactor throat (Stream 3). Either a water makeup stream (Stream 12) or molecular sieve adsorber (S-3) is used to maintain the desired water content. The H_2SO_4 remaining in the raw product stream (Stream 4) is then neutralized by slaked lime (Stream 5). The

vapor fraction (Stream 7) of the neutralizer (R-2) product (Stream 6) is collected via a flash drum (S-1) and sent to THF accumulator (T-1), while the remaining liquid and humin mixture (Stream 8) are separated in the filter (S-2). The humins (Stream 9) are burned to produce heat and electricity in the boiler/turbogenerator if necessary. The filtered liquid (Stream 10) forms the feed to the product distillation columns (D-1 and D-2), where a product stream containing LGO, HMF and LGA is obtained at the bottom (Stream 17) with over 99% recovery. The top product of D-1 (Stream 11) is mainly THF and water azeotrope, which is recycled back to T-1, while the top fraction from D-2 (Stream 16) are the mixture of excessive water and other light components, such as LA, furfural, and FA. This stream can be sent to wastewater treatment or further utilized and processed if necessary. The base design of the process is based on the cellulose feedstock rate of 17,000 kg/hr, which is estimated to be equivalent to 1,000 dry tons of white birch per day.

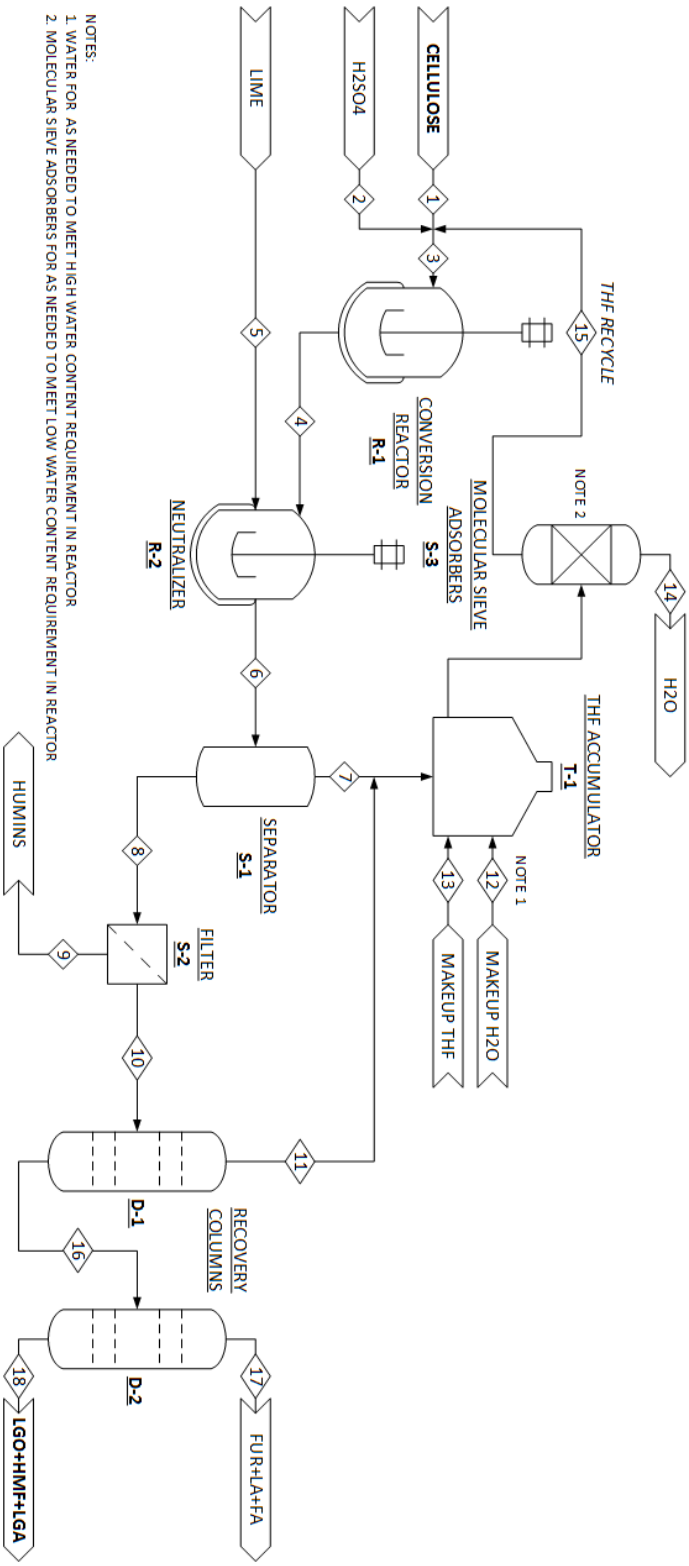


Figure 4.3.7. Overall process flow diagram

Equipment sizing and cost estimation were performed using Aspen Process Economic Analyze (V8.8 Aspen Technology) based on simulation results for the base case design. The cost parameters from using this design basis allows for detail to adjust individual equipment costs based on the scaling variables, thus resulting in the right track of the individual equipment costs on the different cellulose loading and water content. Thus, the installed cost for each equipment IC is calculated using an exponential scaling expression as shown in Eq. (4.3).

$$IC = \gamma PC_0 \left(\frac{SV}{SV_0}\right)^{\beta} \qquad (4.3)$$

PC_0 stands for the base equipment cost, which is the purchase cost of the equipment in the base case design. SV_0 stands for the base scaling value, which can be the flow rate through the equipment in the base case design. β is the scaling factor to calculate the purchased cost at different actual flow rate. γ is the installed factor. All the cost parameters (Table 4.3.4) are presented in 2015$ by using the Chemical Engineering Plant Cost Index. To determine the total capital investment (TCI), other additional direct and indirect costs are added as 46% (ACF) of TCI.(*24*) Therefore, Eq. (4.4) is used to calculate the TCI from IC.

$$TCI = \frac{\sum IC}{(1-ACF)} \qquad (4.4)$$

Then, the TCI is leveled to calculate yearly capital charge to normalize the effect of the capital investment with a capital charge factor (CCF) of 13.7% based on the cost parameters listed in Table S2. [60] The yearly capital charge CC is calculate using Eq. (4.5).

$$CC = CCF \cdot TCI \qquad (4.5)$$

The utility requirements of several process equipment are obtained from Aspen Plus simulations. Table 4.3.4 presents each requirement for such equipment. The utility cost associated with the costs for raw materials (feedstock and makeup chemicals) form the overall variable operating costs. The fixed operating costs, including labor and various overhead items, are assumed to be a fraction (2.5%) of TCI.

Table 4.3.4. Capital costs (2015$) and scaling factors for main equipment.

Equipment	Base cost ($) PC_0	Scaling value SV_0	Units	Scaling basis	Scaling factor β	Installed factor γ	Reference
R-1	48,658,000	601,902	kg/hr	feed	0.60	1.50	(*24*)
R-2	1,377,000	602,330	kg/hr	feed	0.70	1.50	Aspen Plus
S-1	144,000	602,330	kg/hr	feed	0.70	2.00	Aspen Plus
S-2	8,151,000	11,631	kg/hr	solid feed	1.00	1.70	(*24*)
S-3	285,000	16,856	kg/hr	adsorbed water	0.70	2.00	Aspen Plus
D-1	1,790,000	9,117	kmol/hr	vapor flow	0.60	2.40	Aspen Plus

D-2	538,000	129	kmol/hr	vapor flow	0.60	2.40	Aspen Plus
T-1	144,000	590,238	kg/hr	feed	0.70	2.00	Aspen Plus

* Cost parameters used for calculating capital charge factor are listed in **Table S4**.

A process synthesis model was developed to find the optimal combination of cellulose loading and water content. The objective function of the model was to minimize the unit production cost of the mixture of LGO, HMF and LGA since they all can be converted to valuable C6 chemicals under similar condition. A polynomial model was first developed to fit the yields of cellulose dehydration products regarding cellulose loading and water content. Then, material and energy balances were formulated for each equipment. Underwood's method was used to determine the scaling variable for sizing the distillation column, which is able to capture the impact of feed composition on the size and duty of a distillation column. This optimization model was also capable to simultaneously determine whether a water makeup stream (Stream 12) or molecular sieve adsorber (S-3) was required to maintain the optimal water content upon the feed (S-3) to dehydration reactor (R-1).

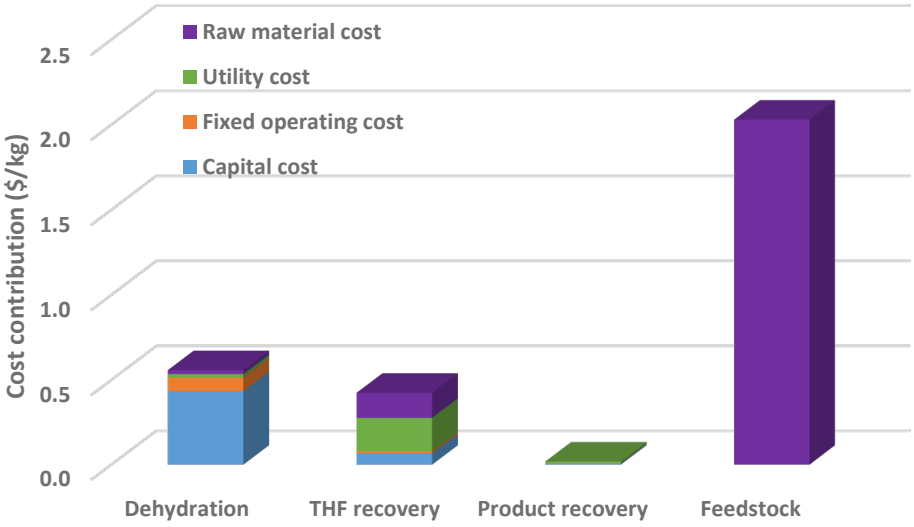


Figure 4.3.8. Cost contribution per process section ($/kg product).

Fourth, the optimal cellulose loading and water content were determined based on the process synthesis model. A 2.2% cellulose loading with 1.4% water content gives the lowest production cost of the LGO, HMF and LGA mixture at $3.032/kg. Figure 4.3.8 illustrates the contribution to the overall production cost by process section and capital, utility, raw material, and fixed operating costs. Cellulose feedstock is the primary cost contributor at $0.825/kg (translating to a cost contribution of $2.030/kg or 67.0 % to overall production cost). The next highest cost driver, contributing $0.434/kg (6.5%) to overall production cost, is the dehydration section due to the capital expense associated with conversion of the dilute cellulose/solvent mixture. The THF recovery section is also a significant cost contributor due to the utility cost ($0.198/kg) for separation and makeup THF ($0.148/kg). Since the feedstock cost weighs heavily on overall production cost, its price variances are further expanded over a larger range of values ($0-2.0/kg) to consider potential cellulose sources.

Figure 4.3.9 shows the minimum overall product cost and its associated optimal cellulose loading and water content as a function of cellulose price. The water content and cellulose loading decrease as the cellulose price increases. The production price increasing almost linearly with an increase in cellulose price. This shows that a higher cellulose price leads to lower cellulose loading and obtain higher product yields. If a relative cheap cellulose ($0.1/kg) is fed to the process, the minimum production cost decreases to $1.0/kg with 4.0% cellulose loading. However, the optimal cellulose loading cannot be beyond 5.0% regardless of the cellulose price, while optimal water content maintains 1.0%~2.0% for the proposed range of cellulose price.

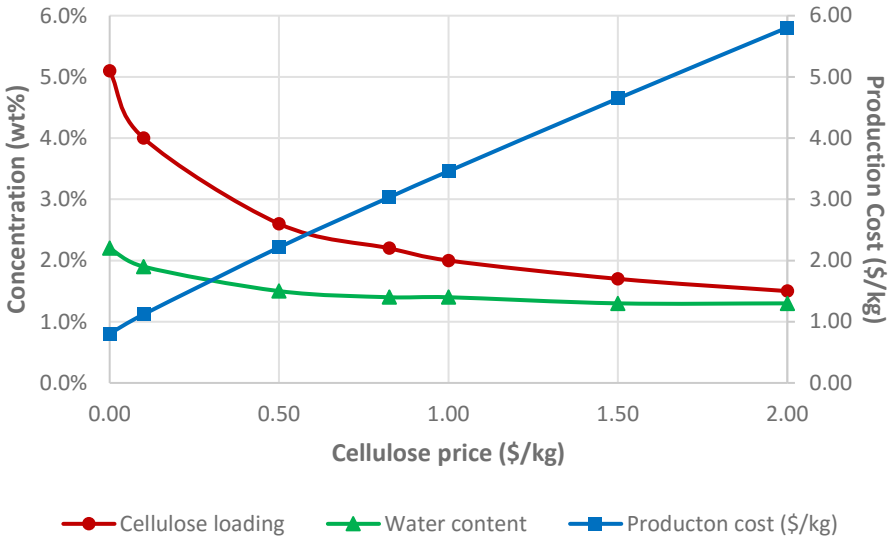


Figure 4.3.9. Optimal design scenarios as a function of cellulose price.

Finally, a sensitivity analysis was performed on the process to understand how the cellulose loading and water content impacts the economics. Figure 4.3.10 illustrates the optimal production cost as function of cellulose loading and water content. The point in the bottom left corner of the plot (2.2% cellulose loading with 1.4% water content) corresponds to the optimal design with

cellulose price at \$0.825/kg, discussed previously. The results show that the overall production cost can vary from \$3.03/kg to \$6.51/kg, while the optimal combination regions for cellulose loading and water content are 1%~3% and 0~5% respectively in order to keep production cost lower than \$3.50/kg. A high cellulose loading (>6.0%) leads to low product yields and therefore high unit production cost (>\$4.0/kg). Furthermore, this contour plot shows the cellulose loading accounts for larger impact on production cost than water content, since the cellulose loading leads to the trade-off between product yields and the costs of the reaction and recovery sections.

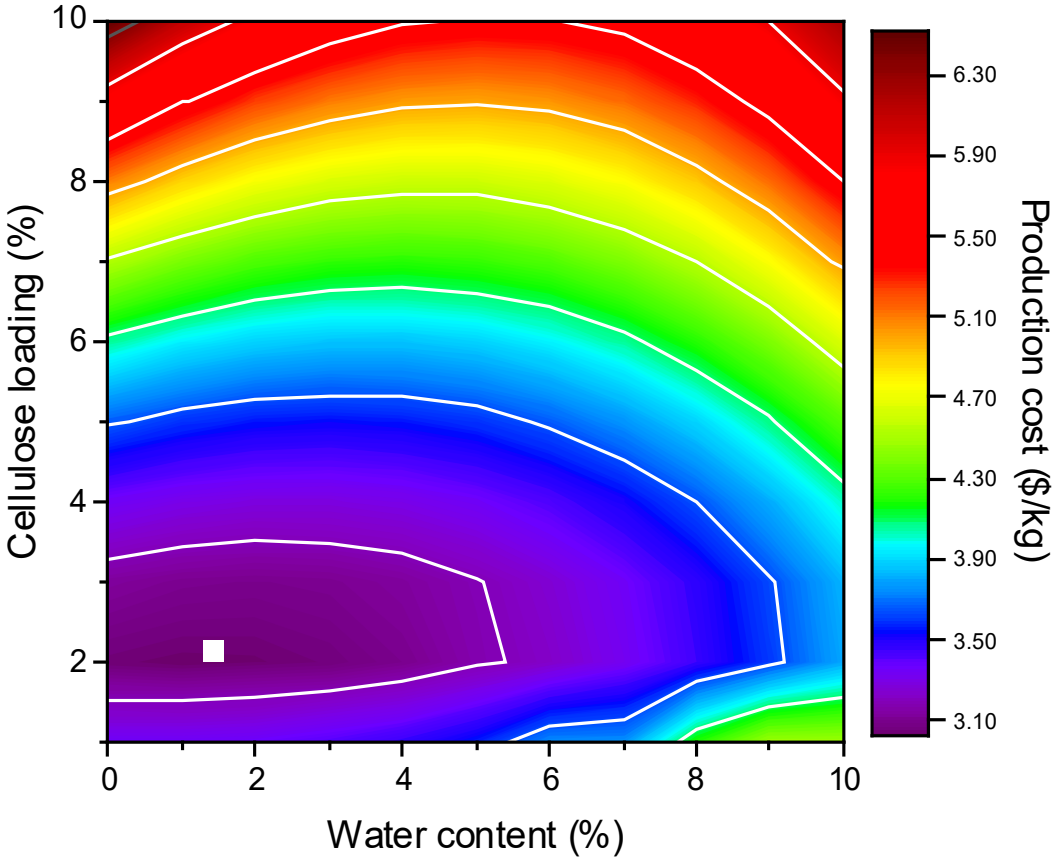


Figure 4.3.10. Production cost (\$/kg) as a function of cellulose loading and water content.

4.4. Conclusions

LGA is the primary product of cellulose decomposition in both pure THF solvent, and in mixtures of THF with dilute water (<5 wt%). LGO and HMF are produced from LGA dehydration in the absence of water. Increasing the water content up to 5 wt% increases the HMF selectivity, increases the glucose selectivity, and decreases the LGO selectivity. The ratio between glucose and LGA stays constant throughout the reaction demonstrating that this reaction is in equilibrium. The increased HMF yield in the presence of low concentrations of H_2O (<5 wt.%) is attributable to a) hydrolysis of LGA to yield the HMF-precursor glucose, and b) the enhanced reactivity of glucose compared to LGA in THF/water mixtures. The enhanced reactivity of glucose in the presence of an aqueous co-solvent is likely owing to preferential solvation of this hydrophilic substrate by water molecules. LGO isomerization is a minor reaction pathway for HMF formation at low water concentrations (<5wt.%). Maximum yields of LGO and HMF are achieved in the presence of 1 or 2.5 wt.% H_2O. Acid stabilizes LGO by hindering addition reactions of the C=C bond, while H_2O accelerates the degradation of LGO. LGO and HMF yields decrease with increased cellulose loadings.

THF, as a solvent with a moderate polarity/dipole moment, gives the highest total yields of LGO and HMF and attenuates the rate of LGO and HMF degradation compared to other organic solvents. Moreover, the low boiling point of THF makes downstream distillation separation more energy-efficient. The LGO/HMF yields are related to the proton dissociation abilities of the acids tested in this study, which decreased accordingly in the order H_2SO_4 > H_3PO_4 > HCOOH. The increased HMF yield in the presence of HCl compared to H_2SO_4 and H_3PO_4 is attributed to the known role of chloride anions solubilizing key transition states in biomass conversion schemes.

Finally, the results of techno-economic analyses for LGO and HMF production from cellulose have shown that the overall production cost can vary from $3.03/kg to $6.51/kg, depending on cellulose loadings and the water content of the solvent system. Optimal process conditions with respect to these two variables are 1%~3% cellulose loading and 0~5% water content, in order to keep production costs lower than $3.50/kg.

4.5. References

1. F. Cao *et al.*, Dehydration of cellulose to levoglucosenone using polar aprotic solvents. *Energy & Environmental Science* **8**, 1808-1815 (2015).
2. Y. Lin, J. Cho, G. a. Tompsett, PR Westmoreland, and GW Huber. *J. Phys. Chem. C* **113**, 20097 (2009).
3. R. Vinu, L. J. Broadbelt, A mechanistic model of fast pyrolysis of glucose-based carbohydrates to predict bio-oil composition. *Energy & Environmental Science* **5**, 9808-9826 (2012).
4. F. Shafizadeh, R. H. Furneaux, T. T. Stevenson, Some reactions of levoglucosenone. *Carbohydrate Research* **71**, 169-191 (1979).
5. S. Helle, N. M. Bennett, K. Lau, J. H. Matsui, S. J. Duff, A kinetic model for production of glucose by hydrolysis of levoglucosan and cellobiosan from pyrolysis oil. *Carbohydrate research* **342**, 2365-2370 (2007).
6. R. Weingarten *et al.*, Selective conversion of cellulose to hydroxymethylfurfural in polar aprotic solvents. *ChemCatChem* **6**, 2229-2234 (2014).
7. M. Käldström *et al.*, Formation of furfural in catalytic transformation of levoglucosan over mesoporous materials. *ChemCatChem* **2**, 539-546 (2010).
8. X. Hu *et al.*, Mediating acid-catalyzed conversion of levoglucosan into platform chemicals with various solvents. *Green chemistry* **14**, 3087-3098 (2012).
9. X. Qi, M. Watanabe, T. M. Aida, R. L. Smith Jr, Selective conversion of D-fructose to 5-hydroxymethylfurfural by ion-exchange resin in acetone/dimethyl sulfoxide solvent mixtures. *Industrial & Engineering Chemistry Research* **47**, 9234-9239 (2008).
10. E. I. Gürbüz *et al.*, Conversion of hemicellulose into furfural using solid acid catalysts in γ-valerolactone. *Angewandte chemie international edition* **52**, 1270-1274 (2013).
11. S. H. Krishna, T. W. Walker, J. A. Dumesic, G. W. Huber, Kinetics of levoglucosenone isomerization. *ChemSusChem* **10**, 129-138 (2017).
12. I. Kolthoff, M. Chantooni Jr, Second dissociation constant of sulfuric acid in acetonitrile and in dimethyl sulfoxide. *Journal of the American Chemical Society* **90**, 5961-5964 (1968).
13. I. M. Kolthoff, M. K. Chantooni Jr, S. Bhowmik, Dissociation constants of uncharged and monovalent cation acids in dimethyl sulfoxide. *Journal of the American Chemical Society* **90**, 23-28 (1968).
14. I. Kolthoff, T. Reddy, Acid-base strength in dimethyl sulfoxide. *Inorganic Chemistry* **1**, 189-194 (1962).

15. D. Y. Murzin, Solvent effects in catalysis: implementation for modelling of kinetics. *Catalysis Science & Technology* **6**, 5700-5713 (2016).
16. G. Strappaveccia *et al.*, γ-Valerolactone as an alternative biomass-derived medium for the Sonogashira reaction. *Green Chemistry* **17**, 1071-1076 (2015).
17. G. G. Engerholm, A. C. Luntz, W. Gwinn, D. O. Harris, Ring Puckering in Five-Membered Rings. II. The Microwave Spectrum, Dipole Moment, and Barrier to Pseudorotation in Tetrahydrofuran. *The Journal of Chemical Physics* **50**, 2446-2457 (1969).
18. F. E. Critchfield, J. A. Gibson Jr, J. L. Hall, Dielectric Constant and Refractive Index from 20 to 35° and Density at 25° for the System Tetrahydrofuran—Water1. *Journal of the American Chemical Society* **75**, 6044-6045 (1953).
19. I. Geoffroy, P. Willmann, K. Mesfar, B. Carré, D. Lemordant, Electrolytic characteristics of ethylene carbonate–diglyme-based electrolytes for lithium batteries. *Electrochimica Acta* **45**, 2019-2027 (2000).
20. D. Fish, J. Smid, Solvation of lithium ions in mixtures of tetraethylene glycol dimethyl ether and propylene carbonate. *Electrochimica Acta* **37**, 2043-2049 (1992).
21. S. Sudo *et al.*, Dielectric Properties of Ethyleneglycol−1,4-Dioxane Mixtures Using TDR Method. *The Journal of Physical Chemistry A* **111**, 2993-2998 (2007).
22. K. R. Enslow, A. T. Bell, The Role of Metal Halides in Enhancing the Dehydration of Xylose to Furfural. *ChemCatChem* **7**, 479-489 (2015).
23. S. H. Mushrif, S. Caratzoulas, D. G. Vlachos, Understanding solvent effects in the selective conversion of fructose to 5-hydroxymethyl-furfural: a molecular dynamics investigation. *Physical Chemistry Chemical Physics* **14**, 2637-2644 (2012).
24. D. M. Alonso *et al.*, Increasing the revenue from lignocellulosic biomass: Maximizing feedstock utilization. *Science advances* **3**, e1603301 (2017).

Chapter 5. Universal Kinetic Solvent Effects in Acid-Catalyzed Reactions of Biomass-Derived Oxygenates.[3]

5.1. Introduction

Acid-catalyzed reactions in the liquid phase are ubiquitous in the production and upgrading of renewable biomass-derived oxygenates,(*1, 2*) which have garnered interest as sources of organic carbon for the production of renewable chemicals and liquid fuels.(*3, 4*) As shown in the Chapters 3 and 4, an important variable to control the reactivity and selectivity for these catalytic processes is the solvent composition, which affects reaction rates,(*5, 6*) product selectivities,(*7, 8*) the stability of desired products,(*9, 10*) and the economics of downstream separations. While extensive knowledge has been accumulated regarding optimal solvent compositions for key applications,(*11*) it is not generally possible to anticipate how mixed-solvent environments will perform in *new* processes *a priori*, because the mechanistic details underlying kinetic solvent effects in multicomponent systems are not well understood.(*12, 13*) Computational efforts in the past decade have relied on *ab initio* quantum chemical methods to quantify the effects of solvent on the barriers to elementary reaction steps. These studies have provided detailed insights in important case studies,(*14, 15*) but broadly applicable design rules have not been developed due (in part) to the limitation of *ab initio* techniques to capture the slow intermolecular re-organizations that constitute solvent-solute interactions.(*16-18*) As a result, optimizing the composition of the

[3] This chapter was adapted from: T. W. Walker, A. K. Chew, X. Li, B. Demir, Z. C Zhang, G. W. Huber, R. C Van Lehn and J. A. Dumesic., Universal kinetic solvent effects in acid-catalyzed reactions of biomass-derived oxygenates. *Energy & Environmental Science* **11**, 617-628 (2018). J. A. Dumesic and T. W. Walker conceived the work. T. W. Walker and X. Li designed and carried out the reaction kinetics studies. A. K . Chew and R. C. Van Lehn design and carried out the molecular dynamics simulations. T. W Walker and A. K. Chew wrote the manuscript. All authors helped edit the manuscript.

liquid phase for new processes is typically effected through exhaustive and laborious experimentation.

In contrast to *ab initio* methods, classical molecular dynamics (MD) simulations can access longer time scales (~μs) and larger length scales (~nm), and at a lower computational cost. MD simulations therefore permit characterization of the solvent environment in the immediate vicinity of a reactant molecule (*i.e.,* the *local solvent domain* of the reactant), which can be compared to the solvent environment far from the reactant (*i.e.,* the *bulk solvent domain*).(*19-21*) For example, MD simulations have examined the preferred configurations of solvent molecules at biomass-relevant reactant surfaces.(*19, 22*) Based on key reaction rate measurements in this study and insights from our recent work,(*23*) we hypothesize that trends in acid-catalyzed reaction rates as a function of solvent environment can be related to the formation and properties of a water-rich or -deficient local domain near the reactant. These properties can be quantified *via* classical MD simulations, and moreover can be determined for a large series of possible reactants and solvent compositions given the computational efficiency of classical MD simulations.

Herein, we report the effects of mixed-solvent environments, consisting of water mixed with a polar aprotic cosolvent, on experimentally determined rates of Brønsted-acid-catalyzed reactions for seven biomass-derived model compounds. Irrespective of the selected cosolvent, consisting of 1,4-dioxane, γ-valerolactone, and tetrahydrofuran, we find that reactants with more available hydroxyl groups become more reactive as the water content of the solvent environment is decreased, and this behavior is true for both hydrolysis and dehydration reactions considered in this study. By contrasting properties between local and bulk solvent domains using MD simulations, we develop a computational approach to predict reaction rates as a function of solvent composition. We predict experimentally determined reaction rates using three computationally

determined observables: (1) the extent of water enrichment in the local solvent domain of the reactant (Γ); (2) the average hydrogen bonding lifetime between water molecules and the reactant (τ); and (3) the fraction of the reactant accessible surface area occupied by hydroxyl groups (δ). To our knowledge, the approach developed in this study provides the first tool of its kind in the context of biomass conversion in multicomponent solvent environments. As such, this study represents a step toward the model-predictive design of liquid-phase biomass conversion technologies. Moreover, this approach demonstrates that contrasting properties between local and bulk solvent domains using MD simulations can provide insight into the kinetics of acid-catalyzed reactions in mixed-solvent environments.

5.2. Methods

5.2.1. Reaction kinetics studies

Reactions were carried out in closed, 10 mL thick-walled glass reactors. In a typical experiment, an appropriate amount of reactant (*e.g.*, xylitol), acid catalyst, water and organic cosolvent (*e.g.*, 1,4-dioxane) were charged to the reactors, which were then sealed and placed in an oil bath at the appropriate temperature. Reactors were removed at times corresponding to the desired reaction time, and quenched in an ice bath at 273 *K*. Reaction products were analyzed using high-performance liquid chromatographs equipped with differential refractometers and photo diode array detectors, or gas chromatographs equipped with flame ionization detectors. All products were quantified using calibration curves with external standards. Conditions for each reaction (temperature, fractional conversion, etc.) were chosen so that each reaction was selective (>90%) to a single product. This procedure allowed for reliable measurements of rate constants based on the rate reactant consumption in a MATLAB-based optimization routine as detailed in the ESI. Trifluoromethane sulfonic (triflic) acid ($pK_{a,\,H2O}$ = -14.7, $pK_{a,\,DMSO}$ = -14.3, $pK_{a,\,MeCN}$ =

0.7) was used as catalyst in all experiments, which has been shown to behave as a strong acid even in pure polar aprotic solvents.(*24, 25*) We thus assume complete dissociation of the acidic proton in all mixed solvent environments, allowing for normalization of the apparent rate constants on a per-proton basis.

5.2.2. Molecular dynamics simulations

Classical molecular dynamics simulations were performed using Gromacs version 2016.(*26*) Reactants and cosolvents were parameterized using the CGenFF/CHARMM36 force fields(*27-29*) while water was modeled using the SPC/E(*30*) model. Solvent mixtures were initially equilibrated in an *NPT* ensemble using a Berendsen barostat at 1 bar and velocity-rescale thermostat at 300 K. The initial simulation box size containing the cosolvent and water was set to $(6\ \text{nm})^3$ in all simulations. A single reactant molecule was then added to the system, equilibrated with the same barostat and thermostat at the temperature of the reaction for 500 ps, followed by a *NPT* MD simulation with Parrinello-Rahman and Nose-Hoover as the barostat and thermostat, respectively, for 200 ns. The accessible surface area, radial distribution functions, and preferential exclusion coefficients were calculated using the final 190 ns of simulation data. For preferential exclusion coefficients, the simulation trajectory was partitioned into two separate trajectories of 95 ns to obtain the error in the calculations as the standard deviation of the calculated values. Simulation analysis was performed using the MDTraj tool box(*31*) and analysis tools developed in-house. A total of 63 MD simulations (~12.6 μs simulation time) was performed.

5.2.3. Preferential exclusion coefficient

Preferential exclusion coefficients were calculated using Equation (5.5) for each configuration in the final 190 ns of simulation data. n_W^L and n_C^L are calculated by counting the total

number of molecules within a cutoff distance of the reactant, where the cutoff between local and bulk domains is determined by the RDF for each system composition (Figure 5.3.5. (b)). Distances are computed between the center of mass of each molecule.

5.2.4. Hydrogen bonding lifetime

Hydrogen bonding lifetimes were calculated based on the Luzar and Chandler approach.(*32*) The kinetics of hydrogen bond formation and breaking is described in Equation (5.1):

$$\frac{dc(t)}{dt} = -kc(t) + k'n(t) \tag{5.1}$$

where k and k' are rate constants for breaking and making hydrogen bonds, respectively. $c(t)$ is the hydrogen bond correlation function that outputs the probability a hydrogen bond is intact at time t given that it was intact at $t = 0$. $n(t)$ is the probability that a hydrogen bond is broken but the hydrogen bonding groups are still within hydrogen bonding distance. The average bonding lifetime (τ_{HB}) could be found by the reciprocal of the forward rate constant ($1/k$) and calculated using Gromacs 5.0.1.(*33, 34*)

5.3. Results and Discussion

5.3.1. Universal effects of reactant and cosolvent properties on reactivity

Figure 5.3.1 shows the seven Brønsted-acid catalyzed reactions considered in this study, including those reported in our prior work.(*23*) For brevity, the reactant abbreviations in Scheme 1 are used throughout this text. Reactions were carried out in pure water, and in aqueous mixtures of the three polar aprotic cosolvents: 1,4-dioxane (DIO); γ-valerolactone (GVL); and

tetrahydrofuran (THF). Reactants in Figure 5.3.1 are organized according to decreasing hydrophilicity, as estimated by the *accessible hydroxyl fraction* (δ), which we have defined as the accessible surface area (ASA) occupied by the (N) hydroxyl groups in a reactant molecule ($ASA_{OH,k}$) normalized by the ASA occupied by the (M) total atoms in the molecule, shown in Equation (5.2).

$$\delta = \frac{\sum_{k=1}^{N} ASA_{OH,k}}{\sum_{l=1}^{M} ASA_{l}} \tag{5.2}$$

We interpret this dimensionless quantity as a qualitative *molecular descriptor* of the reactant hydrophilicity, normalized by its molecular size; larger values of δ indicate more hydrophilic molecules.

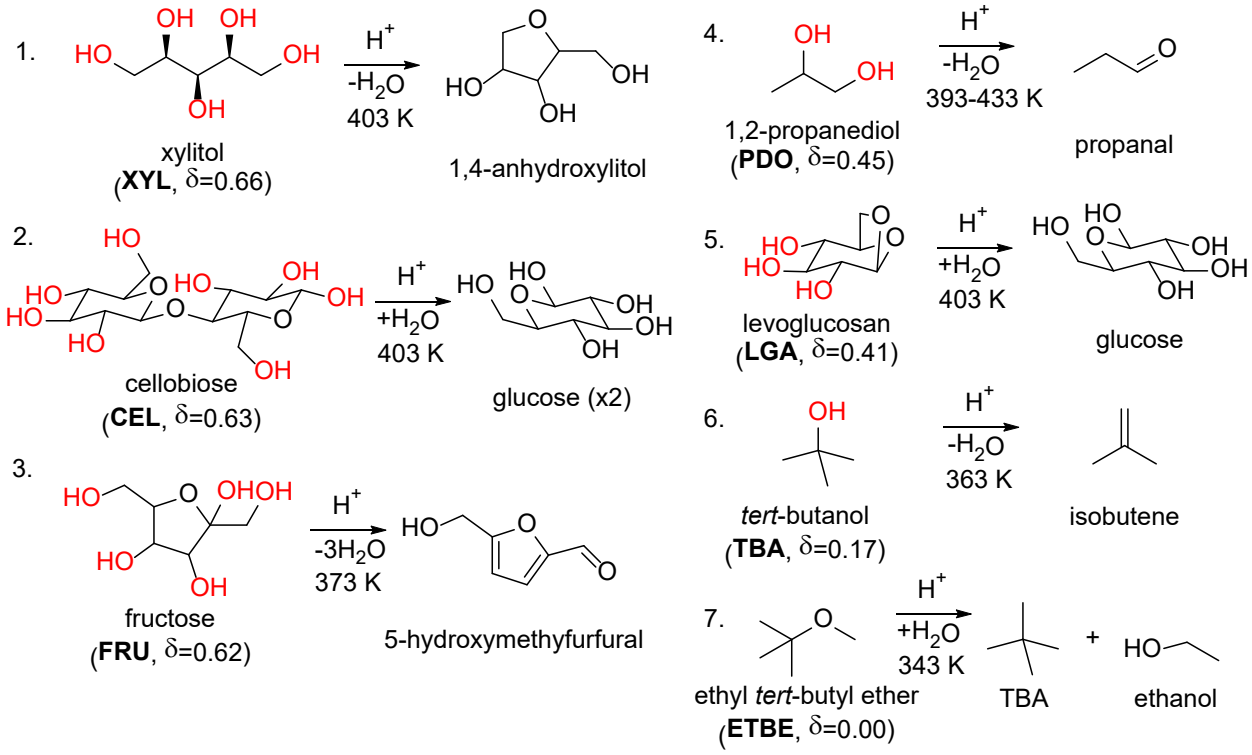


Figure 5.3.1. Brønsted acid-catalyzed reactions of seven model compounds. Rate constants associated with reactions 3, 4, and 6 were taken from prior work.(*23*) Hydroxyl groups are

highlighted in red for emphasis. Reactants are arranged according to decreasing hydrophilicity, as estimated by the δ parameter.

The forward rates of all reactions (r_i) in Figure 5.3.1 are described by apparent first-order kinetics with respect to the concentrations of the reactants (C_i) and acidic protons in solution (C_{H^+}). Accordingly, we measured apparent rate constants in each solvent environment by fitting the experimental reactions kinetics data (obtained in batch reactors) to expressions in the form of Equation (5.3), where ($k^i_{org,j}$) is the apparent rate constant for the i^{th} reaction, and the subscript denotes the identity and composition (in j^{th} mass fraction) of the organic cosolvent.

$$r_i = -\frac{dC_i}{dt} = k^i_{org,j} C_i C_{H^+} \tag{5.3}$$

Figure 5.3.2. shows the measured apparent rate constants for xylitol (XYL) dehydration in DIO/water mixtures ($k^{\mathrm{XYL}}_{\mathrm{DIO},j}$), normalized by the rate constant in pure water ($k^{\mathrm{XYL}}_{\mathrm{H_2O}}$), as a function of the mass fraction of DIO in the solvent environment (m_{DIO}). As the mass fraction of DIO increases, the value of the rate constant for XYL dehydration increases by nearly two orders of magnitude. In general, the values of the rate constants measured in this study are a strong function of solvent composition, and this phenomenon has previously been noted elsewhere.(*5, 6, 23*)

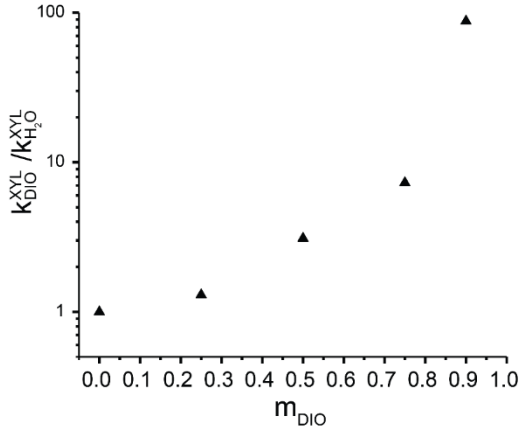


Figure 5.3.2. Apparent rate constant for XYL dehydration normalized by the rate constant in pure water versus the mass fraction of the organic cosolvent in DIO/water mixtures. *Reaction conditions: 75 – 200 mM XYL; 0.03 – 1.3 M trifluoromethane sulfonic (triflic) acid; 403 K.*

To compare reactivity trends across different reactions and solvent environments, the rate constants associated with each of the reactions in Figure 5.3.1 were described in terms of a kinetic solvent parameter ($\sigma^{i}_{org,j}$) as shown in Equation (5.4).

$$\sigma^{i}_{org,j} = \log_{10}\left(\frac{k^{i}_{org,j}}{k^{i}_{\mathrm{H_2O}}}\right) \tag{5.4}$$

Positive kinetic solvent parameters indicate an increase in the rate of reaction in a particular solvent environment compared to the same reaction in pure water, while negative values have a converse implication. Table 5.3.1 presents the kinetic solvent parameters for the rate of XYL dehydration in aqueous mixtures of up to 90 wt% DIO, GVL, and THF. These results demonstrate a general trend of increasing reactivity with decreasing water content for each of the three cosolvents.

Table 5.3.1 Kinetic solvent parameters for the Brønsted-acid-catalyzed dehydration of XYL in mixtures of water with three organic cosolvents. *Reaction conditions: 403 K; 0.015-1.3 M triflic acid. Confidence intervals were calculated at the 95% confidence level.*

$$\boldsymbol{k^{\mathrm{XYL}}_{\mathrm{H_2O}} = 1.04x10^{-4} \pm 9.3x10^{-6}\ L\ mol^{-1}s^{-1}}$$

Mass fraction of the organic phase	$\sigma^{\mathrm{XYL}}_{\mathrm{GVL},j}$	$\sigma^{\mathrm{XYL}}_{\mathrm{DIO},j}$	$\sigma^{\mathrm{XYL}}_{\mathrm{THF},j}$
0.90	2.05 +/- 0.07	1.80 +/- 0.08	1.85 +/- 0.10

0.75	1.02 +/- 0.08	1.02 +/- 0.08	0.74 +/- 0.13
0.50	0.41 +/- 0.09	0.50 +/- 0.08	N/A*
0.25	0.11 +/- 0.10	0.18 +/- 0.08	0.23 +/- 0.07

* Not available due to phase separation between water and organic cosolvent

Figure 5.3.3.(a) presents the kinetic solvent parameters for XYL and *tert*-butanol (TBA) dehydration in aqueous mixtures of DIO, GVL, and THF as a function of the mass fraction of the organic cosolvent (m_{org}). These two reactions represent upper and lower limits in our dataset with respect to reactant hydrophilicity and kinetic behavior. XYL has an accessible hydroxyl fraction (δ) of 0.66 and becomes monotonically more reactive with decreasing water content of the solvent environment. In contrast, TBA has an accessible hydroxyl fraction of 0.17, and its reactivity is a non-monotonic function of solvent composition.(*23*) In general, the extent to which the rate of TBA dehydration is affected by the increasing addition of organic cosolvent is smaller than that of XYL dehydration.

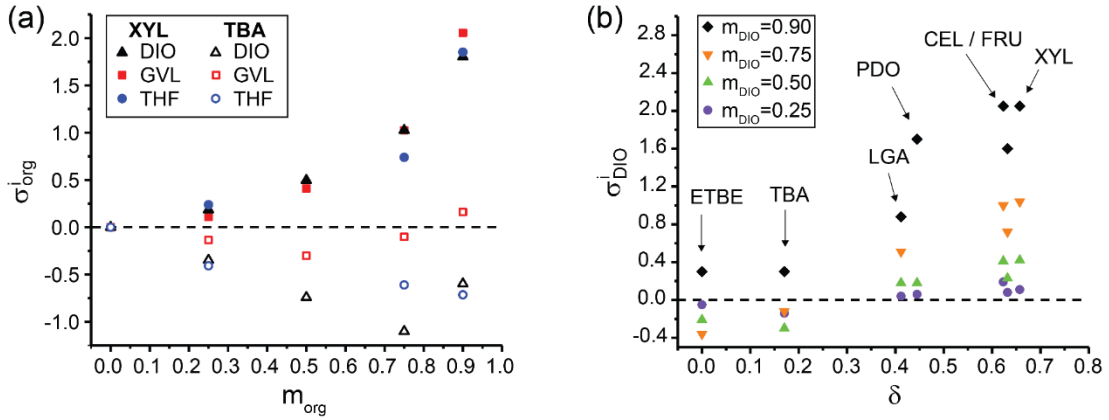


Figure 5.3.3. Kinetic solvent parameters (σ^i_{org}) as a function of: (a) solvent composition in aqueous mixtures of GVL, XYL and THF (open symbols = TBA, closed symbols = XYL), and; (b) the accessible hydroxyl fraction (δ) in DIO/water mixtures.

The comparison of XYL and TBA dehydration suggests that reactivity trends with respect to the water content of the solvent environment may depend on the hydrophilicity of the reactant, as we have postulated elsewhere.(*23*) Figure 5.3.3.(b) presents the kinetic solvent parameters as a function of the accessible hydroxyl fraction for all seven reactions in aqueous DIO mixtures. At a fixed solvent composition, the kinetic solvent parameters generally become more positive, indicating increasing rates of reaction compared to pure water, with an increasing accessible hydroxyl fraction. Similar behavior was observed in GVL and THF mixtures. Note that this behavior is observed for both the dehydrations *and* hydrolysis reactions considered in this study.

5.3.2. Proposed mechanism: reaction rates correlate with formation of water-enriched local solvent domain

To understand the aforementioned trends in reactivity, we note that rates of chemical reactions are controlled by the thermodynamic properties of the reactants, catalyst, and transition states in the elementary steps.(*13*) A quantitative understanding of solvent effects therefore requires knowledge of the reaction mechanism, and a rigorous characterization of the transition-state-solvent interactions in each of the kinetically relevant steps.(*35, 36*) In many cases, however, catalytic reactions occur *via* a sequence of elementary steps where a *single* step is rate-limiting, and the rate is then controlled by the thermodynamic properties of the transition state for this step.(*37-39*) In these cases,(*40*) the reaction mechanism may be analyzed *via* a sequence of quasi-equilibrated steps, such that the reactant and proton are treated as being in equilibrium with the transition state. Furthermore, transition states in acid-catalyzed reactions typically display strong carbocation-like character.(*39, 41*) Accordingly, as shown in Figure 5.3.4.(b), we consider the

acid-catalyzed reactions in this study as being composed of two generalized steps, both of which can be impacted by the solvent:

(i) the transfer of a proton from the bulk domain to the reactant, and;

(ii) the formation of a carbocation-like transition state.

For step (*i*), we have shown that water-enriched *local solvent domains* form around hydroxyl groups in the presence of polar aprotic cosolvents.(*23*) In mixed-solvent environments, hydrophilic reactants thus drive the formation of local solvent domains in which the local density of water molecules near the reactant is greater than the density of water molecules in the bulk solvent, as shown in Figure 5.3.4.(a). A proton is therefore destabilized in the bulk solvent relative to the local domain, because of its higher affinity for water than for the organic phase, leading to a thermodynamic driving force for transferring the proton to the reactant.(*42-44*) Vlachos and co-workers have postulated a similar mechanism for fructose dehydration in DMSO-water mixtures.(*21*)

For step (*ii*), we hypothesize that the stability of the reactant, proton, and transition state are correlated in these local solvent domains, because water molecules that bind strongly to the reactant are preorganized into configurations that stabilize charged transition states. Similar solvent preorganization is thought to contribute to enzyme catalytic efficiency,(*45*) has been speculated as a key effect in the acid-catalyzed glucose to fructose isomerization reaction,(*17*) and may contribute to the enhanced reactivity of hydronium ions in confined environments (zeolites).(*46*)

We now explore whether reactant-solvent-cosolvent interactions can be tuned to deliberately drive the formation of water-enriched local domains near hydrophilic reactants. Water and the hydrophilic reactant are then characterized as being confined to a local domain, with

confinement enhancing reactivity by increasing reactant-proton association, and stabilizing the carbocation-like transition states common to acid-catalyzed reaction mechanisms. With MD simulations, we probe our hypotheses by analyzing the local solvent domain near the reactant and deriving simulation measurables that can be used to predict experimental reaction rates. Simulation snapshots for XYL in pure water and 90 wt% DIO are shown in Figure 5.3.4.(c) and Figure 5.3.4.(d), respectively.

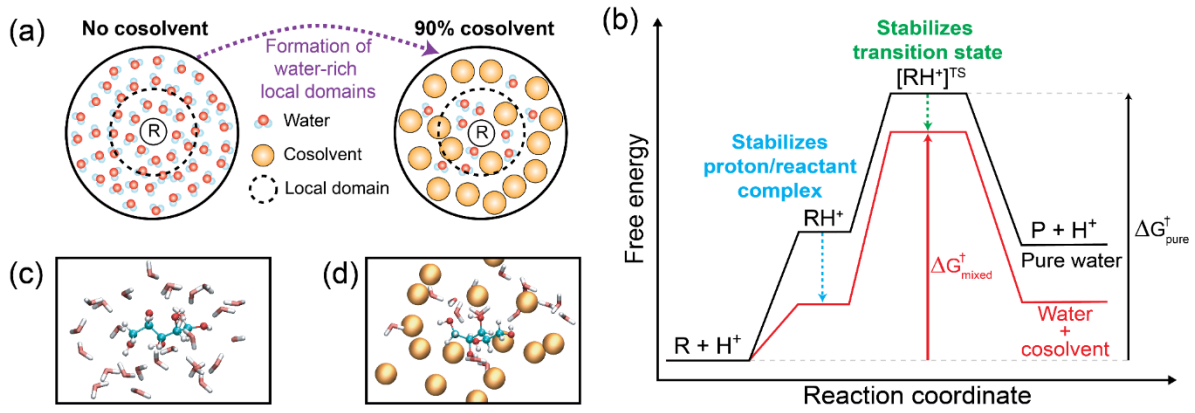


Figure 5.3.4. (a) Role of cosolvent molecules on the distribution of solvent molecules. Favorable interactions with hydrophilic reactants in mixed-solvent environments drive the formation of water-rich local domains around the reactant. While there are fewer water molecules in the local domain relative to pure water, the local water density is enriched relative to the bulk density in the solvent mixture. (b) Proposed effect of cosolvent molecules on a reaction free energy landscape. Stabilization of the proton and transition state in the water-rich local domain, relative to the bulk domain, lowers the apparent free energy barrier for the reaction in a mixed-solvent environment. (c, d) MD simulation snapshots of XYL in (c) pure water and (d) 90 wt% DIO, which is drawn as a single representative bead to match the schematics in (a).

5.3.3. MD simulations: formation of water-enriched local domains in solvent mixtures

Figure 5.3.5 shows the radial distribution function (RDF), which quantifies the density of water molecules at a radius r away from a central molecule normalized by the bulk water density. The RDF is schematically depicted in Figure 5.3.5(a) for a mixed-solvent environment in which XYL is the central molecule. The RDF for XYL in a 90 wt% DIO/water mixture and in pure water (0 wt% DIO/water) is shown in Figure 5.3.5(b). We define the cutoff, r_{cutoff}, between the local and bulk solvent domains as the distance at which the RDF reaches unity.

RDFs between TBA-water and XYL-water in various DIO/water mixtures are shown in Figures 5.3.5(c) and (d), respectively. From the RDFs, we find that the water content in the local solvent domain of each reactant increases compared to the bulk solvent domain when a high concentration of cosolvent is present in the solvent environment. This behavior is apparent from the increase in the magnitude of the first solvation peak for systems containing large concentrations of DIO relative to its magnitude for a pure water system, indicating that water preferentially partitions to the local solvent domain around the reactant. The RDF for XYL has a broader first solvation peak than TBA, which indicates greater water enrichment in the local solvent domain when the reactant has more hydroxyl groups.

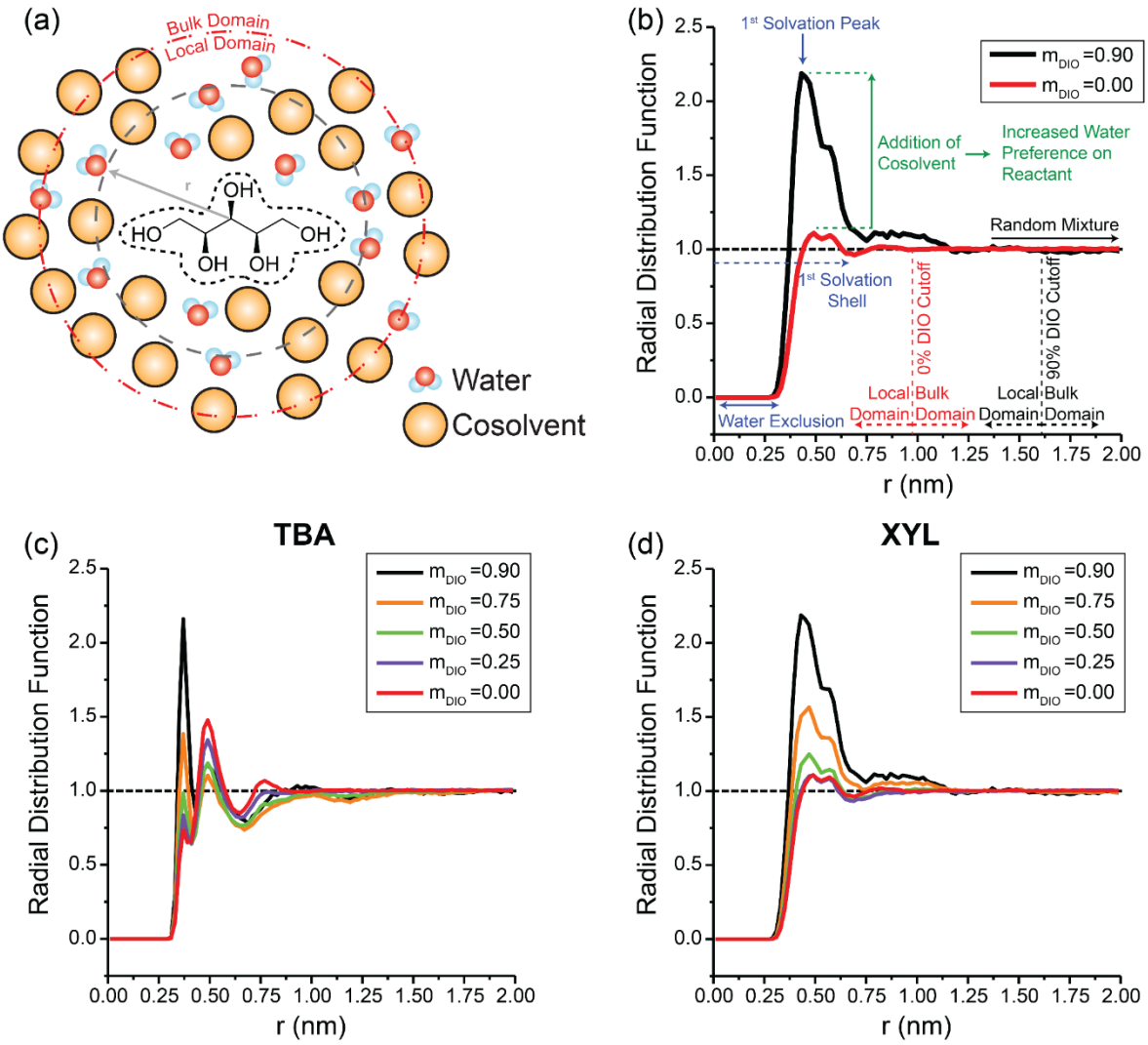


Figure 5.3.5. (a) Schematic depiction of the radial distribution function for XYL. The distance, *r*, is calculated between the center of mass of the reactant and the oxygen atom of each water molecule. (b) Radial distribution function for XYL in 90 wt% DIO and pure water (m_{DIO}=0). The cutoff between local and bulk domain is defined as the distance when the RDF between the reactant and water reaches unity (i.e. a random mixture). (c,d) Radial distribution function for TBA (c) and XYL (d) for various wt% of organic solvent.

5.3.4. Quantifying water enrichment within the local domain of the reactant

Because RDFs are difficult to compare across reactants with different concentrations of cosolvents, we calculated the preferential exclusion coefficient (Γ) as a molecular descriptor to quantify the local domain composition around the reactant. We define Γ as the excess number of cosolvent molecules within the local solvent domain of the reactant relative to the bulk solvent domain. Preferential exclusion coefficients are calculated from the MD simulations according to Equation 5.5, where n_C and n_W are the total number of cosolvent and water molecules, and the superscripts L and B indicate molecules within the local and bulk domains, respectively.(*47-49*)

$$\Gamma = -\left\langle n_C^L - n_W^L \left(\frac{n_C^B}{n_W^B}\right)\right\rangle \tag{5.5}$$

Positive values of Γ indicate that the concentration of cosolvent is lower in the local solvent domain of the reactant than in the bulk solvent domain. Positive Γ is also referred to as preferential hydration,(*50*) because the exclusion of cosolvent indicates that the reactant has a higher affinity for water. Negative values of Γ indicate that the concentration of cosolvent is higher in the local solvent domain of the reactant than in the bulk solvent domain and that the reactant has a higher affinity for the cosolvent.

Preferential exclusion coefficients (Γ) calculated for TBA and XYL in DIO/water and GVL/water mixtures at various cosolvent concentrations are shown as filled lines in Figures 5.3.5(a) and (b). Experimentally determined kinetic solvent parameters (σ) are depicted as dashed lines for comparison. For TBA, Γ is negative with the exception of 90% GVL/water. Conversely, Γ is positive for XYL across the range of cosolvent compositions, which means that XYL preferentially excludes cosolvent and has a higher affinity for water. We find that Γ correlates well with σ, even capturing the non-monotonic behavior in TBA.

Figures 5.3.6(c) and (d) show a strong linear correlation between simulation-derived Γ and the experimentally determined σ, as indicated by Pearson correlation coefficients (Pearson's *r*). Pearson correlation coefficients close to 1 indicate a total positive linear correlation, whereas values near -1 indicate a linear negative correlation, and zero indicates no linear correlation. Pearson correlation coefficients for all reactants in each of the three cosolvent mixtures are summarized in Table 5.3.2. We find that Γ and σ are highly correlated (Pearson's $r \geq 0.80$) for the majority of the systems, with the exception of THF/water systems.

The agreement between simulations and experiments suggests that higher reaction rates correspond to the enrichment of water near the reactant despite differences in reactant hydrophilicity and reaction mechanisms. However, the poor correlation between Γ and σ in some reactant/cosolvent environments suggests that additional parameters that characterize the local solvent domain will improve our understanding of the solvent effects that contribute to experimentally determined reaction rates.

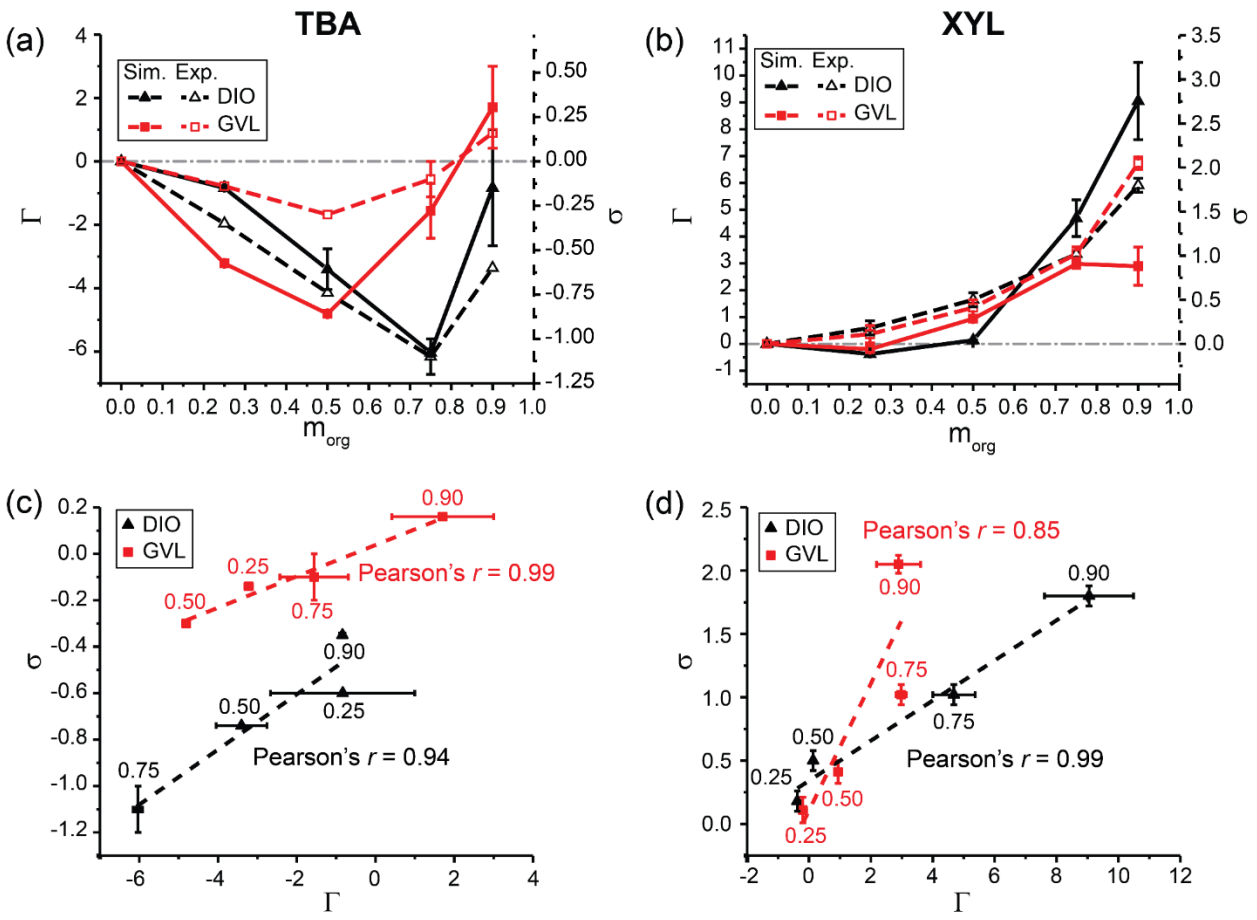


Figure 5.3.6. Relationship between experimentally determined kinetic solvent parameters (σ) and simulated preferential exclusion coefficient (Γ) for (a) TBA and (b) XYL for various wt% of organic cosolvent in GVL/water and DIO/water mixtures. The gray dotted line denotes when σ and Γ are zero. Kinetic solvent parameters are also plotted against Γ for (c) TBA and (d) XYL. The dashed lines in (c) and (d) represent the best-fit line. Data points are labeled with the wt% of the organic cosolvent. Pure water systems have been omitted from parts (c) and (d) because σ and Γ will always be zero.

Table 5.3.2. Pearson correlation coefficient between the kinetic solvent parameters (σ) and the preferential exclusion coefficient (Γ) for various reactants and cosolvents.*

	Reactant†							
Cosolvent	**ETBE**	**TBA**	**LGA**	**PDO**	**FRU**	**CEL**	**XYL**	**All****
DIO	-0.08	0.94	0.84	0.99	0.96	0.98	0.99	0.84
GVL	0.80	0.99	0.91	0.92	0.93	0.80	0.85	0.72

5.3.5. Quantifying reactant-water hydrogen bonding strength

Following the hypothesis that charged transition states may be stabilized by water molecules in the local solvent domain that are preorganized into favorable binding configurations, we next calculated the average reactant-water hydrogen bonding lifetime (τ_{HB}) as a molecular descriptor to quantify the strength of water binding to the reactant in the mixed-solvent systems.(*51*) We expect that hydrogen bonds between the reactant and water are stronger (*i.e.,* longer-lasting) in mixed-solvent environments that have large cosolvent concentrations, because water-water hydrogen bonds are unable to form, increasing the preference of reactant-water hydrogen bonds. Stronger interactions between water and reactant may translate to a lower transition state free energy and thus, an increase in reaction rate.(*52, 53*)

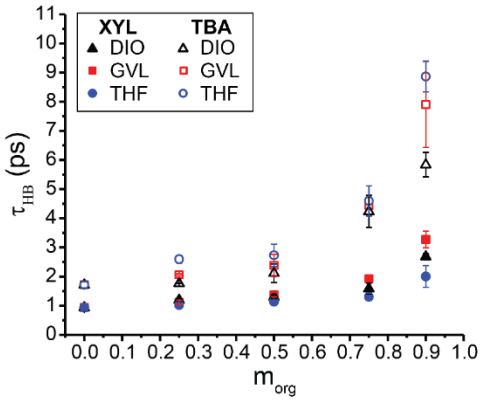


Figure 5.3.7. Average reactant-water hydrogen bond lifetimes (in picoseconds) for TBA and XYL as a function of mass fraction of organic cosolvent in DIO/water, GVL/water, and THF/water mixtures.

Hydrogen bonding lifetimes for TBA and XYL are shown in Figure 5.3.7. When increasing the concentration of the organic cosolvent, hydrogen bonding lifetimes between the reactant and the surrounding water molecules increase monotonically across different cosolvent environments, indicating stronger reactant-water hydrogen bonds. XYL has a lower hydrogen bonding lifetime compared to TBA, possibly due to a higher reaction temperature for XYL (403 K vs. 363 K); higher system temperatures often result in lower hydrogen bonding lifetimes. To remove temperature effects, we define the hydrogen bonding lifetime ratio, τ, by normalizing the hydrogen bonding lifetime in the organic cosolvent mixtures ($\tau_{\mathrm{HB,org}}$) by the hydrogen bonding lifetime of the same reactant in pure water ($\tau_{\mathrm{HB,H_2O}}$), as shown in Equation ().

$$\tau = \frac{\tau_{\mathrm{HB,org}}}{\tau_{\mathrm{HB,H_2O}}} \tag{5.6}$$

A monotonic increase in hydrogen bonding lifetime ratio with respect to cosolvent fraction was observed for reactants across all cosolvent environments, confirming that reactant-water hydrogen

bonding strength generally increases as the amount of available water in the mixture decreases.

Multidescriptor correlation between experimental and simulation results

Since both Γ and τ measure contributions to experimentally determined reaction rates, we explored the use of these descriptors in combination to improve the correlations in Table 5.3.2, as shown in Equation (5.7).

$$\sigma_{\text{pred}} = A + Bb + Cc + \cdots \tag{5.7}$$

A, B, and C are coefficients that quantify the relationship between the simulated molecular descriptors *b* and *c* and experimental data (σ). Coefficients for the multidescriptor correlation are obtained by linear regression between the simulated descriptors and experimental kinetic solvent parameters. The model accuracy is then assessed by using the coefficients to calculate a predicted kinetic solvent parameter, σ_{pred}, and plotting it against the experimentally determined kinetic solvent parameter σ_{exp}.

To determine if hydrogen bonding strength improves the correlation between simulation results and experimental models, we define two models: Equation (5.8) uses a single descriptor, Γ and Equation (5.9) uses two descriptors, Γ and τ.

$$\sigma_{\text{pred}} = A + B(\Gamma) \tag{5.8}$$

$$\sigma_{\text{pred}} = A + B(\Gamma) + C(\tau) \tag{5.9}$$

As shown in Table 5.3.3, the slope between σ_{pred} and σ_{exp} for most reactant/cosolvent systems is close to unity and the root-mean-square error (RMSE) between the predicted and experimental values is lowered when using the two-descriptor model in Equation (5.9) compared to only fitting a single descriptor in Equation (5.8). Therefore, the addition of the hydrogen bonding lifetime

ratio, which characterizes the binding strength between the reactant and water within the local domain, improves the overall correlation between σ_{pred} and σ_{exp}.

Table 5.3.3. Best-fit slope and root-mean-square error (RMSE) between predicted kinetic solvent parameters (σ_{pred}) and experimental kinetic solvent parameters (σ_{exp}).

		1 Descriptor Fit		2 Descriptor Fit	
		Equation (5.8)		Equation ()	
Cosolvent	Reactant	Slope*	RMSE**	Slope*	RMSE**
DIO	ETBE	0.01	0.11	0.17	0.10
	TBA	0.89	0.09	0.99	0.03
	LGA	0.70	0.11	0.98	0.03
	PDO	0.97	0.05	N/A	N/A
	FRU	0.92	0.16	1.00	0.03
	CEL	0.97	0.08	0.97	0.07
	XYL	0.98	0.09	0.98	0.08
GVL	ETBE	0.63	0.14	0.63	0.14
	TBA	0.97	0.03	0.98	0.02
	LGA	0.82	0.14	0.95	0.07
	PDO	0.85	0.29	N/A	N/A
	FRU	0.86	0.27	1.00	0.00
	CEL	0.64	0.35	0.97	0.11
	XYL	0.73	0.39	1.00	0.00

* Highlighted slopes have a value less than 0.90

** Highlighted RMSE has a value greater than the experimental error of ≈0.10

† N/A: Omitted values since only three experimental values are available, resulting in an exact solution to Equation (5.9). THF cosolvent mixture results are omitted for the same reason.

The results in Table 5.3.3 suggest that the two-descriptor correlation model can accurately reproduce reaction rates as a function of cosolvent concentration for a single reactant in a single solvent mixture. To probe if Γ and τ can be used to predict reaction performance for a series of various reactants, we calculated best-fit parameters for Equation (5.9) using the combined data for all seven reactants in DIO/water mixtures and used the resulting two-descriptor correlation model to calculate values of σ. Comparing the calculated and experimental values of σ resulted in a slope of 0.71 and RMSE of 0.36. We further found that incorporating a reactant-specific descriptor, the accessible hydroxyl fraction (δ), led to the improved multidescriptor correlation model described in Equation (5.10).

$$\sigma_{\text{pred}} = A + B(\Gamma) + C(\tau) + D(\delta) \tag{5.10}$$

The three molecular descriptors are physically motivated *vis-à-vis* the generalized reaction mechanism proposed above. The preferential exclusion coefficient captures the extent to which the reactant accumulates excess water in its immediate vicinity, which creates a thermodynamic driving force for the transfer of a proton from the bulk phase to the reactant, thereby initiating the acid-catalyzed reaction mechanism. The hydrogen bonding lifetime ratio captures information regarding the binding strength of water to the reactant, which we interpret as a measure of the ability of the encapsulated water cluster to stabilize the carbocation-like transition states common to acid-catalyzed reactions, as shown in Figure 5.3.4.. The accessible hydroxyl fraction captures information relating to the percentage of the surface area of the reactant molecule occupied by

hydroxyl groups, which we interpret as a qualitative descriptor of the reactant's hydrogen bonding capacity, normalized by its molecular size.

Figure 5.3.8(a) compares kinetic solvent parameters calculated using the multidescriptor correlation model in Equation (5.10) to experimentally determined values for DIO/water mixtures, affording a slope of 0.88 and RMSE of 0.24 with few false positive/negatives. As the correlation is maintained across several different reactants, we suggest that acid-catalyzed reactions behave similarly in each of these cosolvent mixtures, where knowledge of simulation-derived Γ, τ, and δ can predict experimental σ_{exp} with accuracy. To demonstrate the predictive power of the model in Equation (), we selected one reactant (the *test set*), fit the parameters in Equation 5.10 using the remaining six reactants (the *training set*), and then calculated kinetic solvent parameters for the test set reactant using the parameters derived from the training set. This procedure assesses the ability of the multidescriptor correlation model to predict kinetic solvent parameters for reactants that are not used to determine model parameters. Figure 5.3.8(b) shows the results of this procedure for DIO/water mixtures using FRU as the test set reactant. We find that the best-fit slope and RMSE between experimental and predicted kinetic solvent parameters are nearly the same when using the six-reactant training set compared to using all seven reactants. The test set RMSE is 0.13, which is lower than the average RMSE of the training set. These results confirm the robustness and predictability of the multidescriptor correlation model.

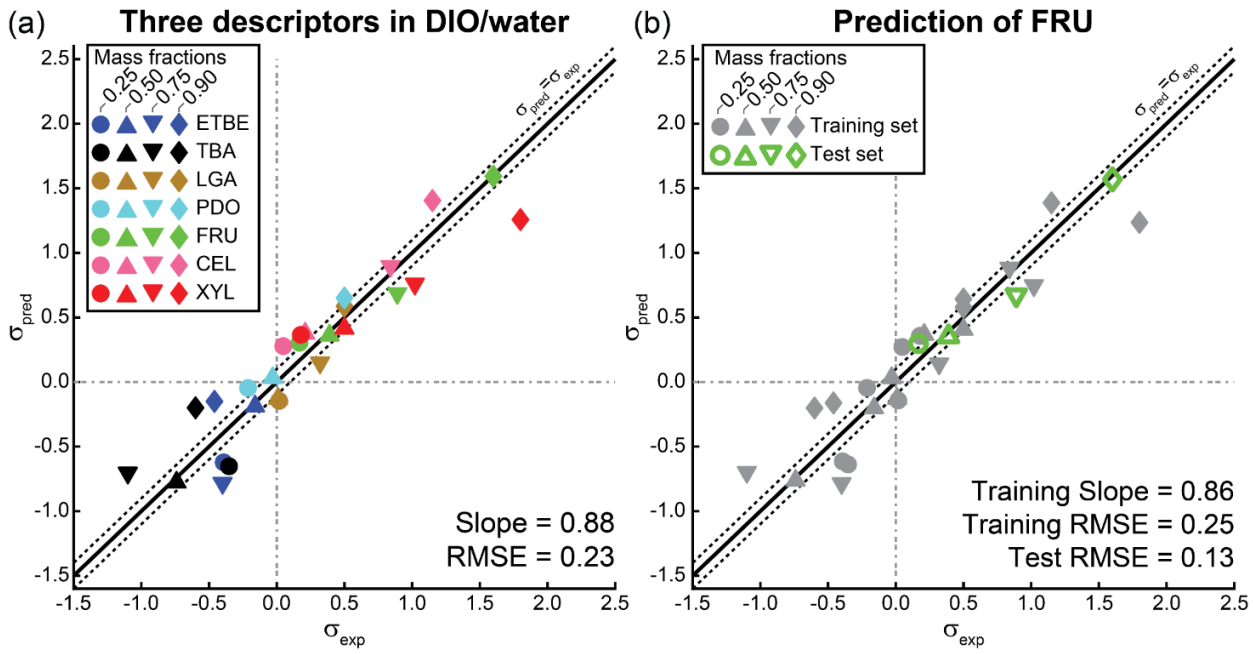


Figure 5.3.8. (a) Comparison of kinetic solvent parameters calculated using the multidescriptor correlation model (σ_{pred}) to experimentally determined values (σ_{exp}) for all seven reactants in DIO/water mixtures. Each reactant has four data points for 0.25, 0.50, 0.75, and 0.90 mass fractions of DIO, with the exception of PDO. The slope of the best-fit line for all the data points and the average root-mean-squared error (RMSE) between the values of σ_{pred} and σ_{exp} are shown at the bottom right. The solid black line indicates a perfect correlation ($\sigma_{pred} = \sigma_{exp}$) and dotted lines are drawn at $\sigma_{exp} = 0$ and $\sigma_{pred} = 0$ to help visualize false positive/negative predicted values. Lines above and below the $\sigma_{pred} = \sigma_{exp}$ line are shifted by ±0.10, denoting the approximate experimental error. (b) Prediction of kinetic solvent parameters using FRU as a test set with all other reactants taken as a training set. The slope of the best-fit line and RMSE between the values of σ_{pred} and σ_{exp} for the training and test sets are shown at the bottom right.

Table 5.3.4 summarizes the model coefficients and error for the different cosolvent environments. Simulated parameters were re-scaled to values between 0 and 1 so that the values of the coefficients in Equation 5.10) can be compared. DIO/water mixtures were found to have the best correlation when fitting. GVL/water and THF/water cosolvent environments have the highest error. The coefficients differ across cosolvents, suggesting that the simulated parameters contribute to a different degree when changing cosolvent environments. In general, Γ has the largest weight in the multidescriptor correlation model for DIO/water and GVL/water systems, indicating that the formation of a water-rich local domain is the dominant contributor to reactivity, in agreement with the correlations in Table 5.3.2. Γ has a smaller weight in THF/water system further supported by low correlations between Γ and σ. The low correlation implies that there may be other descriptors that could be included to improve the model, such as additional cosolvent-specific descriptors (*e.g.*, dielectric constants, size, dipole moment, etc.).(*54*)

Table 5.3.4. Coefficients of the multidescriptor correlation model describing the rates of all seven reactions in each of the three cosolvent mixtures with the best-fit slope and root-mean-square error (RMSE) between predicted (σ_{pred}) and experimentally determined (σ_{exp}) kinetic solvent parameters. The descriptors $\tilde{\Gamma}$, $\tilde{\tau}$, and $\tilde{\delta}$ are equal to Γ, τ, and δ but re-scaled to values between 0 and 1 so the coefficients are comparable.

$$\sigma_{\text{pred}} = A + B(\tilde{\Gamma}) + C(\tilde{\tau}) + D(\tilde{\delta})$$

Cosolvent	A	B	C	D	Slope	RMSE
DIO	-1.497	1.489	1.270	1.025	0.88	0.23
GVL	-1.389	1.542	0.851	1.077	0.70	0.37
THF	-0.825	0.501	1.510	0.302	0.53	0.58

5.4. Conclusions

We have analyzed the effects of three polar aprotic cosolvents, in aqueous mixtures of varying composition, on the acid-catalyzed reactions of seven biomass-derived model compounds. General trends in reactivity, as expressed by changes in apparent rate constants as a function of solvent composition, were correlated to simulation-derived observables obtained from classical molecular dynamics simulations. We find that the presence of organic cosolvents in the solvent mixture leads to the formation of water-enriched local solvent domains near hydrophilic reactants and increases the strength of hydrogen bonding between reactants and local water molecules. These effects are quantified by molecular descriptors describing: (1) the local density of water near the reactant (Γ); (2) the average hydrogen-bond lifetime between the reactant and neighboring water molecules (τ); and (3) the accessible surface area occupied by hydroxyl groups on the reactant (δ). By combining these three observables in a multiple linear regression scheme, we have

developed a multidescriptor correlation model that predicts rate constants as a function of solvent composition. This development represents an important step toward the computational design of new liquid-phase biomass conversion processes, informed by a first principles approach.

5.5. Acknowledgments

This work was supported in part by: the Department of Energy, Office of Energy Efficiency and Renewable Energy (EERE), under Award Number DE-EE0006878; by the U.S. Department of Energy, Office of Basic Energy Sciences; and by the DOE Great Lakes Bioenergy Research Center (http://www.glbrc.org). The U.S. Department of Energy, Office of Sciences, Office of Biological and Environmental Research support the GLBRC, through the cooperative agreement BER DE-FC02-07ER64494 between the Board of Regents of the University of Wisconsin and the U.S. Department of Energy. T.W. Walker acknowledges that this material is based on work supported by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin–Madison with funding from the Wisconsin Alumni Research Foundation. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. This work also used the computing resources and assistance of the UW-Madison Center for High Throughput Computing (CHTC) in the Department of Computer Sciences. The CHTC is supported by UW-Madison, the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation, the Wisconsin Institutes for Discovery, and the National Science Foundation, and is an active member of the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science. We acknowledge Liam Witteman for help in carrying

out the reaction kinetics studies. We thank Dr. Max A. Mellmer for thoughtful discussions regarding organic solvent effects in acid-catalyzed reactions.

5.6. References

1. J. N. Chheda, G. W. Huber, J. A. Dumesic, Liquid-phase catalytic processing of biomass-derived oxygenated hydrocarbons to fuels and chemicals. *Angewandte Chemie International Edition* **46**, 7164-7183 (2007).
2. A. Corma, S. Iborra, A. Velty, Chemical Routes for the Transformation of Biomass into Chemicals. *Chemical Reviews* **107**, 2411-2502 (2007).
3. Y. Román-Leshkov, C. J. Barrett, Z. Y. Liu, J. A. Dumesic, Production of dimethylfuran for liquid fuels from biomass-derived carbohydrates. *Nature* **447**, 982-985 (2007).
4. M. Stöcker, Biofuels and Biomass-To-Liquid Fuels in the Biorefinery: Catalytic Conversion of Lignocellulosic Biomass using Porous Materials. *Angewandte Chemie International Edition* **47**, 9200-9211 (2008).
5. M. A. Mellmer, D. M. Alonso, J. S. Luterbacher, J. M. R. Gallo, J. A. Dumesic, Effects of γ-valerolactone in hydrolysis of lignocellulosic biomass to monosaccharides. *Green Chemistry* **16**, 4659-4662 (2014).
6. M. A. Mellmer *et al.*, Solvent effects in acid-catalyzed biomass conversion reactions. *Angewandte chemie international edition* **53**, 11872-11875 (2014).
7. Y. Román-Leshkov, J. N. Chheda, J. A. Dumesic, Phase modifiers promote efficient production of hydroxymethylfurfural from fructose. *Science* **312**, 1933-1937 (2006).
8. Z. Wei, Y. Li, D. Thushara, Y. Liu, Q. Ren, Novel dehydration of carbohydrates to 5-hydroxymethylfurfural catalyzed by Ir and Au chlorides in ionic liquids. *Journal of the Taiwan Institute of Chemical Engineers* **42**, 363-370 (2011).
9. M. A. Mellmer, J. M. R. Gallo, D. Martin Alonso, J. A. Dumesic, Selective production of levulinic acid from furfuryl alcohol in THF solvent systems over H-ZSM-5. *ACS Catalysis* **5**, 3354-3359 (2015).
10. Y. J. Pagan-Torres, T. Wang, J. M. R. Gallo, B. H. Shanks, J. A. Dumesic, Production of 5-hydroxymethylfurfural from glucose using a combination of Lewis and Brønsted acid catalysts in water in a biphasic reactor with an alkylphenol solvent. *Acs Catalysis* **2**, 930-934 (2012).
11. L. Shuai, J. Luterbacher, Organic solvent effects in biomass conversion reactions. *ChemSusChem* **9**, 133-155 (2016).
12. E. Cesarotti, R. Ugo, L. Kaplan, A discussion of the different kinds of solute-solute and solute-solvent interactions acting in homogeneous catalysis by transition metal complexes. *Coordination Chemistry Reviews* **43**, 275-298 (1982).
13. R. J. Madon, E. Iglesia, Catalytic reaction rates in thermodynamically non-ideal systems. *Journal of Molecular Catalysis A: Chemical* **163**, 189-204 (2000).
14. S. Behtash, J. Lu, E. Walker, O. Mamun, A. Heyden, Solvent effects in the liquid phase hydrodeoxygenation of methyl propionate over a Pd (1 1 1) catalyst model. *Journal of catalysis* **333**, 171-183 (2016).

15. S. Caratzoulas, D. G. Vlachos, Converting fructose to 5-hydroxymethylfurfural: a quantum mechanics/molecular mechanics study of the mechanism and energetics. *Carbohydrate Research* **346**, 664-672 (2011).
16. R. S. Assary, T. Kim, J. J. Low, J. Greeley, L. A. Curtiss, Glucose and fructose to platform chemicals: understanding the thermodynamic landscapes of acid-catalysed reactions using high-level ab initio methods. *Physical Chemistry Chemical Physics* **14**, 16603-16611 (2012).
17. V. Choudhary *et al.*, Insights into the interplay of Lewis and Brønsted acid catalysts in glucose and fructose conversion to 5-(hydroxymethyl) furfural and levulinic acid in aqueous media. *Journal of the American Chemical Society* **135**, 3997-4006 (2013).
18. A. Curioni *et al.*, Density Functional Theory-Based Molecular Dynamics Simulation of Acid-Catalyzed Chemical Reactions in Liquid Trioxane. *Journal of the American Chemical Society* **119**, 7218-7229 (1997).
19. V. Vasudevan, S. H. Mushrif, Insights into the solvation of glucose in water, dimethyl sulfoxide (DMSO), tetrahydrofuran (THF) and N,N-dimethylformamide (DMF) and its possible implications on the conversion of glucose to platform chemicals. *RSC Advances* **5**, 20756-20763 (2015).
20. B. Mostofian *et al.*, Local phase separation of co-solvents enhances pretreatment of biomass for bioenergy applications. *Journal of the American Chemical Society* **138**, 10869-10878 (2016).
21. S. H. Mushrif, S. Caratzoulas, D. G. Vlachos, Understanding solvent effects in the selective conversion of fructose to 5-hydroxymethyl-furfural: a molecular dynamics investigation. *Physical Chemistry Chemical Physics* **14**, 2637-2644 (2012).
22. O. K. Abou-Zied, O. I. K. Al-Shihi, *Steady-state and time-resolved fluorescence investigation of 2-pyridone and 3-pyridone in solution and their specific binding to human serum albumin*. SPIE BiOS (SPIE, 2008), vol. 6867.
23. M. A. Mellmer *et al.*, Solvent-enabled control of reactivity for liquid-phase reactions of biomass-derived compounds. *Nature Catalysis* **1**, 199-207 (2018).
24. E. Raamat *et al.*, Acidities of strong neutral Brønsted acids in different media. *Journal of Physical Organic Chemistry* **26**, 162-170 (2013).
25. A. Trummal, L. Lipping, I. Kaljurand, I. A. Koppel, I. Leito, Acidity of Strong Acids in Water and Dimethyl Sulfoxide. *The Journal of Physical Chemistry A* **120**, 3663-3669 (2016).
26. S. Páll, M. J. Abraham, C. Kutzner, B. Hess, E. Lindahl. (Springer International Publishing, Cham, 2015), pp. 3-27.
27. R. B. Best *et al.*, Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain $\chi 1$ and $\chi 2$ Dihedral Angles. *Journal of Chemical Theory and Computation* **8**, 3257-3273 (2012).
28. K. Vanommeslaeghe *et al.*, CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* **31**, 671-690 (2010).
29. W. Yu, X. He, K. Vanommeslaeghe, A. D. MacKerell Jr., Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *Journal of Computational Chemistry* **33**, 2451-2468 (2012).
30. H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma, The missing term in effective pair potentials. *The Journal of Physical Chemistry* **91**, 6269-6271 (1987).

31. Robert T. McGibbon *et al.*, MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **109**, 1528-1532 (2015).
32. A. Luzar, D. Chandler, Hydrogen-bond kinetics in liquid water. *Nature* **379**, 55-57 (1996).
33. H. J. C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **91**, 43-56 (1995).
34. B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **4**, 435-447 (2008).
35. J. Zhang, A. Das, R. S. Assary, L. A. Curtiss, E. Weitz, A combined experimental and computational study of the mechanism of fructose dehydration to 5-hydroxymethylfurfural in dimethylsulfoxide using Amberlyst 70, PO43−/niobic acid, or sulfuric acid catalysts. *Applied Catalysis B: Environmental* **181**, 874-887 (2016).
36. S. H. Mushrif, V. Vasudevan, C. B. Krishnamurthy, B. Venkatesh, Multiscale molecular modeling can be an effective tool to aid the development of biomass conversion technology: A perspective. *Chemical Engineering Science* **121**, 217-235 (2015).
37. T. D. Swift *et al.*, Kinetics of Homogeneous Brønsted Acid Catalyzed Fructose Dehydration and 5-Hydroxymethyl Furfural Rehydration: A Combined Experimental and Computational Study. *ACS Catalysis* **4**, 259-267 (2014).
38. G. R. Akien, L. Qi, I. T. Horváth, Molecular mapping of the acid catalysed dehydration of fructose. *Chemical Communications* **48**, 5850-5852 (2012).
39. A. J. Bennet, M. L. Sinnott, Complete kinetic isotope effect description of transition states for acid-catalyzed hydrolyses of methyl .alpha.- and .beta.-glucopyranosides. *Journal of the American Chemical Society* **108**, 7287-7294 (1986).
40. J. A. Dumesic, Analyses of Reaction Schemes Using De Donder Relations. *Journal of Catalysis* **185**, 496-505 (1999).
41. J. K. Lee, A. D. Bain, P. J. Berti, Probing the Transition States of Four Glucoside Hydrolyses with 13C Kinetic Isotope Effects Measured at Natural Abundance by NMR Spectroscopy. *Journal of the American Chemical Society* **126**, 3769-3776 (2004).
42. C. P. Kelly, C. J. Cramer, D. G. Truhlar, Single-Ion Solvation Free Energies and the Normal Hydrogen Electrode Potential in Methanol, Acetonitrile, and Dimethyl Sulfoxide. *The Journal of Physical Chemistry B* **111**, 408-422 (2007).
43. M. D. Tissandier *et al.*, The Proton's Absolute Aqueous Enthalpy and Gibbs Free Energy of Solvation from Cluster-Ion Solvation Data. *The Journal of Physical Chemistry A* **102**, 7787-7794 (1998).
44. C. Kalidas, G. Hefter, Y. Marcus, Gibbs Energies of Transfer of Cations from Water to Mixed Aqueous Organic Solvents. *Chemical Reviews* **100**, 819-852 (2000).
45. A. Warshel, Energetics of enzyme catalysis. *Proceedings of the National Academy of Sciences* **75**, 5250-5254 (1978).
46. S. Eckstein *et al.*, Influence of Hydronium Ions in Zeolites on Sorption. *Angewandte Chemie International Edition* **58**, 3450-3455 (2019).
47. M. Kang, P. E. Smith, Preferential interaction parameters in biological systems by Kirkwood–Buff theory and computer simulation. *Fluid Phase Equilibria* **256**, 14-19 (2007).

48. C. P. Schneider, B. L. Trout, Investigation of Cosolute−Protein Preferential Interaction Coefficients: New Insight into the Mechanism by Which Arginine Inhibits Aggregation. *The Journal of Physical Chemistry B* **113**, 2050-2058 (2009).
49. D. Shukla, B. L. Trout, Preferential Interaction Coefficients of Proteins in Aqueous Arginine Solutions and Their Molecular Origins. *The Journal of Physical Chemistry B* **115**, 1243-1253 (2011).
50. I. L. Shulgin, E. Ruckenstein, Local Composition in the Vicinity of a Protein Molecule in an Aqueous Mixed Solvent. *The Journal of Physical Chemistry B* **111**, 3990-3998 (2007).
51. N. Smolin, R. Winter, Effect of Temperature, Pressure, and Cosolvents on Structural and Dynamic Properties of the Hydration Shell of SNase: A Molecular Dynamics Computer Simulation Study. *The Journal of Physical Chemistry B* **112**, 997-1006 (2008).
52. J. E. Sutton, D. G. Vlachos, A Theoretical and Computational Analysis of Linear Free Energy Relations for the Estimation of Activation Energies. *ACS Catalysis* **2**, 1624-1634 (2012).
53. S. Wang *et al.*, Universal transition state scaling relations for (de)hydrogenation over transition metals. *Physical Chemistry Chemical Physics* **13**, 20760-20765 (2011).
54. M. J. Kamlet, J. L. M. Abboud, M. H. Abraham, R. W. Taft, Linear solvation energy relationships. 23. A comprehensive collection of the solvatochromic parameters, .pi.*, .alpha., and .beta., and some methods for simplifying the generalized solvatochromic equation. *The Journal of Organic Chemistry* **48**, 2877-2887 (1983).

Chapter 6. Effects of Mixed Solvent Environments on the Selectivity of Acid-Catalyzed Dehydration Reactions.[4]

6.1. Introduction

In the previous chapter, we utilized classical molecular dynamics (MD) simulations to estimate the rates of acid-catalyzed reactions of biomass-analogous model oxygenates in mixtures of water and polar aprotic cosolvents,(*1*) based on insights obtained from prior experiments and *ab initio* simulations.(*2, 3*) These simulation-derived estimates of reactivity translate into predictions of selectivity for series reactions in which an intermediate rather than a terminal product is desired. For example, the acid-catalyzed dehydration of fructose to afford 5-hydroxymethylfurfural (HMF), a platform chemical for fuel and other commodity chemicals production,(*4, 5*) can be difficult to control due to the subsequent hydrolysis of HMF to form levulinic and formic acids.(*6*) Inclusion of dimethyl sulfoxide (DMSO) has been shown to increase HMF selectivity, in part by preventing the subsequent, undesirable hydrolysis step.(*7*) However, the reaction networks underlying liquid-phase processes are often comprised of reactions occurring both in series and in parallel. Therefore, it is desirable to understand how solvent composition affects not just the rates of acid-catalyzed reactions, but also the selectivity of reactions occurring in parallel.

The mechanisms by which solvent molecules influence acid-catalyzed reaction selectivities can be broadly understood in terms of solvent effects that we divide into three

[4] This chapter was adapted from: A. K. Chew, T. W. Walker, J. Euclid, L. Wittman, G. W. Huber, J. A. Dumesic and R. C. Van :Lehn, "Effects of Mixed Solvent Environments on the selectivity of Acid-Catalyzed Dehydration Reactions," ACS Catalysis (2019), *in press.* A. K. Chew and R. C. Van :Lehn conceived the work. T. W. Walker and L. Witteman designed and carried out the reaction kinetics experiments. A. K Chew Designed and carried out the molecular dynamics simulations. J. Euclid carried out the ALAMO calculations. T. W. walker and A. K. Chew wrote the manuscript. All authors edited the manuscript.

categories (*8-10*): equilibrium solvation, (*3, 11-14*) cosolvent participation, (*15, 16*) and solvent dynamics. (*3, 14, 15, 17*) Equilibrium solvation refers to the effect of the solvent environment on the stabilities (*i.e.*, relative free energies) of the reactant, transition, and product states.(*12, 18*) For example, polar aprotic cosolvents can lower the activation energy barriers for certain biomass conversion processes by stabilizing key transition states, resulting in enhanced reaction rates compared to the same reactions in pure water.(*9, 13*) Cosolvent participation refers to the ability of cosolvent molecules to direct reaction selectivity by participating in reaction steps or sterically "shielding" specific reactive sites around the reactant.(*7*) For example, classical MD simulations suggest that the competition between DMSO and water molecules for hydroxyl groups on fructose promotes conversion to HMF by protecting fructose from undesirable parallel reactions and shielding HMF from subsequent rehydration.(*7*) Solvent dynamics refers to the effect of solvent relaxation on the stabilization of charged transition states (ubiquitous in acid-catalyzed reaction mechanisms) that polarize nearby solvent molecules. For example, the slow orientational relaxation of methanol compared to water increases the activation energy barrier for the isomerization of glucose to fructose by leading to poor solvation of the transition state.(*15*) These examples illustrate the possible processes by which the solvent environment can influence reaction selectivity but have been determined for a limited set of reactant-solvent combinations. Developing an understanding of solvent effects that can apply across a range of mixed-solvent environments remains a challenge.

Herein, we investigate the hypothesis that solvent composition can be systematically varied to influence the selectivity of model acid-catalyzed reactions to afford well-characterized products, and that characterizing the stability and local solvent environments of reactants and products using classical MD simulations can be used anticipate these effects. We demonstrate that the selectivity

of Brønsted acid-catalyzed 1,2-propanediol (PDO) dehydration can be modulated by tuning the composition of six different cosolvent-water mixtures with various mass fractions of cosolvent, affording low-to-moderate yields of propanal (PRO) or acetone (ACE). We find that ACE is formed as the major product in aqueous solutions of DMSO, whereas all other solvent environments prefer the formation of PRO over ACE with the rate of PRO formation depending on the specific cosolvent. Furthermore, we find that MD-derived reactant and product solvation free energies, which quantify the effect of equilibrium solvation on selectivity, can explain changes to the rate of PRO formation. We then generalize to reaction rate trends for representative polyols in 90 wt% GVL- and DMSO-water mixtures. Finally, analysis of the unique spatial distribution of solvent molecules around PDO in DMSO-water mixtures suggests that DMSO competes with water for reactive sites which may lead to changes in the reaction mechanism to favor the production of ACE. While *ab initio* methods are necessary to confirm the reaction mechanisms underlying observed trends in selectivity, these results represent a step toward anticipating how solvent composition affects the selectivity of reactions occurring in parallel using computationally efficient simulation methods.

6.2. Methods

6.2.1. Reaction kinetics studies

Reactions were carried out in closed, 10 mL thick-walled glass reactors. In a typical experiment, an appropriate amount of reactant (*i.e.* PDO), acid catalyst, water and organic cosolvent (*e.g.* DMSO) were charged to the reactors, which were then sealed and placed in an oil bath at 433 K. Reactors were removed at the desired reaction time and quenched in an ice bath at 273 K. The conversion of the reactant was monitored as a function of reaction time using a Shimadzu high-performance liquid chromatograph equipped with a differential refractometer.

Products were quantified using a Shimadzu gas chromatograph equipped with a flame ionization detector. All products were quantified using calibration curves with external standards. Rate constants for the conversion of reactants and the formation products were derived from Equation 1 using the method of initial rates. Trifluoromethane sulfonic (triflic) acid ($pK_{a,\ H2O}$ = -14.7, $pK_{a,\ DMSO}$ = -14.3) was used as catalyst in all experiments. Triflic acid behaves as a strong acid even in pure polar aprotic solvents.(*19, 20*) We thus assume complete dissociation of the acidic proton in all mixed-solvent environments, allowing for normalization of the apparent rate constants on a per-proton basis.

The sum of reaction selectivities is typically less than 100% in all solvent systems displayed tested in this report. This is due to the formation of a range of condensation products, such as 2-ethyl-4-methyl-1,3-dioxolane, 2-methoxy-1,3-dioxolane, 4-methyl-1,3-dioxolane and 1,3-dioxolane. In our prior work, we have also detected the presence of multiple degradation products such as 3-hydroxyl-2-methyl-pentanal, 2-ethyl-2-butenal, heptanal and 2-(1-methylethoxy)-1-propanol using gas chromatography-mass spectrometry analysis.(*3*) However, due to difficulties in isolating and quantifying these species, their formation rates were not investigated in this study.

6.2.2. Molecular dynamics simulations

All classical molecular dynamics simulations were performed using Gromacs version 2016.(*21*) Reactant, products, and cosolvents were parameterized using the CGenFF/CHARMM36 force fields.(*22-24*) Water was modeled using the Single Point Charged/Extended (SPC/E) model.(*25*) We initialized simulations of mixed-solvent environments following the protocol discussed in our previous work.(*1*) The initial simulation box dimensions were set to $(6\ nm)^3$ in all simulations and water and cosolvent molecules were added at the desired composition. The system

was equilibrated for 5 ns at 300 K using a velocity-rescale thermostat and 1 bar using a Berendsen barostat.in a *NPT* simulation for 5 ns.

Solvation free energies were computed from a series of stochastic dynamics simulations, as described in Ref. (*26*) Starting from an equilibrated solvent system (as described above), a reactant or product is added to the system. The total potential of the system was defined as a function of Lennard-Jones (λ_{LJ}) and electrostatic (λ_{elec}) potential parameters, shown in Equation 6.1:

$$U(\lambda_{LJ}, \lambda_{elec}) = U_{M,solv}^{LJ}(\lambda_{LJ}) + U_{M,solv}^{elec}(\lambda_{elec}) + U_{M}^{bonded} + U_{M}^{nonbonded} + U_{solv}^{bonded} + U_{solv}^{nonbonded} \quad (6.1)$$

where $U_{M,solv}^{LJ}$ and $U_{M,solv}^{elec}$ are the LJ and electrostatic potentials between solute and solvent, U_{M}^{bonded} and $U_{M}^{nonbonded}$ are intramolecular bonded and non-bonded potentials of the solute, and U_{solv}^{bonded} and $U_{solv}^{nonbonded}$ are the bonded and non-bonded potentials between all solvent molecules.(*27*) λ_{LJ} and λ_{elec} vary between 0 and 1 and modify the magnitude of the associated interactions; values of 0 correspond to a species in the vapor phase with no interactions with the solvent environment, while values of 1 correspond to a species in the liquid phase with complete interactions with the solvent environment. Seventeen independent simulations were performed for each solvation free energy: fourteen in which $\lambda_{elec} = 0.00$ and $\lambda_{LJ} = 0.00$, 0.00922, 0.04794, 0.11505, 0.260634, 0.31608, 0.43738, 0.56262, 0.68392, 0.79366, 0.88495, 0.95206, 0.99078, or 1.00, and three in which $\lambda_{LJ} = 1.00$ and $\lambda_{elec} = 0.25$, 0.75, or 1.00. These values represent a 12-point Gaussian sequence which was used previously to parameterize ion models [63, 64].(*28, 29*) For each simulation, the system was energy minimized with the steepest descent algorithm and equilibrated for 100 ps at constant temperature followed by 2 ns at constant temperature and constant pressure using the Berendsen barostat. An 11-ns production simulation at constant

temperature and pressure was then performed with the Parrinello-Rahman barostat. All simulations were performed at 433.15 K and 1 bar. Energy differences computed between all pairs of windows were collected every 0.2 ps and solvation free energies were computed with the Multistate Bennett Acceptance Ratio(*30*) method using the python alchemical-analysis tool.(*31*) The last 10 ns of each production simulation were split into two 5 ns trajectories and treated as two independent trials. All solvation free energy results and error bars are reported as the average and standard deviation of the two trials, respectively.

Spatial distribution maps were generated using simulations with the reactant restrained and the solvent free to explore reactive sites. A single reactant was added to an equilibrated solvent system and then the system was equilibrated again for 500 ps at 433.15 K using a velocity-rescale thermostat and 1 bar using a Berendsen barostat. The system was simulated for 300 ns for PDO and 50 ns for *cis*- and *trans*-isomers of 1,2-cyclohexanediol at the same temperature and pressure, controlled by the Parrinello-Rahman barostat and Nose-Hoover thermostat, respectively. We extracted the most likely configuration of the reactant from unbiased simulations and restrained the reactant's atomic positions with a force constant of 10,000 kJ/(mol-nm) to eliminate rotational degrees of freedom. We then performed an additional 200 ns simulation with the restrained reactant in solution, where the final 190 ns of the trajectory was used to calculate spatial distribution maps. Simulation analysis was performed using the *MDTraj* package(*32*) and analysis tools developed in-house.

6.3. Results

6.3.1. Proposed reaction mechanism for the acid-catalyzed dehydration of PDO

Figure 6.3.1A depicts a possible mechanism for the Brønsted acid-catalyzed dehydration of PDO in the gas phase over a solid acid catalyst.(*33*) Depending upon which hydroxyl moiety is removed in the form of water, this reaction affords either PRO or ACE as the major product, passing through a secondary- or primary-carbocation-like intermediate, respectively. If subsequent conversion of the products can be neglected in the acid-catalyzed dehydration of PDO, then the selectivity of this reaction depends uniquely upon the relative rates at which corresponding products are formed; or, equivalently, upon the relative energy differences between the reactant, transition, and product states. Accordingly, PRO is the major observable product in the gas phase because a secondary carbocation is more stable than a primary carbocation.(*34*) However, liquid solvents can alter the reaction mechanism by stabilizing structures that may otherwise be unstable in the gas phase.(*9, 11*) Therefore, liquid-phase reactions may proceed through a different mechanism, such as a reaction mechanism that is more concerted rather than stepwise, or a reaction mechanism with intermediates that are chemically distinct from those shown in Figure 6.3.1A. Following these insights, we now explore whether the composition of mixed-solvent environments can be tuned to modulate the relative rates of PDO dehydration to selectively afford either PRO or ACE. To this end, we examine the effects of mixed-solvent environments composed of water and one of six polar aprotic cosolvents; their chemical structures are shown in Figure 6.3.1B.

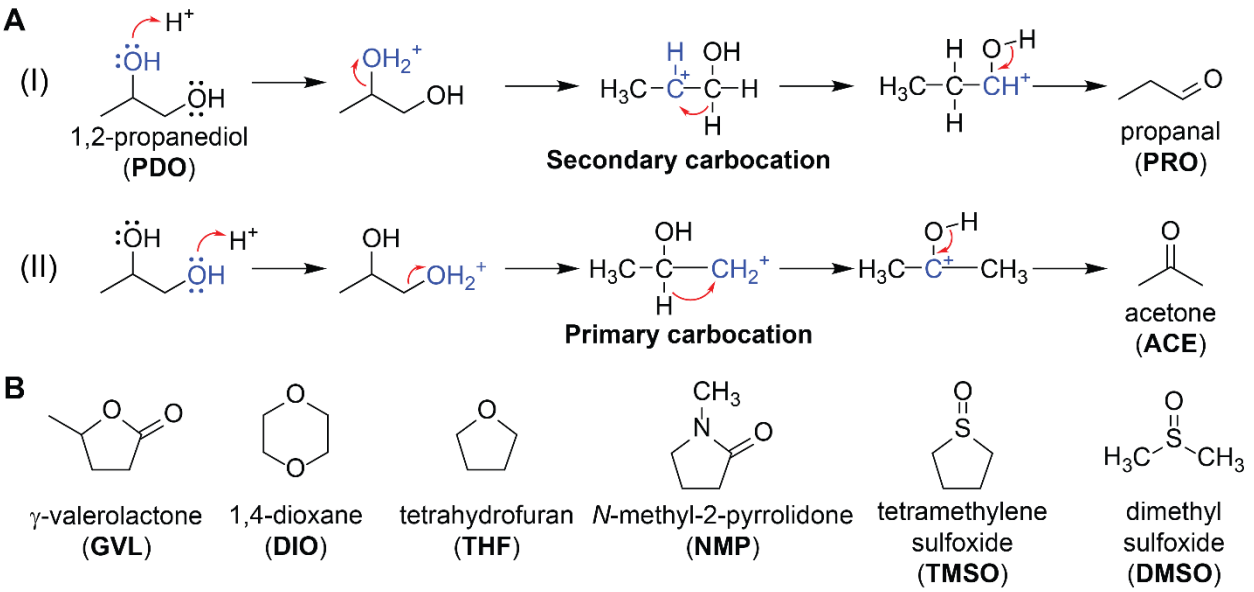


Figure 6.3.1. A Proposed mechanism for Brønsted acid-catalyzed dehydration of 1,2-propanediol (PDO) to afford either propanal (PRO) or acetone (ACE) in the gas phase over a solid acid catalyst.(*34*) **B** Polar aprotic cosolvents used for this study.

6.3.2. Product selectivities for the Brønsted acid-catalyzed dehydration of 1,2-propanediol

We carried out PDO dehydration in batch reactors over a triflic acid catalyst and monitored reactant conversion as a function of time. The kinetics of PDO dehydration have previously been shown to be of approximately first order with respect to the concentration of the reactant and acidic protons in solution.(*3*) Accordingly, Equation 6.2 relates the initial rate of reactant conversion (r_i) to the molar concentration of reactant i (C_i), the molar concentration of acidic protons (C_{H^+}) in solution, and the apparent first-order rate constant (k_{app}) at a fixed solvent composition and temperature:

$$r_i = -\frac{dC_i}{dt} = k_{app} C_i C_{H^+} \tag{6.2}$$

Simultaneously, the formation of product p (*i.e.*, PRO or ACE) was monitored as a function of time, and the selectivity to each product (S_p) was quantified as the ratio of the initial rate of product formation (r_p) to the initial rate of reactant conversion (r_i) or, equivalently, as a ratio of rate constants (Equations 6.3-6.4):

$$r_p = \frac{dC_p}{dt} = k_p C_i C_{H^+} \tag{6.3}$$

$$S_p = \frac{r_p}{r_i} = \frac{k_p}{k_{app}} \tag{6.4}$$

The rate constant for the formation of each product is therefore obtained by multiplying k_{app} by the selectivity to either ACE or PRO.

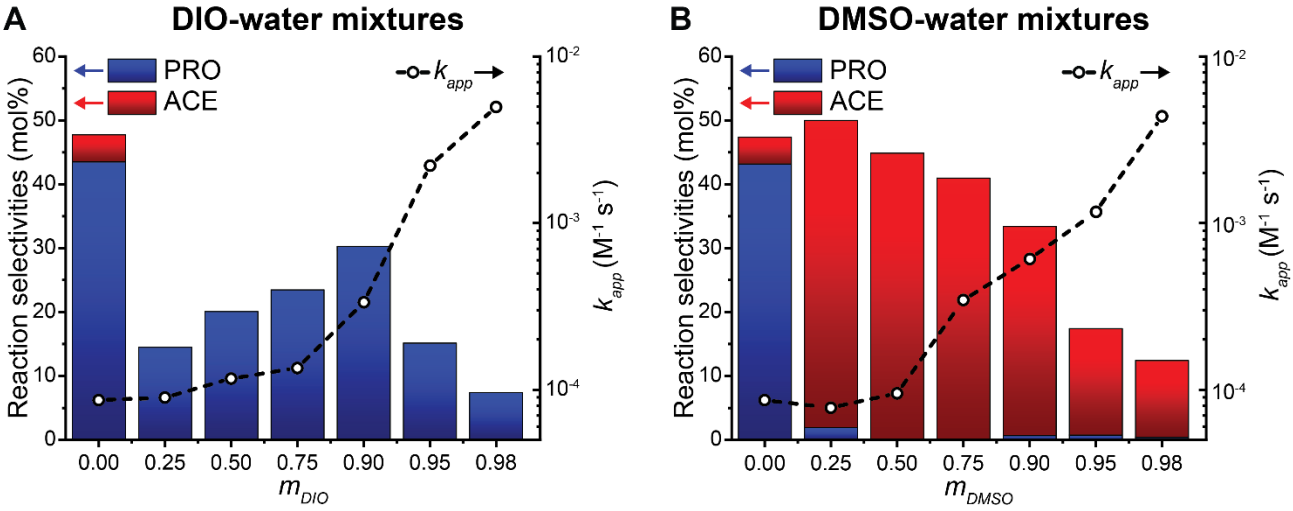


Figure 6.3.2. Apparent rate constants (k_{app}, dashed lines) and selectivities to propanal (PRO, blue columns) and acetone (ACE, red columns) for Brønsted-acid-catalyzed 1,2-propanediol (PDO) dehydration in mixtures of water with (**A**) 1,4-dioxane (DIO) and (**B**) dimethyl sulfoxide (DMSO) as a function of the mass fraction of the organic component (m_{DIO} or m_{DMSO}). Rate constants are derived from Equation 1 and selectivities are derived from Equation 3. The standard error in selectivities is +/- 5 mol%. Reaction conditions: ~20 mol% PDO conversion; 433 K; 0.4-0.005 M

triflic acid; 0.01 M PDO; 90-150 min reaction time; 500 rpm stirring rate, 2 mL total solvent volume.

Figure 6.3.2 shows the selectivities and apparent rate constants for PDO dehydration in water and in aqueous mixtures with DIO or DMSO. The reaction selectivities do not sum to one due to the formation of condensation products that are difficult to quantify (see Methods); thus, we focus on the effect of the solvent composition on the formation of PRO and ACE. In pure water, PRO is formed with 41% selectivity on a molar basis and ACE is formed with 9% selectivity. In both solvent systems, the value of the apparent rate constant for PDO conversion increases monotonically with the mass fraction of the organic phase, an effect that was explored in prior work.(*1, 3*) At 25 wt% DIO, the selectivity to ACE drops to zero and the selectivity to PRO decreases to 15% (Figure 6.3.2A). Upon further addition of DIO, the selectivity to PRO partially recovers, passing through a maximum of 30% at 90 wt% DIO. The selectivity to PRO trends towards zero for mass fractions of DIO above 90 wt%, indicating that water might play a role in the reaction mechanism that affords this product. This observation is consistent with the prior studies indicating that protic solvents such as water can facilitate acid-catalyzed reaction mechanisms by stabilizing carbocation-like intermediates.(*3, 15*) In contrast, almost no PRO is formed at 25 wt% DMSO, whereas the selectivity to ACE increases to 49% (Figure 6.3.2B). The further addition of DMSO results in a monotonic decrease in the selectivity to ACE, again indicating that this reaction is at least partially facilitated by water.

To understand the trends in selectivity, we now probe the role of DMSO-water mixtures in altering the relative rates of PRO and ACE formation. Figure 6.3.3 shows the apparent rate constants for PDO consumption, PRO formation, and ACE formation as a function of solvent composition in DMSO-water mixtures. As the mass fraction of DMSO is increased from 7 to 25

wt%, the overall rate of PDO conversion remains roughly constant, while the rate of ACE formation increases by a factor of four. In contrast, the rate of PRO formation is decreased by a factor of more than 30. As the mass fraction of DMSO is increased from 25 to 90 wt%, the rate constants for PDO consumption and the formation of both products generally increase together.

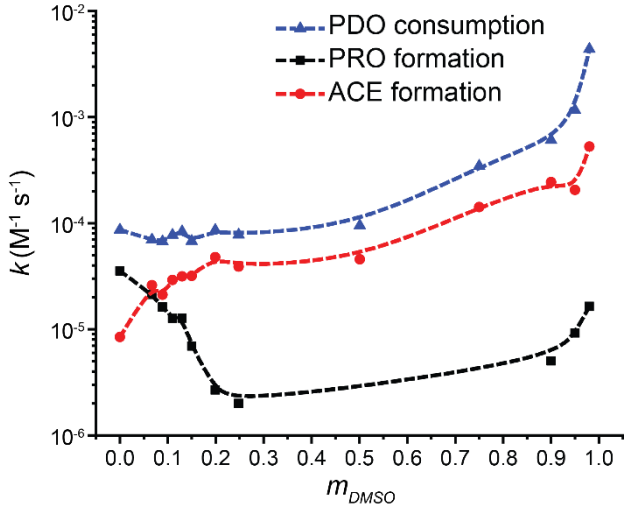


Figure 6.3.3. Apparent rate constants for reactant conversion and product formation for 1,2-propanediol (PDO) dehydration in mixtures of water with dimethyl sulfoxide (DMSO) as a function of the mass fraction of DMSO (m_{DMSO}). Reaction conditions: ~20 mol% conversion; 433 *K*; 0.4-0.005 M triflic acid; 0.1 M PDO; 150 min reaction time; 500 rpm stirring rate, 2 mL total solvent volume. Dashed lines are visual aids.

6.3.3. Relating selectivity trends to tabulated cosolvent properties

Based on these results, we propose that DMSO alters the selectivity of PDO dehydration by suppressing the rate of PRO formation; or equivalently, by increasing the activation energy barrier for PRO formation. We sought to determine if this behavior could generalize to other cosolvents based on the hypothesis that solvent polarizability or the dielectric constant could relate to activation energy barriers that lead to a specific product.(*10*) Previous work has also shown that

more nucleophilic (basic) or polar solvents, such as DMSO, can facilitate distinct reaction mechanisms compared to pure water.(*35*) For example, strongly nucleophilic solvents can facilitate concerted solvolysis of tertiary carbons *via* formation of partial covalent bonds between the transition state and polarized solvent molecules.(*36*) A concerted mechanism for PDO dehydration would preclude the need to pass through an unstable primary carbocation to afford ACE (Figure 6.3.1), which might explain the aforementioned shift in ACE selectivity in DMSO-water mixtures(*37*) and the decrease in PRO formation. We thus carried out the acid-catalyzed dehydration of PDO in four other cosolvent-water mixtures at 90 wt% cosolvent to determine the relationship between cosolvent polarity and basicity to selectivity. Selected cosolvents include □-valerolactone (GVL), tetrahydrofuran (THF), *N*-methyl-2-pyrrolidone (NMP), and tetramethylene sulfoxide (TMSO) because these cosolvents have dielectric constants that lie between those of DIO (2.20) and DMSO (46.46) (Table 6.3.1). To facilitate comparison of relative reaction rates between mixed-solvent environments, Equation 6.5 defines the product kinetic solvent parameter (σ^p with p indicating a specific product) based on the rate of product formation in the mixed-solvent environment (k^p_{org}) and the rate of product formation in pure water ($k^p_{H_2O}$):

$$\sigma^p = \log_{10}\left(\frac{k^p_{org}}{k^p_{H_2O}}\right) \tag{6.5}$$

$\sigma^p > 0$ indicates increased product formation in a particular mixed-solvent environment compared to the same reaction in pure water, while $\sigma^p < 0$ has the converse implication.

Table 6.3.1 presents apparent rate constants for acid-catalyzed PDO dehydration, selectivities to PRO and ACE, and σ^{PRO} for each of the cosolvent-water mixtures. Entries in Table 6.3.1 are organized in descending order based on σ^{PRO}. Based on the hypothesis that more polar or basic cosolvents might facilitate a distinct mechanism for PDO dehydration compared to

pure water, Table 6.3.1 also displays the cosolvents' dielectric constant (*38, 39*) and basicity, as expressed by the pK_a's of their protonated conjugate acids (pK_a*).(*40*) Larger dielectric constants indicate a more polar solvent and pK_a* values below -1.71 indicate solvents that are more basic than pure water. With the exception of DMSO-water mixtures, PRO is the preferred product in all solvent systems as expected based on the reaction mechanism. Furthermore, neither the product selectivities nor σ^{PRO} correlate with the dielectric constant or the basicity of the organic cosolvent.(*41*) These observations illustrate that properties of mixed-solvent environments cannot generally be understood in terms tabulated descriptors of the pure individual components.

Table 6.3.1. Apparent rate constants (k_{app}), product kinetic solvent parameters (σ^{PRO}), and selectivities of Brønsted acid-catalyzed 1,2-propanediol (PDO) dehydration in mixed-solvent environments. The mass fraction of the organic component was 90 wt% in all experiments (excluding pure water). Rate constants are derived from Equations 1 and 2. Selectivities are derived from Equation 3. The standard error in selectivities is +/- 5 mol%. Reaction conditions: ~20 mol% conversion; 433 K; 0.4 M triflic acid; 0.1 M PDO; 150 min reaction time; 500 rpm stirring rate, 2 mL total solvent volume.

Cosolvent	pK_a*	Dielectric constant	k_{app} $(M^{-1}ks^{-1})$	Conversion $(mol\%)$	Selectivity $(mol\%)$		σ^{PRO}
					PRO	ACE	
GVL	-7.00	36.47	0.880	55	68	0	1.24
DIO	-3.00	2.20	0.336	12	34	0	0.52
THF	-2.05	7.40	0.434	23	8	0	0.48
NMP	-0.71	32.16	0.084	24	52	1	0.09
TMSO	--	42.84	0.281	14	3	0	-0.44
DMSO	-1.8	46.46	0.670	37	1	33	-0.57
pure water	**-1.71**	**80.10**	**0.084**	**9**	**41**	**9**	**--**

**basicity is quantified as the pK_a of the solvent's protonated conjugate acid. For example, -2.05 is the pK_a of the protonated ether group on THF.*

6.3.4. Equilibrium solvation: quantifying reactant and product solvation free energies

Having demonstrated that the selectivity of PDO dehydration in mixed-solvent environments cannot be rationalized by the pure cosolvents' tabulated bulk properties, we now use MD to probe the molecular bases by which mixed-solvent environments might influence selectivity to ACE and PRO. We hypothesize that for the mixed-solvent environments considered here, solvent relaxation is fast relative to the timescales associated with reaction elementary steps, such that the activation energy of the reaction for the formation of a specific product is correlated with the difference between the free energies of the reactant and product. Similar hypotheses, all of which follow from the Hammond postulate, lead to the typical linear free energy relationships that relate changes in reaction rates to changes in the energy of reactant and product states in catalytic processes.(*42, 43*) This hypothesis is equivalent to assuming that equilibrium solvation effects dictate selectivity and suggests that changes in selectivity can be related to the free energies of the reactants and products without knowledge of the reaction mechanism or explicitly modeling the transition state itself. Figure 6.3.4A schematically illustrates the hypothesized effect of a mixed-solvent environment on a reaction free energy landscape that would influence selectivity to a particular product. Furthermore, *ab initio* MD simulations have shown that the PDO dehydration mechanism that forms PRO is characterized by a late transition state (*3*) that is analogous to the late transition state formed in the acid-catalyzed conversion of fructose to HMF.(*14*) This finding emphasizes the importance of quantifying product stability because in many instances the structure of a late transition state more closely resembles the product than the reactant.

To investigate the equilibrium solvation mechanism, we use classical MD simulations to quantify the solvation free energy of each product and reactant in each experimentally studied mixed-solvent environment. The solvation free energy of a species is defined as the change in free

energy associated with transferring the species from vacuum to solvent; the difference in the solvation free energy of a species in two solvent environments is the change in the free energy associated with transferring the species from one solvent environment to the other. Figure 6.3.4B schematically defines the thermodynamic cycle used to determine the influence of the solvent environment on the reactant and product free energies. The free energy of transferring species s from water to a mixed-solvent environment ($\Delta G_s^{H_2O \rightarrow org}$) is related to the solvation free energy of *s* in pure water ($G_s^{H_2O}$) and the solvation free energy of *s* in a mixed-solvent environment (G_s^{org}) by Equation 6.6:

$$\Delta G_s^{H_2O \rightarrow org} = G_s^{org} - G_s^{H_2O} \tag{6.6}$$

A negative value of $\Delta G_s^{H_2O \rightarrow org}$ indicates that s is more favorably solvated in the mixed-solvent environment than in pure water. We then quantify the free energy difference between the reactant (r) and product (p) in a mixed-solvent environment ($\Delta\Delta G$) by performing four solvation free energy calculations that are related by Equation 7.7:

$$\Delta\Delta G = \Delta G_p^{H_2O \rightarrow org} - \Delta G_r^{H_2O \rightarrow org} \tag{6.7}$$

If $\Delta\Delta G < 0$, the difference between the free energy of the product and the reactant is more negative in the mixed-solvent environment than in pure water, indicating that the product is stabilized to a greater degree than the reactant (Figure 6.3.4A). Therefore, we hypothesize that a negative value of $\Delta\Delta G$ for a product in a given mixed-solvent environment should increase its rate of formation compared to pure water due to equilibrium solvation, and conversely a positive value of $\Delta\Delta G$ for a product should decrease its rate of formation. This hypothesis is qualitatively consistent with the free energy trends observed in mixed-solvent environments for *tert*-butanol and PDO dehydration using *ab initio* MD simulations.(*3*)

Figure 6.3.4C plots $\Delta\Delta G$ for PRO and ACE against experimentally determined values of σ^{PRO} in 90 wt% organic cosolvents. $\Delta\Delta G$ for PRO formation is negative for aqueous mixtures containing 90 wt% GVL, DIO, THF, or NMP. For these same mixtures, σ^{PRO} is positive, indicating that reaction rates for PRO formation are enhanced relative to pure water and agreeing with the simulation prediction. Conversely, $\Delta\Delta G > 0$ and $\sigma^{PRO} < 0$ for 90 wt% TMSO and DMSO mixtures, indicating that in these systems equilibrium solvation effects predict the suppression of PRO formation (Figure 6.3.3). Figure 6.3.4D plots the correlation between $\Delta\Delta G$ and σ^{PRO} for each of the solvent mixtures from Figure 6.6.3 and for aqueous mixtures containing 25, 50, and 75 wt% DIO. As hypothesized, $\Delta\Delta G$ and σ^{PRO} exhibit a negative linear correlation (Pearson's $r = -0.81$) with a root-mean-squared error (RMSE) of 0.33 between values of σ^{PRO} predicted using the linear correlation and experimentally determined values. These results support the hypothesis that equilibrium solvation can account for the effect of mixed-solvent environments on the rate of PRO formation. However, we found that $\Delta\Delta G$ does not capture the non-monotonic behavior found in σ^{PRO} across different mass fractions of DIO and DMSO and $\Delta\Delta G$ for PRO is more negative than ACE for all solvents, suggesting that PRO should be the preferred product compared to ACE. This data thus does not explain the experimental finding that ACE is preferentially formed over PRO in DMSO-water mixtures, suggesting that the formation of ACE in the presence of DMSO is not captured by equilibrium solvation effects as will be further discussed below.

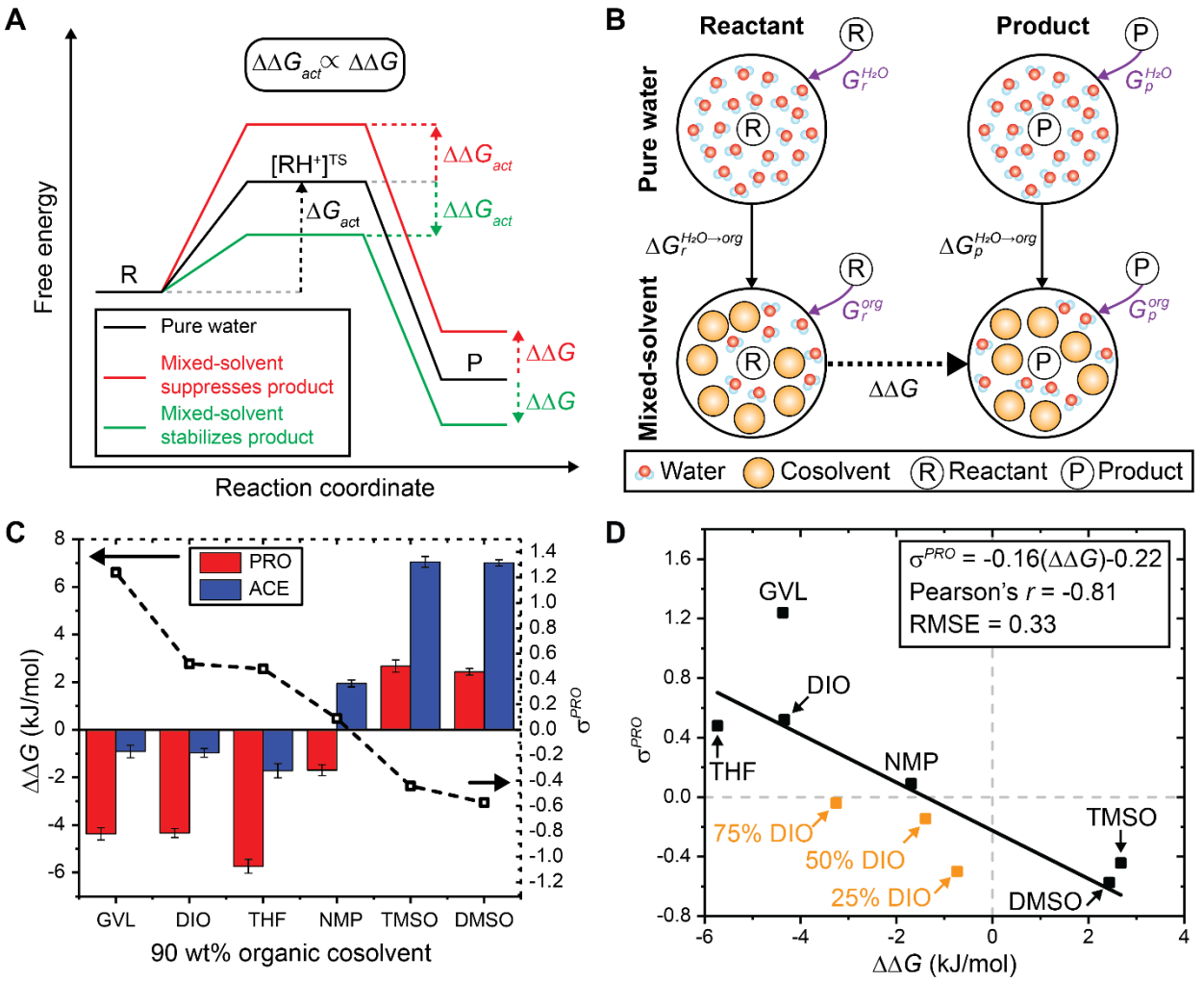


Figure 6.3.4. A Hypothesized effect of mixed-solvent environment on the free energies of reactant, transition, and product states. The change in the relative free energy between the reactant and product states ($\Delta\Delta G$) is proportional to the change in the activation energy ($\Delta\Delta G_{act}$) for a reaction in a mixed-solvent environment compared to the same reaction in pure water. **B** Thermodynamic cycle to calculate the free energy difference between a reactant and product in a mixed-solvent environment relative to pure water. Purple arrows indicate solvation free energies computed from MD simulations which are used to calculate the transfer free energies indicated by filled black arrows. The dashed black arrow indicates $\Delta\Delta G$. **C** Comparison of simulated $\Delta\Delta G$ for PRO (red bar) and ACE (blue bar) and experimental kinetic solvent parameter for PRO formation (σ^{PRO},

dashed black lines) in 90 wt% organic cosolvents. **D** Correlation between σ^{PRO} and $\Delta\Delta G$ for 90 wt% mass fraction of organic solvent (black) and various wt% mass fractions of DIO (orange).

6.3.5. Equilibrium solvation effects extended to other dehydration reactions

To determine if $\Delta\Delta G$ can quantify product formation for different dehydration reactions, we performed the acid-catalyzed dehydration of three other representative polyols in 90 wt% GVL- and DMSO-water mixtures: *cis*-1,2-cyclohexanediol, *trans*-1,2-cyclohexanediol, and *trans*-1,3-cyclohexenediol; dehydration reaction schemes are shown in Figure 6.3.5A. GVL and DMSO were selected as cosolvents because they correspond to the highest and lowest rates of PRO formation from PDO dehydration (Table 6.3.1). Table 6.3.2 shows the rates and the selectivities of these reactions towards the dehydration products in 90 wt% GVL- and DMSO-water mixtures over a triflic acid catalyst at 433 K. For all reactions, we find that the selectivity to the dehydration products in Table 6.3.2 is higher in 90 wt% GVL than 90 wt% DMSO and $\Delta\Delta G$ is more negative in 90 wt% GVL than in 90 wt% DMSO. Figure 6.3.5B shows the correlation between σ^{p} (where p refers generically to the dehydration product of each reaction) and $\Delta\Delta G$ for each different reaction in 90 wt% GVL- and DMSO-water mixtures. σ^{p} and $\Delta\Delta G$ are again negatively correlated for each separate reaction, agreeing with the results for PDO dehydration (Figure 6.3.4). This result indicates that reactant and product solvation free energies in mixed-solvent environments can provide insight into the selectivity for a dehydration product.

Table 6.3.2. Dehydration of polyols over triflic acid in 90 wt% GVL-water and DMSO-water mixtures. Reaction conditions: 0.1 M reactant, 433 K, 0.05-0.6 M TfOH, 500 rpm stirring, reaction time 0-120 min.

Reaction	Solvent	k_{app} $(M^{-1}ks^{-1})$	Conversion $(mol\%)$	Selectivity $(mol\%)$[a]	σ^p	$\Delta\Delta G$ $(kJ\ mol^{-1})$
1	GVL	0.88	55	68	1.23	-4.37
	DMSO	0.67	37	1	-0.71	2.44
2	GVL	45.74	56	70	1.85	-3.01
	DMSO	1.52	31	27	-0.04	0.98
3	GVL	3.13	18	3	1.54	-2.43
	DMSO	0.34	11	1	-0.08	2.65
4[b]	GVL	68.19	33	99	2.27	-5.14
	DMSO	0.32	44	25	-0.65	-1.32

[a]dehydration products shown in Figure 6.3.5A.

[b]consists of equal *cis* and *trans* isomers of 1,3-cyclohexanediol. The $\Delta\Delta G$ value is the average result from both isomers.

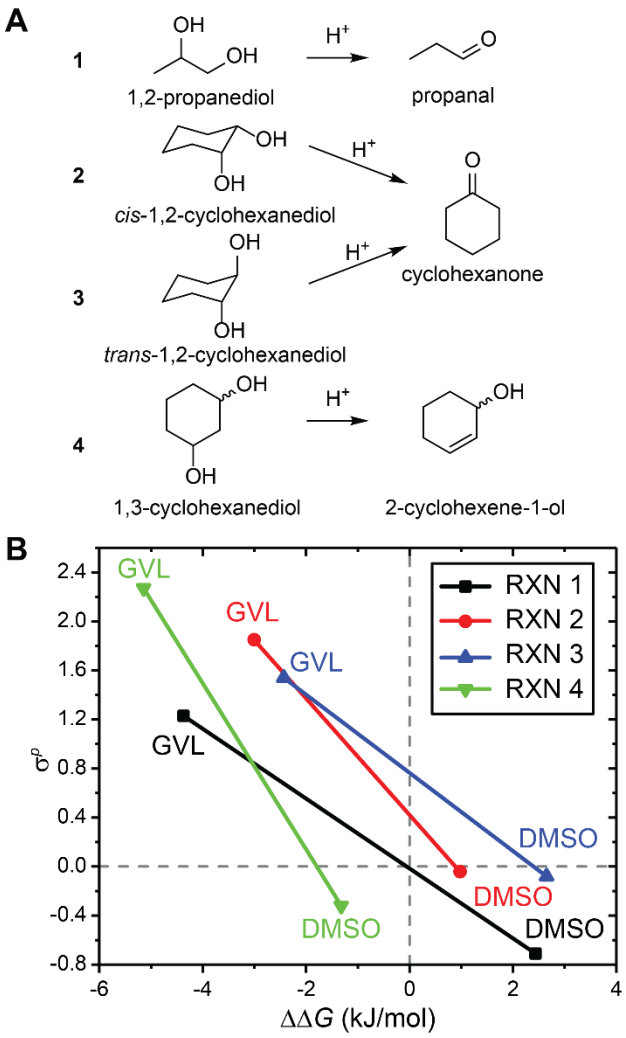


Figure 6.3.5. A Four acid-catalyzed dehydration reactions of representative polyols. **B** Correlation between experimental kinetic solvent parameter for dehydration product formation (σ^p) and $\Delta\Delta G$ for each of the four reactions in 90 wt% GVL- and DMSO-water mixtures.

6.3.6. Participation of DMSO in dehydration reactions

Whereas DMSO-water mixtures are the only solvent systems tested which afford ACE as a dehydration product from PDO, solvation free energy calculations show that $\Delta\Delta G$ is more negative for ACE than PRO in all solvent mixtures (Figure 6.3.4C). This indicates that equilibrium

solvation effects alone cannot explain the formation of ACE in DMSO-water mixtures. Alternatively, we hypothesize that DMSO molecules can participate directly in the reaction mechanism, adopting specific spatial positions around the reactant to stabilize transition states that lead to ACE. Indeed, the ability of DMSO to mediate distinct reaction mechanisms compared to other polar aprotic solvents has been previously reported. For instance, *ab initio* molecular dynamics simulations show that DMSO can act as a proton acceptor to mediate the proton transfer steps in glucose dehydration reactions.(*44*) As noted previously, DMSO is also thought to promote fructose conversion to HMF by shielding HMF from rehydration.(*7*) These processes both require that DMSO molecules localize near the reactant in mixed-solvent environments to facilitate interactions with transition states.

To probe the hypothesis that DMSO facilitates a distinct reaction pathway by participating directly in the mechanism, Figure 6.3.6 shows spatial distribution maps that quantify the density of both water and DIO or DMSO molecules in the three-dimensional volume near PDO. Each density is normalized by the density of the bulk solvent, and spatial distribution maps are computed for PDO in pure water, 90 wt% DIO, and 90 wt% DMSO with normalized density values ranging from 1.50-3.00 for water (shown in red) and from 1.30 to 1.50 for DMSO (shown as blue). These density values were selected to show spatial regions enriched in either water or cosolvent molecules relative to the composition of the bulk solvent.

In pure water, there is slight enrichment of water around the hydroxyl groups of PDO, indicating a near-uniform distribution of water molecules as expected. In 90 wt% DIO, however, the spatial distribution maps display two distinct, water-enriched features associated with each hydroxyl moiety. We interpret these features to indicate water solvating both the *hydrogen bonding donor and acceptor facets* of the hydroxyl groups in this solvent system as compared to pure water.

This observation is consistent with our previous work in which water-enriched local domains were shown to form around hydrophilic reactants in the presence of DIO.(*1*) In 90 wt% DMSO, however, there is enrichment of *both water and DMSO* molecules around the hydroxyl groups and, importantly, the water-enriched features corresponding to the hydrogen bonding acceptor facets of the hydroxyl groups are missing.

To connect this solvation behavior the unique reactivity of PDO in the DMSO-water system, we note that the hydrogen bonding acceptor facet on a hydroxyl group would corresponding to the site at which an acidic proton is transferred from a hydronium ion to the reactant.(*7*) Therefore, in DMSO-water mixtures, it follows that the reaction mechanism underlying PDO dehydration may not be initiated in this manner. Finally, we note that DMSO itself can act as a base,(*45*) and that the stability of charged intermediates in base-catalyzed reaction mechanisms follow opposite trends as compared to their acid-catalyzed counterparts (a primary carbanion is relatively stable compared to a primary carbocation).(*46*) Following these points, we conjectured that DMSO might suppress the acid-catalyzed reaction mechanism that affords PRO by occluding the hydronium ion's access to the hydroxyl groups, while simultaneously acting as a base to remove the primary OH group in the form of water. To probe this hypothesis, we tested the dehydration of 1-butanol dehydration and of PDO, in 90 wt% DMSO-water mixtures, in the presence and absence of a TfOH acid catalyst, respectively. However, in the presence of an acid catalyst, none of the corresponding dehydration product (butene) was formed from 1-butanol; and in the absence of an acid catalyst, PDO underwent no reaction. Thus, we reject the hypothesis that DMSO is acting as a base catalyst to afford ACE from PDO. All things considered then, DMSO's ability to form ACE from PDO in mixed solvent environments can only be explained by a mechanism which is consistent with the following observations:

1. DMSO participates in the reaction mechanism, but an acid catalyst is still required;
2. DMSO is able to remove a primary hydroxyl group in the form of water, but only when a second hydroxyl group is vicinal to the reaction site, and;
3. the mechanism is likely not initiated by a proton transfer.

One possible mechanism consistent with these observations is displayed in Figure 6 below. A strong acid activates DMSO,(*47*) which adds to the primary alcohol group of PDO, displacing a proton in the process. The primary alcohol is selectively attacked in this step as its pKa is lower than that of a secondary alcohol group.(*48*) Finally, the acidic proton catalyzes a semipinacol rearrangement(*49*) to yield the final products. To probe this proposed mechanism, we measured the kinetics of glycerol dehydration in 90 wt% GVL- and DMSO-water mixtures at 433 K over a triflic acid catalyst. In the DMSO-water mixtures, hydroxyacetone (a primary dehydration product analogous to ACE) was afforded in near quantitative yields. In contrast, no hydroxyacetone was formed in GVL, supporting the fact that DMSO can facilitate distinct reactivity compared to other polar aprotic cosolvents, an in particular is able to facilitate the acid-catalyzed removal of a primary, vicinal hydroxyl group.

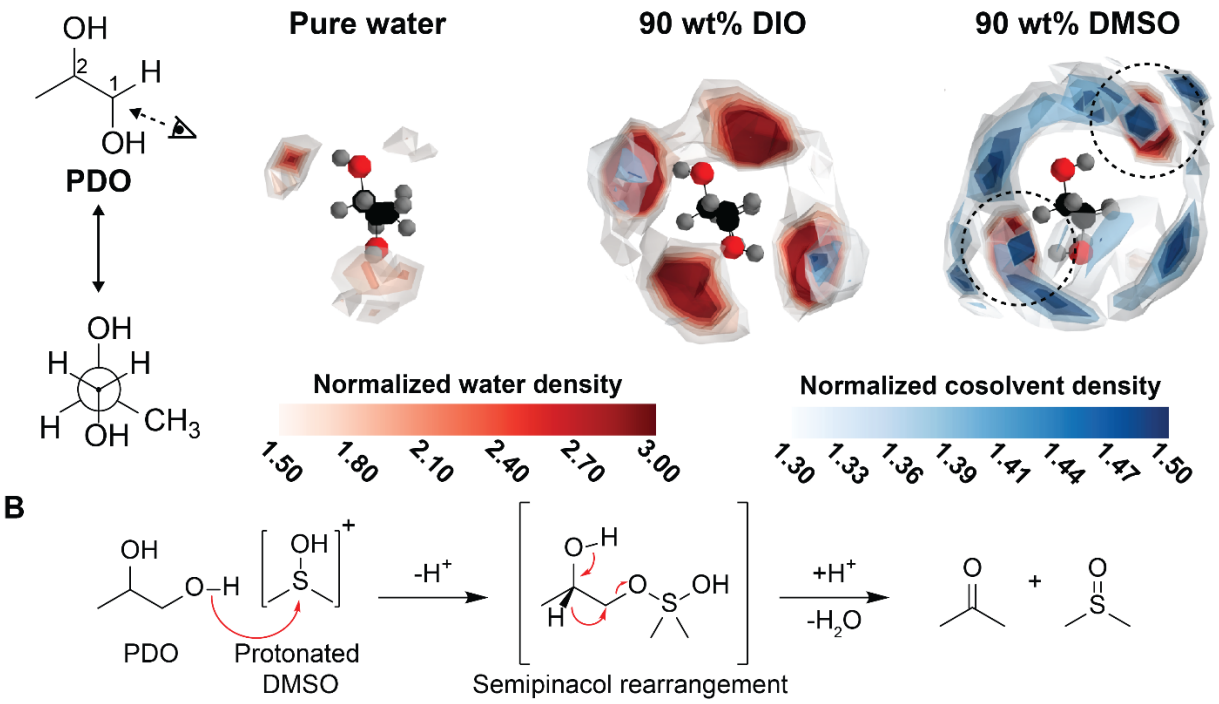


Figure 6.3.6. Spatial distribution maps of 1,2-propanediol in pure water, 90 wt% DIO, and 90 wt% DMSO. 1,2-propanediol is positioned so the view angle is along the C1-C2 bond as in the projection diagram at left. Isovalues between 1.50-3.00 are shown for water in red. Isovalues between 1.30-1.50 are shown for cosolvent in blue. Dashed lines are drawn to emphasize the competition of water and DMSO around the hydroxyl groups of PDO.

6.3.7. Solvent effects on dehydration of 1,2-cyclohexanediol stereoisomers

In unexpected finding highlighted in Table 6.3.2 is that the rate constant for the dehydration of *cis*-1,2-cyclohexanediol is an order of magnitude larger than the rate constant for the dehydration of *trans*-1,2-cyclohexanediol, even though both reactions yield cyclohexanone. This observation suggests that the dehydration of the two isomers proceeds *via* a concerted mechanism, such as the mechanism shown in Figure 6.3.77A, in which the stereochemistry of the substituents participating in the reaction has a significant effect on the reaction rate. In contrast, 1,3-cyclohexanediol was obtained as an equilibrium mixture of *cis*- and *trans*-isomers, both of which

were converted to the corresponding dehydration product at equal rates. This observation suggests that the latter reaction proceeds through a stepwise mechanism, where the stereochemistry of the reactant has no effect upon the stability of the transition state between the reactant and products.

To visualize differences in the solvent environment that could lead to improved reaction performance for *cis*-1,2-cyclohexanediol dehydration reaction, Figure 6.3.7 shows spatial distribution maps of *cis*- and *trans*-1,2-cyclohexanediol in pure water, 90 wt% GVL, and 90 wt% DMSO. The solvation structure of the *cis*-isomer shares a distinct water domain between the hydroxyl groups in 90 wt% GVL, which could help stabilize a transition state associated with a concerted mechanism. This shared water domain is not found in the *trans*-isomer case, potentially explaining the observed difference in reaction rate between the two stereoisomers. Because Table 2 shows that dehydration reactions are suppressed in DMSO for both stereoisomers, this behavior suggests that the equilibrium solvation effect probed by the MD simulations generalizes to reactions proceeding through both concerted and sequential mechanisms. Moreover, the spatial distribution maps again show that DMSO molecules, but not GVL molecules, compete with water to solvate hydroxyl groups around both stereoisomers. This behavior is again consistent with the solvation of PDO shown in Figure 6 and indicates that the unique effects of DMSO generalize to stereoisomers. Detailed *ab initio* calculations will be required to probe the importance of the shared water domain in facilitating acid-catalyzed dehydration reactions and is a subject of future work.

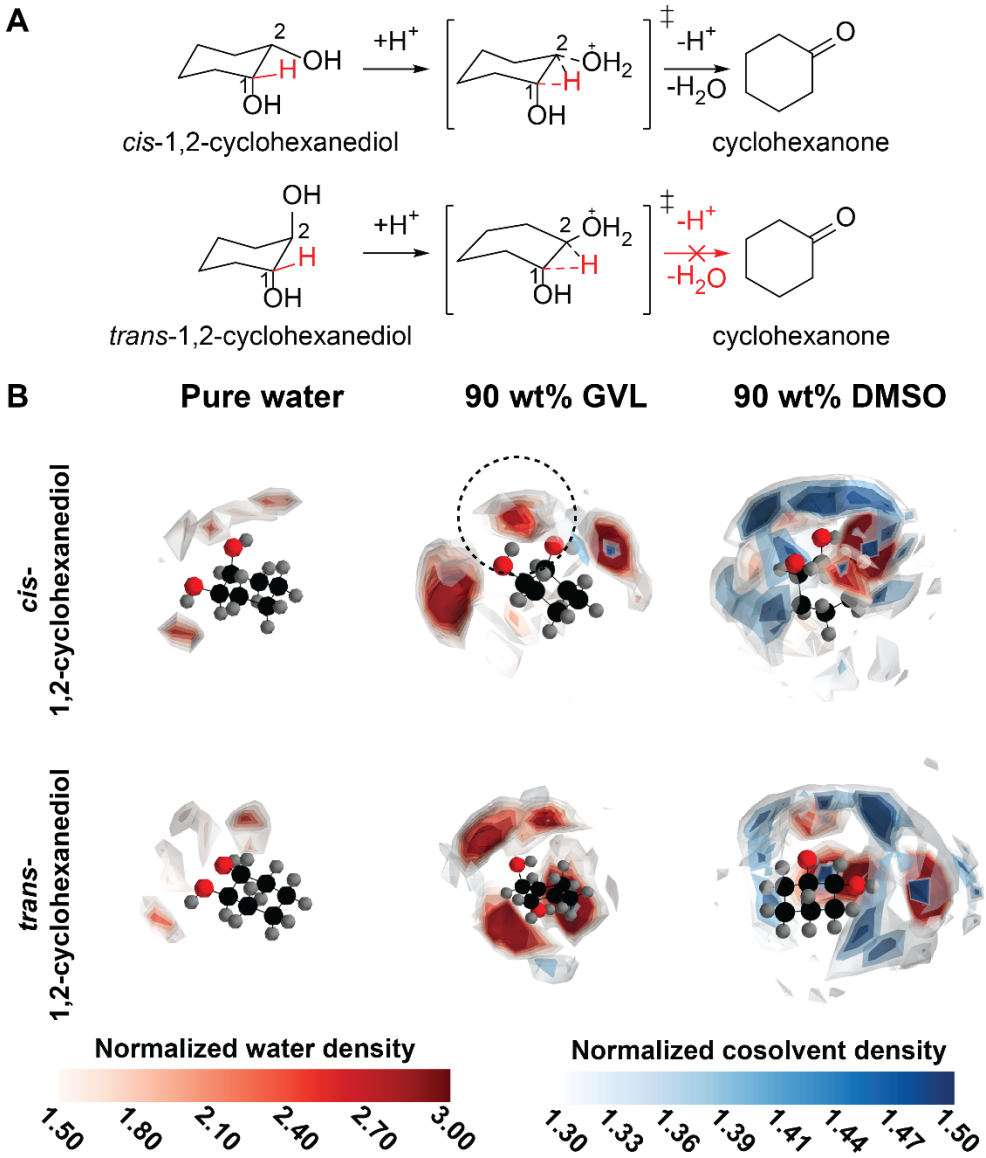


Figure 6.3.7. A Proposed concerted mechanism for the conversion of *cis*- and *trans*-1,2-cyclohexanediol to afford cyclohexanone. **B** Spatial distribution maps of *cis*- and *trans*-1,2-cyclohexanediol in pure water, 90 wt% GVL, and 90 wt% DMSO. Isovalues between 1.50-3.00 are shown for water in red. Isovalues between 1.30-1.50 are shown for cosolvent in blue. Dashed lines are drawn to emphasize the shared local water domain around the *cis*-isomer that may better facilitate a concerted mechanism; this domain is not found in the *trans*-isomer.

6.4. CONCLUSIONS

We have studied the Brønsted acid-catalyzed dehydration of PDO as a model compound for biomass conversion. Reactions were carried out in pure water and in aqueous mixtures of □-valerolactone, 1,4-dioxane, tetrahydrofuran, *N*-methyl-2-pyrrolidone, tetramethylene sulfoxide, and dimethyl sulfoxide at 433 K over a triflic acid catalyst to demonstrate that the selectivity of PDO dehydration can be modulated by solvent composition. PRO is the major product in all mixed-solvent environments, except in mixtures of water with DMSO, in which ACE is the major project. We find that the presence of DMSO in the solvent system suppresses the rate of propanal formation. We attribute this suppression behavior to increases in activation energy barriers for transition states corresponding to PRO formation in mixed-solvent environments. We probe changes to these activation barriers by assuming that the reactant and product states of 1,2-propanediol and propanal, respectively, are correlated in energy with their corresponding transition states. Using classical molecular dynamics simulations, we use solvation free energy calculations to quantify the free energy difference between the reactant and product in a mixed-solvent environment relative to water ($\Delta\Delta G$) and find that this free energy difference correlates with the experimental rate of propanal formation in five cosolvent systems. $\Delta\Delta G$ also captures trends in the suppression of the main dehydration product in three other acid-catalyzed dehydration reactions: *cis*-1,2-cyclohexanediol, *trans*-1,2-cyclohexanediol, and 1,3-cyclohexanediol. However, $\Delta\Delta G$ does not predict the preference for acetone formation in DMSO-water mixtures, suggesting that equilibration solvation of the reactant and product states does not determine the primary product. Analysis of the spatial distribution of solvent molecules around 1,2-propanediol instead shows that DMSO uniquely competes with water for reactive sites, which can lead to changes in mechanisms that favor steps towards forming acetone. Similar analysis of the local solvent environment near

cis- and *trans*-1,2-cyclohexanediol supports the hypothesis that dehydration of these stereoisomers involves a concerted mechanism, with the decreased reaction rates for both isomers in DMSO suggesting that the above insights apply to both sequential and concerted reaction mechanisms.

These findings show that MD simulations can capture trends of the suppression of the main dehydration product. This methodology represents a step toward the rational design of mixed-solvent environments for liquid-phase reaction schemes and has the potential to alleviate the time-intensive exercise of experimentation that typically accompanies the development of new processes. Estimating reaction performance by measuring the product and reaction state stabilities in mixed-solvent environments relies upon the assumption that these states are correlated with transition state energies. Therefore, we caution against over-interpretation of classical MD and note that further *ab initio* studies would be needed to fully quantify equilibrium or non-equilibrium solvent effects in these systems. However, from the perspective of solvent screening, classical MD appears to be sufficient to capture trends and can be a useful guide for *ab initio* studies or experimental investigations.

6.5. ACKNOWLEDGEMENTS

This work was supported by the DOE Great Lakes Bioenergy Research Center (http://www.glbrc.org). The U.S. Department of Energy, Office of Sciences, Office of Biological and Environmental Research support the GLBRC, through the cooperative agreement BER DE-FC02-07ER64494 between the Board of Regents of the University of Wisconsin and the U.S. Department of Energy. T.W. Walker acknowledges that this material is based on work supported by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin–Madison with funding from the Wisconsin Alumni Research Foundation. This work

used the computing resources and assistance of the UW-Madison Center for High Throughput Computing (CHTC) in the Department of Computer Sciences. The CHTC is supported by UW-Madison, the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation, the Wisconsin Institutes for Discovery, and the National Science Foundation, and is an active member of the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science.

6.6. References

1. T. W. Walker *et al.*, Universal kinetic solvent effects in acid-catalyzed reactions of biomass-derived oxygenates. *Energy & Environmental Science* **11**, 617-628 (2018).
2. M. A. Mellmer *et al.*, Solvent effects in acid-catalyzed biomass conversion reactions. *Angewandte chemie international edition* **53**, 11872-11875 (2014).
3. M. A. Mellmer *et al.*, Solvent-enabled control of reactivity for liquid-phase reactions of biomass-derived compounds. *Nature Catalysis* **1**, 199-207 (2018).
4. J. N. Chheda, G. W. Huber, J. A. Dumesic, Liquid-phase catalytic processing of biomass-derived oxygenated hydrocarbons to fuels and chemicals. *Angewandte Chemie International Edition* **46**, 7164-7183 (2007).
5. A. Corma, S. Iborra, A. Velty, Chemical Routes for the Transformation of Biomass into Chemicals. *Chemical Reviews* **107**, 2411-2502 (2007).
6. Y. Román-Leshkov, J. N. Chheda, J. A. Dumesic, Phase modifiers promote efficient production of hydroxymethylfurfural from fructose. *Science* **312**, 1933-1937 (2006).
7. S. H. Mushrif, S. Caratzoulas, D. G. Vlachos, Understanding solvent effects in the selective conversion of fructose to 5-hydroxymethyl-furfural: a molecular dynamics investigation. *Physical Chemistry Chemical Physics* **14**, 2637-2644 (2012).
8. J. J. Varghese, S. H. Mushrif, Origins of complex solvent effects on chemical reactivity and computational tools to investigate them: a review. *Reaction Chemistry & Engineering* **4**, 165-206 (2019).
9. G. Shrivastav, T. S. Khan, M. Agarwal, M. A. Haider, A Car–Parrinello Molecular Dynamics Simulation Study of the Retro Diels–Alder Reaction for Partially Saturated 2-Pyrones in Water. *The Journal of Physical Chemistry C* **122**, 11599-11607 (2018).
10. N. Guo, S. Caratzoulas, D. J. Doren, S. I. Sandler, D. G. Vlachos, A perspective on the modeling of biomass processing. *Energy & Environmental Science* **5**, 6703-6716 (2012).
11. S. Caratzoulas, T. Courtney, D. G. Vlachos, Hybrid Quantum Mechanics/Molecular Mechanics-Based Molecular Dynamics Simulation of Acid-Catalyzed Dehydration of Polyols in Liquid Water. *The Journal of Physical Chemistry A* **115**, 8816-8821 (2011).
12. G. Giorgianni *et al.*, Effect of the Solvent in Enhancing the Selectivity to Furan Derivatives in the Catalytic Hydrogenation of Furfural. *ACS Sustainable Chemistry & Engineering* **6**, 16235-16247 (2018).

13. S. Gupta, M. I. Alam, T. S. Khan, N. Sinha, M. A. Haider, On the mechanism of retro-Diels–Alder reaction of partially saturated 2-pyrones to produce biorenewable chemicals. *RSC Advances* **6**, 60433-60445 (2016).
14. S. Caratzoulas, D. G. Vlachos, Converting fructose to 5-hydroxymethylfurfural: a quantum mechanics/molecular mechanics study of the mechanism and energetics. *Carbohydrate Research* **346**, 664-672 (2011).
15. S. H. Mushrif, J. J. Varghese, C. B. Krishnamurthy, Solvation dynamics and energetics of intramolecular hydride transfer reactions in biomass conversion. *Physical Chemistry Chemical Physics* **17**, 4961-4969 (2015).
16. V. Vasudevan, S. H. Mushrif, Insights into the solvation of glucose in water, dimethyl sulfoxide (DMSO), tetrahydrofuran (THF) and N,N-dimethylformamide (DMF) and its possible implications on the conversion of glucose to platform chemicals. *RSC Advances* **5**, 20756-20763 (2015).
17. M. K. Hazra, B. Bagchi, Non-equilibrium solvation dynamics in water-DMSO binary mixture: Composition dependence of non-linear relaxation. *The Journal of Chemical Physics* **149**, 084501 (2018).
18. L. Shuai, J. Luterbacher, Organic solvent effects in biomass conversion reactions. *ChemSusChem* **9**, 133-155 (2016).
19. E. Raamat *et al.*, Acidities of strong neutral Brønsted acids in different media. *Journal of Physical Organic Chemistry* **26**, 162-170 (2013).
20. A. Trummal, L. Lipping, I. Kaljurand, I. A. Koppel, I. Leito, Acidity of Strong Acids in Water and Dimethyl Sulfoxide. *The Journal of Physical Chemistry A* **120**, 3663-3669 (2016).
21. S. Páll, M. J. Abraham, C. Kutzner, B. Hess, E. Lindahl. (Springer International Publishing, Cham, 2015), pp. 3-27.
22. R. B. Best *et al.*, Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain $\chi1$ and $\chi2$ Dihedral Angles. *Journal of Chemical Theory and Computation* **8**, 3257-3273 (2012).
23. K. Vanommeslaeghe *et al.*, CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* **31**, 671-690 (2010).
24. W. Yu, X. He, K. Vanommeslaeghe, A. D. MacKerell Jr., Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *Journal of Computational Chemistry* **33**, 2451-2468 (2012).
25. H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma, The missing term in effective pair potentials. *The Journal of Physical Chemistry* **91**, 6269-6271 (1987).
26. A. K. Chew, R. C. Van Lehn, Quantifying the Stability of the Hydronium Ion in Organic Solvents With Molecular Dynamics Simulations. *Frontiers in Chemistry* **7**, (2019).
27. D. Shivakumar *et al.*, Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *Journal of Chemical Theory and Computation* **6**, 1509-1519 (2010).
28. D. Horinek, S. I. Mamatkulov, R. R. Netz, Rational design of ion force fields based on thermodynamic solvation properties. *The Journal of Chemical Physics* **130**, 124507 (2009).
29. D. J. Bonthuis, S. I. Mamatkulov, R. R. Netz, Optimization of classical nonpolarizable force fields for OH^- and $H3O^+$. *The Journal of Chemical Physics* **144**, 104503 (2016).

30. M. R. Shirts, J. D. Chodera, Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics* **129**, 124105 (2008).
31. P. V. Klimovich, M. R. Shirts, D. L. Mobley, Guidelines for the analysis of free energy calculations. *Journal of Computer-Aided Molecular Design* **29**, 397-411 (2015).
32. Robert T. McGibbon *et al.*, MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **109**, 1528-1532 (2015).
33. D. Zhang, S. A. I. Barri, D. Chadwick, Dehydration of 1,2-propanediol to propionaldehyde over zeolite catalysts. *Applied Catalysis A: General* **400**, 148-155 (2011).
34. M. Akiyama, S. Sato, R. Takahashi, K. Inui, M. Yokota, Dehydration–hydrogenation of glycerol into 1,2-propanediol at ambient hydrogen pressure. *Applied Catalysis A: General* **371**, 60-66 (2009).
35. W. P. Jencks, Ingold Lecture. How does a reaction choose its mechanism? *Chemical Society Reviews* **10**, 345-375 (1981).
36. J. P. Richard, M. M. Toteva, T. L. Amyes, What Is the Stabilizing Interaction with Nucleophilic Solvents in the Transition State for Solvolysis of Tertiary Derivatives: Nucleophilic Solvent Participation or Nucleophilic Solvation? *Organic Letters* **3**, 2225-2228 (2001).
37. K. Nakamura, Y. Osamura, Theoretical study of the reaction mechanism and migratory aptitude of the pinacol rearrangement. *Journal of the American Chemical Society* **115**, 9112-9120 (1993).
38. F. W. Fowler, A. R. Katritzky, R. J. D. Rutherford, The correlation of solvent effects on physical and chemical properties. *Journal of the Chemical Society B: Physical Organic*, 460-469 (1971).
39. C. Wohlfarth, M. D. Lechner, B. Wohlfarth, *Surface Tension of Pure Liquids and Binary Liquid Mixtures*. (Springer Berlin Heidelberg, 1997).
40. A. Bagno, G. Scorrano, Acid-base properties of organic solvents. *Journal of the American Chemical Society* **110**, 4577-4582 (1988).
41. M. J. Kamlet, J. L. M. Abboud, M. H. Abraham, R. W. Taft, Linear solvation energy relationships. 23. A comprehensive collection of the solvatochromic parameters, .pi.*, .alpha., and .beta., and some methods for simplifying the generalized solvatochromic equation. *The Journal of Organic Chemistry* **48**, 2877-2887 (1983).
42. J. E. Sutton, D. G. Vlachos, A Theoretical and Computational Analysis of Linear Free Energy Relations for the Estimation of Activation Energies. *ACS Catalysis* **2**, 1624-1634 (2012).
43. S. Wang *et al.*, Universal transition state scaling relations for (de)hydrogenation over transition metals. *Physical Chemistry Chemical Physics* **13**, 20760-20765 (2011).
44. X. Qian, Mechanisms and Energetics for Acid Catalyzed β-d-Glucose Conversion to 5-Hydroxymethylfurfurl. *The Journal of Physical Chemistry A* **115**, 11740-11748 (2011).
45. D. Martin, A. Weise, H.-J. Niclas, The Solvent Dimethyl Sulfoxide. *Angewandte Chemie International Edition in English* **6**, 318-334 (1967).
46. J. P. Richard, T. L. Amyes, M. M. Toteva, Formation and Stability of Carbocations and Carbanions in Water and Intrinsic Barriers to Their Reactions. *Accounts of Chemical Research* **34**, 981-988 (2001).

47. G. Rasul, G. K. S. Prakash, G. A. Olah, Protonated and Methylated Dimethyl Sulfoxide Cations and Dications. DFT/GIAO-MP2 NMR Studies and Comparison with Experimental Data1. *The Journal of Organic Chemistry* **65**, 8786-8789 (2000).
48. F. Franks, D. J. G. Ives, The structural properties of alcohol–water mixtures. *Quarterly Reviews, Chemical Society* **20**, 1-44 (1966).
49. Z.-L. Song, C.-A. Fan, Y.-Q. Tu, Semipinacol Rearrangement in Natural Product Synthesis. *Chemical Reviews* **111**, 7523-7556 (2011).

Chapter 7. Rational Design of Mixed Solvent Environments for Acid-Catalyzed Biomass Conversion Processes.[5]

7.1. Introduction

Toward developing a framework to effect the rational design of solvent compositions, in Chapters 5 and 6, we investigated the mechanistic role of mixed solvents systems in mediating rates and selectivities of liquid-phase, acid-catalyzed hydrolysis and dehydration reactions for several biomass-derived model compounds. We found that inclusion of polar aprotic cosolvents in mixtures with water leads to the formation of *water-enriched local domains* around hydrophilic reactants, which draws the acid catalyst to these regions due to preferential catalyst-water interactions(*1*) and results in enhanced reaction rates compared to pure water(Figure 7.1.1). We then generalized these insights into a predictive framework that allows for the rates of acid-catalyzed reactions of biomass-derived oxygenates to be estimated as a function of the water content of the solvent systems using computationally efficient classical molecular dynamics (MD) simulations and machine learning tools.(*1-3*) We also demonstrated how MD simulations can also estimate the selectivity of reactions occurring in parallel by computing the stability, or solvation free energies, between the reactant and product states as a function of solvent composition.(*4*)

In this Chapter, we develop a generalized workflow that combines these MD tools for the down selection of mixed solvent systems in acid-catalyzed biomass conversion processes, using minimal experimentation and computationally efficient methods. We demonstrate the usage of this

[5] This chapter was adapted from a manuscript in prep by the same name. T. W. Walker and G. W. Huber conceived the work. T. W. Walker and A. K. Chew designed the combined experimental and computational solvent screening workflow described herein. T. W. Walker and N. Frelka carried out the reaction kinetics studies. A. K. Chew designed and carried out the MD simulations, machine learning experiments and solvation free energy calculations. T. W. Walker and A. K. Chew wrote the manuscript. T. W. Walker, A. K. Chew, Reid C. Van Lehn, J. A. Dumesic and G. W. Huber edited the manuscript.

workflow by analyzing two case studies: the acid-catalyzed dehydration of cyclohexanol to cyclohexene, and the partial dehydration of fructose to 5-hydroxymethylfurfural.

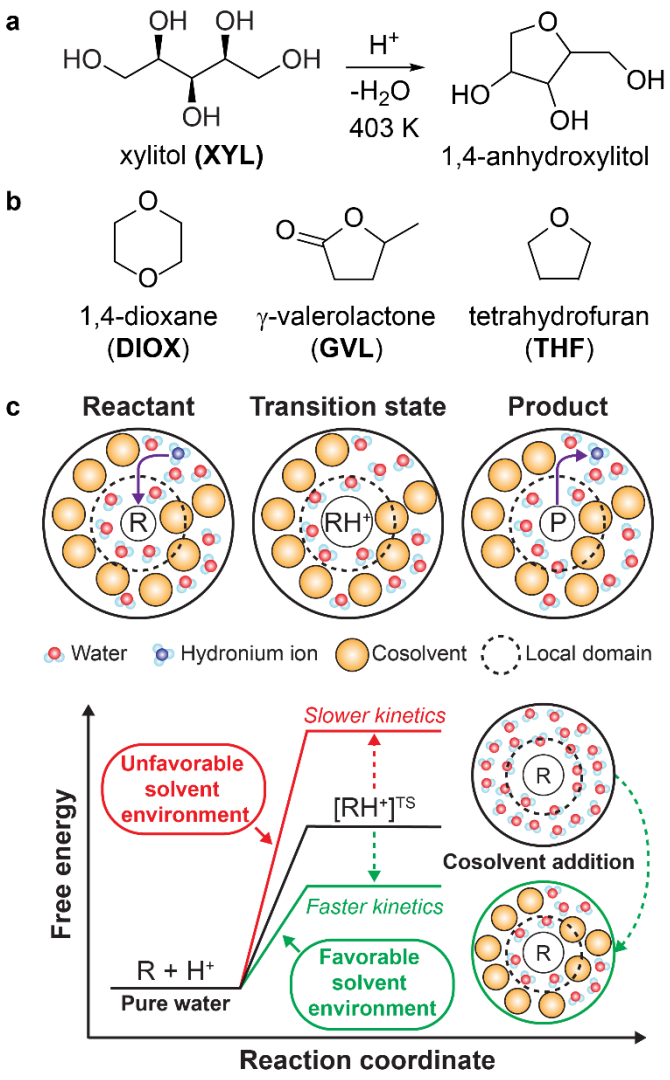


Figure 7.1.1. Overview of acid-catalyzed reactions in mixed-solvents for biomass conversion. (**a**) Acid-catalyzed reaction example of xylitol (XYL) dehydration to afford 1-4-anhydroxylitol. (**b**) Example of three organic, polar aprotic cosolvents. (**c**) Schematic of acid-catalyzed reactions in mixed-solvent environment that proceeds through a charged transition state (TS), formed from the protonation of the reactant by a hydronium ion catalyst. Hypothesized effect of mixed-solvent

environments on the free energy landscape of acid-catalyzed reactions. The schematic illustrates the formation of a local solvent domain around the reactant in a mixed-solvent environment that modifies the reaction free energy landscape, thus affecting reaction kinetics. These images were reproduced with permission of (*1-4*)

7.2 Materials, methods and definitions

7.2.1. Materials

Solvents (tetrahydrofuran, 99% with 200 ppm BHT; gamma-valerolactone, 98%+; acetone, 99% anhydrous; dioxane, 99%+ with 200 ppm BHT; and dimethyl sulfoxide, 98%+) were obtained from vendors and used as received. Fructose, cyclohexane, phenol, 5-hydroxymethylfurfural, levulinic acid, cyclohexanol and cyclohexene (all reagent grade) were obtained from Sigma and used as received. Triflic acid (99%) was obtained from Acros Organics and used as received.

7.2.2. Reaction kinetics experiments

Reaction kinetics studies of fructose dehydration, 5-hydroxymethylfurfural hydrolysis, and cyclohexanol dehydration were carried out in glass reactors. Briefly, an appropriate amount of reactant material (*e.g.,* 1 wt% fructose) were dissolved in the desired solvent system (*e.g.*, 90% tetrahydrofuran with 10 wt% water) containing an appropriate amount of dissolved acid catalyst (*e.g.*, 0.1 M). This reaction mixture was then charged into closed, 10 mL thick-walled glass reactors equipped with magnetic stir bars, which were then submerged in an oil bath at the desired reaction temperature (*e.g.*, 130°C). Reactors were agitated using a magnetic stir plate rotated at 500 rpm. Reactors were then removed at intervals corresponding to the desired reaction time points, and quenched in an ice bath at 0°C.

The contents of the reactors were filtered with a 0.2 micron PTFE syringe filter, and analyzed using high-performance liquid chromatographs (HPLC, for fructose, 5-hydroxymethylfurfual, and levulinic acid) or gas-chromatographs (GC, for cyclohexanol and cyclohexene). Analytes were quantified using calibration curves using known external standards for. The gas chromatograph was a Shimadzu GC-2010 equipped with a flame ionization detector and a RTX-VMS column. The HPLC was a Shimadzu 1020 series equipped with a refractive index detector and a photodiode array. The column was a proton-based Aminex HP87-X ion exclusion column with a mobile phase of 0.005 M sulfuric acid in water at a flow rate of 0.6 mL/min.

Triflic acid was used as an acid catalyst in all experiments, as it has been shown to fully dissociate even in organic solvents, which allows kinetic parameters (*rate constants*) to be estimated by normalizing the experimentally determined reaction rate on an accurate per-proton basis using Equation 7.1. r_i is the rate at which the reactant i is consumed in units of $mol\ L^{-1}\ s^{-1}$, C_i is the molar concentration of the i^{th} species, C_{H+} is the molar concentrations of excess protons in solution (assumed to be equal to the molar concentration of dissolved triflic acid in these experiments), and k_i is the rate constant associated with this reaction step at a fixed solvent composition and temperature.

$$r_i = -\frac{dC_i}{dt} = k_i C_i C_{H^+} \tag{7.1}$$

The rate constant associated with formation of the corresponding product (j) from reactant (i) is then shown in Equation 7.2.

$$r_j = \frac{dC_j}{dt} = k_j C_i C_{H^+} \tag{7.2}$$

Herein, we report rate constants measured in the kinetic regime, so that subsequent conversion of the product (j) can be neglected, and the selectivity (S_j) of the reaction going from the reactant (i)

to the product (*j*) is the ratio of initial rates or, equivalently, the ratio of rate constants as shown in Equation 7.3.

$$S_j = \frac{r_j}{r_i} = \frac{k_j}{k_i} \tag{7.3}$$

Conversion (*X*) of species (*i*) is defined in Equation 7.4.

$$X_i = 1 - \frac{C_i^{final}}{C_i^{initial}} \tag{7.4}$$

Finally, to facilitate comparison of reactivities across multiple mixed solvent systems, we quantify solvent effects on acid-catalyzed reaction rates in terms of a kinetic solvent parameter (σ), which we have defined in our previous work as Equation 7.5, where k_i^{mixed} is the rate constant that describes the conversion rate for reactant (*i)* in the cosolvent-water mixture, and $k_i^{H_2O}$ is the rate constant for the same reaction in pure water, at the same temperature and pressure. Positive σ values indicate an increase the rate of the corresponding reaction step in the mixed solvent system compared to pure water; negative values have a converse implication.

$$\sigma_i = \log\left(\frac{k_i^{mixed}}{k_i^{H_2O}}\right) \tag{7.5}$$

7.2.4. Classical molecular dynamics simulations

All classical MD simulations were performed using the Gromacs version 2016 software (*5*). Reactant, products, and cosolvents were parameterized using the CGenFF/CHARMM36 force fields (*6-8*). Water was modeled using the Single Point Charged/Extended (SPC/E) model (*9*). A $(6\ \text{nm})^3$ simulation box was initialized with water and cosolvent molecules at the desired

composition. The system was equilibrated in a *NPT* simulation for 5 ns at T = 300 K and P = 1 bar with a velocity-rescale thermostat and Berendsen barostat. Then, a single reactant molecule was added to the system and equilibrated with the same barostat and thermostat for 500 ps. 4 ns *NPT* productions were performed at the reaction temperature and P = 1 bar using a Nóse-Hoover thermostat and Parrinello-Rahman barostat. The 4 ns trajectory was partitioned into two 2 ns trajectories. Each partition was used to generate voxel representations as described in Ref. (*3*) and summarized here. For each simulation configuration, a 3D histogram was generated with the system centered on the center-of-mass of the reactant. The histogram covered a $(4\ \text{nm})^3$ volume that was divided into a 20×20×20 grid of bins corresponding to $(0.2\ \text{nm})^3$ volume elements. For each bin, normalized occurrences of reactant, cosolvent, and water molecules were stored in three separate channels to obtain a 20×20×20×3 grid of voxels for a single MD configuration. These grid values were averaged using 2 ns of simulation data, equivalent to 200 MD configurations, to generate a single voxel representation that captures the spatial distribution of atoms within the system. Voxel representations were then inputted into a pre-trained 3D convolutional neural network called SolventNet, as described in Ref. (*3*), which outputs the predicted kinetic solvent parameters. Each 2 ns partition was treated as two independent trials, where the average predicted kinetic solvent parameter is reported and the standard deviation of the predictions is the error.

Solvation free energies were computed from a series of stochastic dynamics simulations following the procedure described in Ref. (*1*) and (*4*). Starting from an equilibrated solvent system, a reactant or product was added to the system. The total potential of the system was defined as a function of Lennard-Jones (λ_{LJ}) and electrostatic (λ_{elec}) potential parameters, shown in Equation 7.6:

$$U(\lambda_{LJ}, \lambda_{elec}) = U_{M,solv}^{LJ}(\lambda_{LJ}) + U_{M,solv}^{elec}(\lambda_{elec}) + U_{M}^{bonded} + U_{M}^{nonbonded} + U_{solv}^{bonded} + U_{solv}^{nonbonded} \quad (7.6)$$

where $U_{M,solv}^{LJ}$ and $U_{M,solv}^{elec}$ are the LJ and electrostatic potentials between solute and solvent, U_{M}^{bonded} and $U_{M}^{nonbonded}$ are intramolecular bonded and non-bonded potentials of the solute, and U_{solv}^{bonded} and $U_{solv}^{nonbonded}$ are the bonded and non-bonded potentials between all solvent molecules (*10*). λ_{LJ} and λ_{elec} vary between 0 and 1 and modify the magnitude of the associated interactions; values of 0 correspond to a species in the vapor phase with no interactions with the solvent environment, while values of 1 correspond to a species in the liquid phase with complete interactions with the solvent environment. Seventeen independent simulations were performed for each free energy: fourteen in which electrostatic potential parameters were turned off ($\lambda_{elec} = 0$) and the VDW potential parameters were varied ($\lambda_{LJ} = 0.00$, 0.00922, 0.04794, 0.11505, 0.260634, 0.31608, 0.43738, 0.56262, 0.68392, 0.79366, 0.88495, 0.95206, 0.99078, or 1.00), three in which VDW potential parameters were on ($\lambda_{LJ} = 1.00$) and electrostatic potential parameters were varied $\lambda_{elec} =$ 0.25, 0.75, or 1.00. For each simulation, the system was energy minimized with the steepest descent algorithm and equilibrated for 100 ps at constant temperature followed by 2 ns at constant temperature and constant pressure using the Berendsen barostat. A 5 ns production simulation at constant reaction temperature and pressure (P = 1 bar) was then performed with the Parrinello-Rahman barostat. Energy differences computed between all pairs of windows were collected every 0.2 ps and solvation free energies were computed with the Multistate Bennett Acceptance Ratio (*11*) method using the python alchemical-analysis tool (*12*).

All simulations were performed using a leapfrog integrator with a 2-fs timestep. Verlet lists were generated using a 1.2 nm neighbor list cutoff. Van der Waals interactions were modeled with a Lennard-Jones potential that was smoothly shifted to zero between 1.0 nm and 1.2 nm.

Electrostatic interactions were calculated using the smooth Particle Mesh Ewald method with a short-range cutoff of 1.2 nm, grid spacing of 0.12 nm, and 4^{th} order interpolation. Bonds were constrained using the LINCS algorithm. All thermostats used a 1.0 ps time constant and all barostats used a 5.0 ps time constant with an isothermal compressibility of 5.0×10^{-5} bar^{-1}.

7.3. Computational design tools and a general procedure for screening mixed solvent systems for biomass conversion processes.

Liquid-phase, acid-catalyzed transformations of biomass-derived materials are often characterized by a complex network of reaction steps occurring both in series and in parallel (*13*). In these cases, the general challenge in effecting the selective transformation of the raw material into a single, desired product is to selectively enhance the rates of the desirable reaction steps over the undesirable ones. In principle, this challenge can be addressed by modulating the properties of the liquid solvent (*14, 15*), which is connected to the free-energy landscape that controls the kinetics of each reaction step through a series of non-covalent solvent-solute interactions (solvation energies). However, few predictive frameworks exist to anticipate these effects, and this is particularly true for mixtures of water with organic cosolvents, wherein the relevant mechanistic details are myriad, complex, and not well understood.

Herein, we propose a general process to screen mixed water-cosolvent systems for acid-catalyzed biomass conversion reactions using a synthesis of the experimental and computational approaches described in our prior work, which can reduce the overall experimental burden associated with designing new solvent systems using empirical screening methods alone. This process is outlined in Figure 2 and is composed of four key steps:

(1) Establishing the reaction network in a reference solvent system and pre-selecting an initial library of possible cosolvent-water systems;

(2) Using MD and machine-learning-based tools to screen the initial library of candidate water-cosolvent systems, differentiating those that promote the rates of the desired reaction steps;

(3) Using solvation free energy calculations to estimate the thermodynamic selectivity preference for a desired product in mixed solvent systems, and;

(4) Experimental validation of model-predicted, best-performing mixed solvent systems.

We discuss these four generalized steps in more detail, and then demonstrate the procedure outlined in Figure 7.3.1 by analyzing two case studies: (1) the conversion of cyclohexanol to cyclohexene and (2) the conversion of fructose (FRU) to 5-hydroxymethylfurfural (HMF).

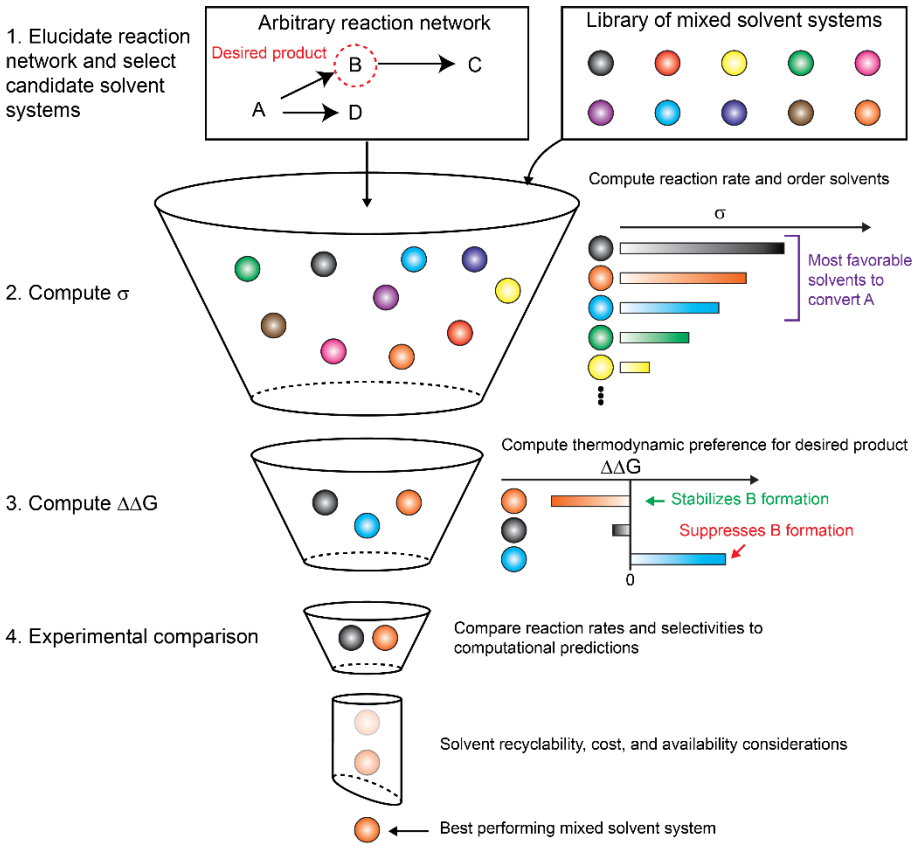


Figure 7.3.1. Process to screen candidate solvent compositions for biomass conversion processes. An arbitrary reaction network for a desired product is drawn from experiments; as an example, product B is desired from reactant A. Polar aprotic cosolvents are selected to mix with water and test the effects of solvent composition on reaction performance. Kinetic solvent parameters (σ) are predicted using molecular dynamics simulations in conjunction with SolventNet, described in Figure 7.3.2, below. The top performing solvents are then tested to see if the reaction selectively forms product B by calculating relative solvation free energies (ΔΔG).

Negative values of ΔΔG indicates that product B is more stabilized in the mixed-solvent environment. The top mixed-solvent environment for the conversion of A and production of B is then selected for experimental testing. Design considerations, such as solvent recyclability, cost, and availability are then used to select the best performing mixed-solvent environment for a specific process.

7.3.1. Step 1: Establish reaction network and pre-select candidate solvent systems

The first step in the solvent design process is to fully elucidate the reaction network underlying the desired chemical transformations in a reference solvent system. An obligatory reference solvent would be pure water, owing to its affordability and ease of handling. However, the reference solvent may contain an organic cosolvent; the minimum criterion is that the reactants and products are soluble at the desired concentrations. For example, cyclohexanol is soluble in pure water, but a minimum amount of organic cosolvent (*e.g.*, ~75 wt% THF) is required to solubilize its dehydration product (cyclohexene). In such cases, this minimum, threshold organic cosolvent content could dictate the reference state.

The reaction network underlying the desired transformation should include all discernable reaction steps between the reactant and products. Accordingly, kinetic parameters at a fixed reaction temperature (rate constants as described in Equations 7.1 and 7.2) should be obtained for all known steps in the reaction network. It should be noted that by "reaction steps" we do not mean *elementary steps,* but rather transformations between a single pair of quantifiable reactants and products. This reaction kinetics model is constituted by rate laws that describe the rates of each reaction step as a function of temperature and reactant concentrations (*16, 17*), and forms the basis for the MD-enabled screening process described below.

Finally, once the reaction kinetics model is formulated in a reference solvent system, a library of candidate cosolvent-water systems is proposed. Candidate cosolvent-water systems at this stage are selected based on simple design criteria such as: compatibility with the anticipated reaction conditions (*e.g.*, thermal stability at the anticipated reaction temperature); miscibility with water; toxicity limits; affordability or availability, and so on. Multiple cosolvent-water compositions may be considered for each organic cosolvent, or a base-case cosolvent composition can be specified to expedite the MD-screening process outlined in the next step. If a minimum amount of organic cosolvent is required in the reference solvent system, then this cosolvent-water composition can be used. If pure water is used as a reference solvent system then, as above, the base-case cosolvent-water composition can be fixed based on the solubility limits of the reactants and products. For example, fructose is highly soluble in pure water, but is only minimally soluble in solvent systems containing more than 90 wt% of common organic solvents (*e.g.* THF).

7.3.2. Step 2: Compute σ to screen mixed solvent systems for improved reactivity

Once the reaction network and library of candidate cosolvent-water systems have been determined, MD and machine learning based tools are used to sort cosolvent-water mixtures based on which are anticipated to best enhance the rate of the desired reaction steps to the greatest extent. In recent work, we have established methods to analyze solvent effects on the rates of acid-catalyzed biomass conversion reactions in cosolvent-water mixtures using classical MD simulations. For each mixed solvent system of interest, one reactant molecule is added, and simulated at the desired reaction temperature and pressure in the *NPT* ensemble. We analyzed the simulation-derived atomic positions of the reactant, water, and cosolvent molecules for each solvent composition, and correlated them to experimentally determined kinetic solvent parameters.

We found that in the presence of organic solvents, water localizes around hydrophilic, oxygen-rich biomass compounds, which drives the acid catalyst to the local domain of the reactant from preferred catalyst-water interactions and results in enhanced reaction rates.(*1*) We have shown that multiple MD observables that describe the localization of water or cosolvent around the reactant can be combined into a regression model to predict reaction rates.(*2*) These models are analogous to linear free energy relationships, in that they are informed by first-principles insights and are broadly applicable, but they are essentially empirical in nature.

We improved on these MD-derived reaction rate predictions by developing a machine learning environment to capture the complex relationship between the spatial configurations of reactant, water, and cosolvent molecules and experimentally determined kinetic solvent parameters (σ), outlined in Figure 3a,b. Of the different machine learning tools (*e.g.* support vector machines, deep neural networks), convolutional neural networks (CNNs) have been found to be the best performer in identifying spatial patterns within 2D images. Thus, we first converted atomic positions from MD simulations into an input representation that could be used for a CNN, called voxel representations. These representations are normalized 3D histograms of water, cosolvent, and reactant atomic positions that are stored in red, blue, and green channels (Figure 7.3.2a). We then input the voxel representations into a 3D CNN, called SolventNet (Figure 7.3.2b){A. K. Chew, 2019 #39}, which is trained on experimental kinetic solvent parameters for 7 biomass-relevant reactants (ethyl *tert*-butyl ether, *tert*-butanol, levoglucosan, 1,2-propandiol, cellobiose, FRU, and XYL), 3 cosolvents (Figure 7.1.1b), and 4 cosolvent mass fractions (25, 50, 75, 90 wt%). We found that SolventNet outperformed other machine learning techniques and human-engineered MD observables in predicting reaction rates. SolventNet also only requires 2 ns production simulations to make a prediction, requiring less than an hour in a supercomputing environment for

a single reactant-solvent combination. Therefore, we use SolventNet to rapidly screen through mixed solvent systems and predict solvent-induced changes in reaction rates, as expressed by σ, for biomass-derived oxygenates as a function of solvent system.

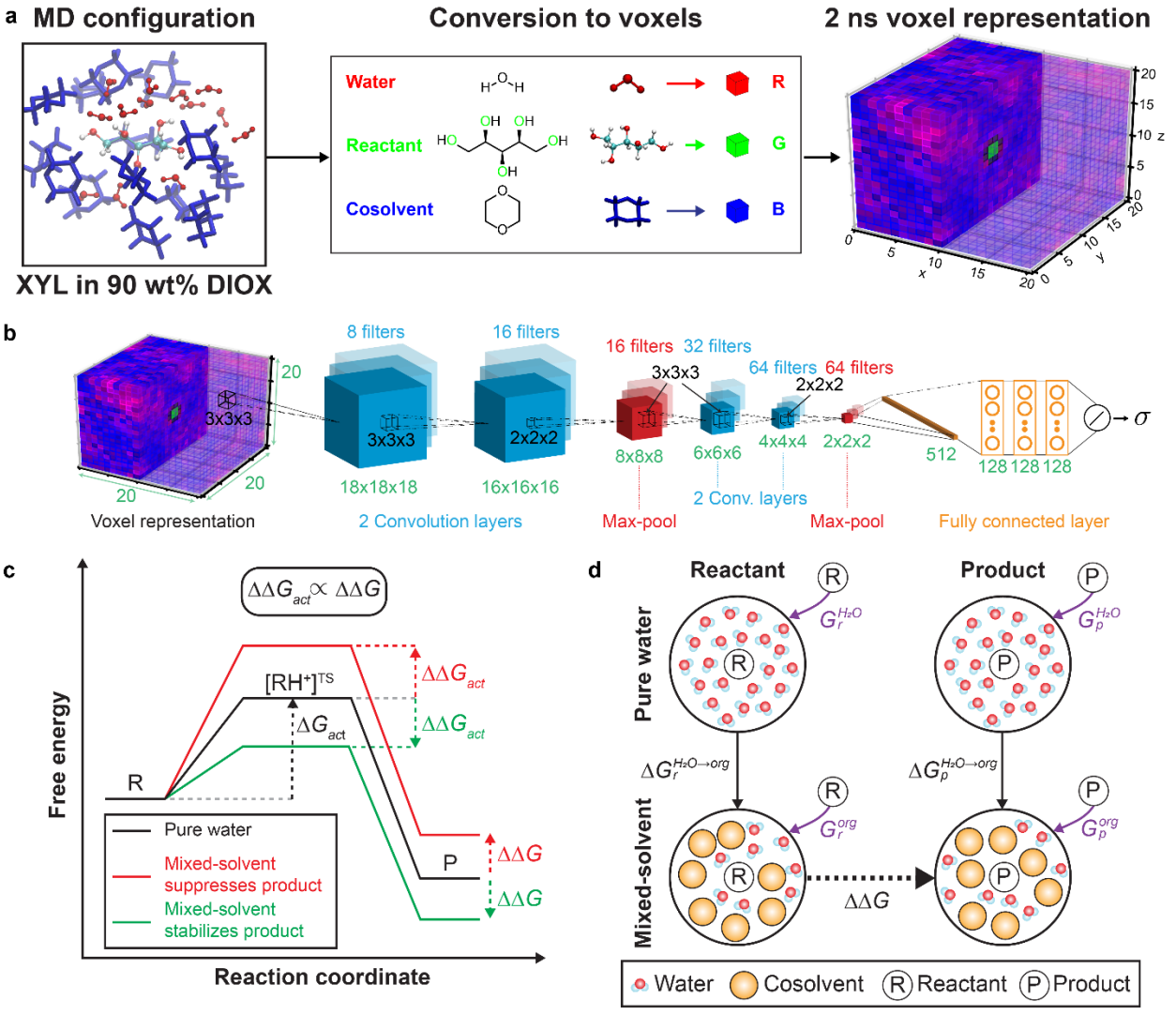


Figure 7.3.2. Computational tools used to predict reaction rates and selectivities. (a) Conversion of atomic positions obtained from molecular dynamics simulation trajectories into a voxel representation using XYL in 90 wt% DIO as an example. For each MD configuration, a 20×20×20 grid of $(0.2\ \text{nm})^3$ volume elements was centered on the reactant. Voxel representations are visualized by showing the water channel in red, the reactant channel in green, and the cosolvent

channel in blue. Half of the voxels are transparent to illustrate the solvent distribution around the reactant. (**b**) Architecture of SolventNet, a 3D CNN that inputs voxel representations and outputs the predicted kinetic solvent parameter (σ). (**c**) Hypothesized effect of mixed-solvent environments on the free energy landscape of reactant, transition, and product states. The change in the relative free energy between the reactant and product states ($\Delta\Delta G$) is proportional to the change in the activation energy ($\Delta\Delta G_{act}$) for a reaction in a mixed-solvent environment compared to the same reaction in pure water. The free energies are drawn relative to the reactant state in pure water. (**d**) Thermodynamic cycle to calculate the free energy difference between a reactant and product in a mixed-solvent environment relative to pure water. Purple arrows indicate solvation free energies computed from MD simulations which are used to calculate the transfer free energies indicated by filled black arrows. The dashed black arrow indicates $\Delta\Delta G$. These images were reproduced with permission of (*3, 4*).

7.3.3. Step 3: Compute $\Delta\Delta G$ to estimate selectivities within reaction networks

SolventNet, as described above, affords rapid predictions of forward reaction rates as a function of solvent compositions. Importantly, however, SolvetnNet does not directly probe selectivity toward a specific product. Accordingly, we have developed a separate framework using solvation free energy calculations to estimate the selectivity of parallel reactions as a function of solvent composition, outlined in Figure 7.3.2c,d. These perturbation-based free energy calculations are computationally more demanding than the aforementioned machine-learning tools, which limits their applicability as a high-throughput screening tool *vis-à-vis* SolventNet. However, once the larger library of candidate water-cosolvent systems has been prescreened using SolventNet to differentiated those which may best the promote overall reactivity of a particular

reactant molecule, solvation free energy calculations can be used to interrogate a manageable subset of anticipated "best performers" for selectivity towards a desired product. These methods are described briefly, now.

Based on *ab initio* molecular dynamics studies(*18*) we hypothesized that the modulation of solvent composition could affect the reaction energy barriers towards specific products, and we estimate changes in these energy barriers by measuring the relative stability of the reactant and product states in mixed solvent systems (Figure 7.3.2c). We define the stability of reactant and product states with solvation free energy calculations, which measures the free energy associated with transferring a reactant or product molecule from vacuum to a solvent environment. We quantify the free energy difference between reactant (r) and product (p) in mixed solvent environments ($\Delta\Delta G$) by performing four solvation free energies that related by Equation 7 and shown in Figure 3d.

$$\Delta\Delta G = \Delta G_p^{H_2O \to org} - \Delta G_r^{H_2O \to org} = \left(G_P^{org} - G_R^{org}\right) - \left(G_P^{H_2O} - G_R^{H_2O}\right) \quad (7)$$

If $\Delta\Delta G < 0$, the free energy difference between product and reactant state is more negative in mixed solvent systems compared to pure water, indicating that the product state is stabilized to a greater degree than the reactant. By evoking standard postulates, we expect that $\Delta\Delta G$ values capture the suppression or enhancement of a main dehydration product without having to model the reaction mechanism or the catalyst explicitly. As mentioned above, since $\Delta\Delta G$ values take a significantly longer time to compute, requiring ~85 ns of simulation time (~12 hours on a supercomputer) for a single reactant-solvent combination, initial screening with SolventNet is necessary to lower the necessary number of $\Delta\Delta G$ calculations.

7.3.4. Step 4. Probe selected solvent systems using experiments

The result of the MD-enabled solvent screening process outlined in Steps 2 and 3 above is the full list of original candidate water-cosolvent systems, with corresponding, SolventNet-generated predictions rate enhancements for each step in the reaction mechanism as compared to the same reactions in pure water. An additional output is a subset of the candidate solvent systems, selected based on the SovlentNet-generated predictions of reactivity, with corresponding, solvation-free-energy-enabled predictions as to which cosolvent systems may best enhance the selectivity of each reaction step toward a desired product. In principle, these ordered subsets of the original candidate cosolvent systems can then be ranked in descending order with respect to their predicted abilities to enhance the rate and selectivity of the desired reaction step.

This ordered list of candidate solvent systems is then used to guide experiments toward confirming best performers, as determined by those solvent systems which maximize selectivity toward a desired product. Once the most selective solvent system has been confirmed by experimentation, other reaction conditions (such as temperature or reactant concentration) can be varied to check for the effect of these parameters on reaction performance. If desired, this optimized set of reaction conditions (temperature, reactant concentration *etc*.) can be used as a *new* reference solvent system, and the solvent selection process can be repeated from step one, forming a feedback loop to further improve reaction performance or map out elements of the process-conditions-vs-performance space for techno-economically-driven design purposes.

7.4. Case Studies

7.4.1. Case study 1: cyclohexanol dehydration to cyclohexene

The Brønsted-acid-catalyzed dehydration of cyclohexanol affords cyclohexene, as displayed in Figure 7.4.1a. This reaction proceeds *via* a known sequence of elementary steps, so that some of the factors which control the reactivity of cyclohexanol are understood from first principles.(*19, 20*) However, the reactive, carbocation-like intermediates underlying this reaction mechanism readily participate in a series of side reactions, so that the selectivity to cyclohexanol is often less than 100% (*21*). Together, these factors make cyclohexanol dehydration of general interest as a probe reaction to understand the factors which control liquid-phase, acid catalyzed biomass conversion reactions (*22*). Accordingly, we now demonstrate how the tools described above can be used to anticipate the effects of mixed solvent systems in modulating the selectivity of cyclohexanol dehydration to afford cyclohexene.

Cyclohexanol and cyclohexene are both converted to unaccountable degradation products (humins) in the presence of an acid catalysts, but for simplicity only the formation of cyclohexene is considered here. The rate constant associated with this reaction was obtained in pure water at 160°C and is in agreement with the values reported elsewhere (*19, 21*). Water was chosen as a reference state solvent system for this example, though it should be noted that the product (cyclohexene), is insoluble in water. In THF-, DIO-, dimethyl sulfoxide (DMSO)-, GVL-, and acetone (ACE)-water mixtures, a minumum amount of ~75 wt% organic cosolvent was added to water to achieve complete mixing of a 1 wt% cyclohexene solution. Accrdingly, to facilitate the compuational solvent screening process, the mass fraction of the organic phase for each candidate cosolvent-water mixture is fixed at 75 wt% of the organic phase, as a lower limit. We consider 75

wt% mixtures of ten common organic cosolvents, which are displayed in Figure 7.4.14b. Note that the abbreviations in Figure 7.4.1b will now be used throughout.

Following the computation screening steps outlined in Figure 7.1.1, we first sort this initial library of ten cosolvent-water systems by SolventNet-predicted σ values for cyclohexanol dehydration as shown in Figure 7.4.1c. The universally negative σ values indicate that the reactivity of cyclohexanol is suppressed in all mixed solvent systems compared to pure water. Despite the suppression of reactivity, a minimum amount of organic cosolvent is necessary to facilitate the solubilization of cyclohexene. In this context, the solvent systems which might best promote the reactivity of cyclohexanol are therefore those with the *least negative* corresponding values of σ. Following this criterion, THF, GVL and ACE are anticipated to best promote the reactivity of cyclohexanol, whereas DMSO and DIOX are anticipated to suppress the reactivity of cyclohexanol.

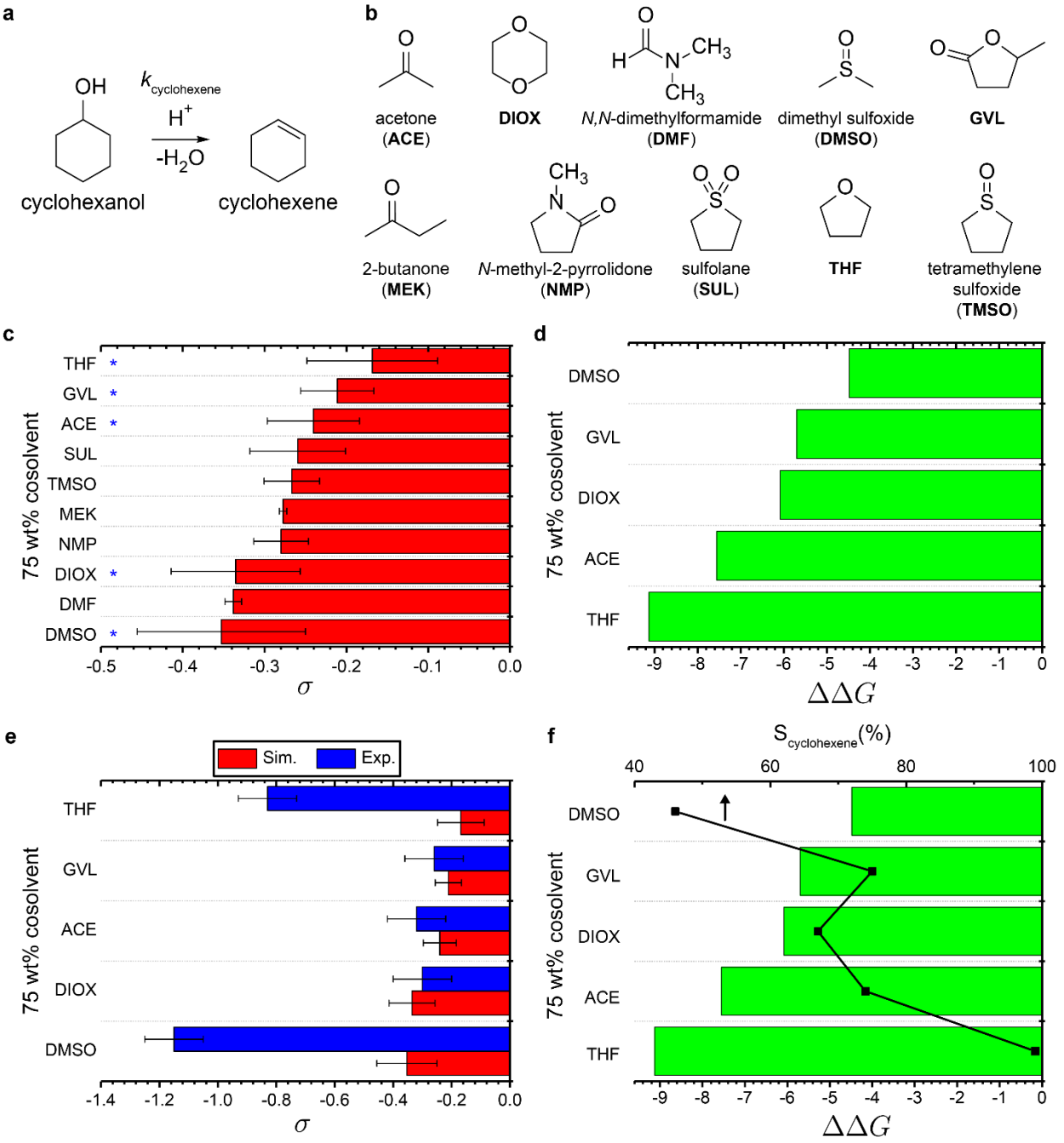


Figure 7.4.1. Case study of cyclohexanol conversion to cyclohexene. (**a**) Acid-catalyzed conversion of cyclohexanol to cyclohexene. (**b**) Ten organic, polar aprotic cosolvents considered as the as the initial cosolvent library. (**c**) Kinetic solvent parameters predicted by SolventNet for 75 wt% organic cosolvents. Blue asterisks indicate solvent systems that are representative of good cosolvents (THF, GVL, and ACE), and poor cosolvents (DIOX and DMSO). (**d**) Relative solvation

free energies ($\Delta\Delta G$) between product (cyclohexene) and reactant state (cyclohexanol) in 75 wt% organic cosolvents relative to pure water. (**e**) Comparison between kinetic solvent parameters predicted from SolventNet (red) and experimentally determined (blue). (**f**) Comparison between $\Delta\Delta G$ and experimental percent selectivity towards cyclohexene.

As discussed above, SolventNet-predicted changes in reactivity (as expressed by σ values) apply only to the *overall reactivity* of the cyclohexanol, and cannot explicitly distinguish between changes in *selecitivty* towards a desired proudct (*e.g.*, cyclohexene). To anticipate which mixed solvent systems promote reactivity toward a desired product, we therefoer interrogate a subset of these ordered cosolvent-water systems using the more compoutationally demanding solvation free energy calculations, as described above. Note that in practice we would prescribe analyzing solvation free energies for only those solvent systesms that are predicted to most enhance the overall reactivity of cyclohexanol, as expressed by σ. In this example, a subset of candiate sovlent systems that meets this criterion would be THF, ACE and GVL. Herein, however, we select DMSO, DIO, GVL, ACE and THF, which span the full range of SolventNet-predicted σ values, to test whether trends in overall reactiivty correlate with trends in selectivity for a given reaciton step.

Figure 7.4.1d shows solvent-enduced reaction free energy changes (ΔΔG's) between cyclohexanol and cyclohexene in 75 wt% DMSO, DIO, GVL, ACE, and THF. Negative ΔΔG values are obtained for all five mixed solvent systems, indicating that the product state is stabilized relative to the reactant state in all mixed solvent systems relative to pure water. Importantly, 75 wt% THF has the most negative ΔΔG, implying that this solvent mixture might effect the highest selectivity to cyclohexene, whereas 75 wt% DMSO might effect the lowest selectivity. Note that

the ordering of the solvents based on σ and ΔΔG values (Figures 4b and 4c, respectively) are not the exactly the same. This is generally expected, since σ only considers the conversion of the reactant without any consideration of the product, whereas ΔΔG considers the stability of a specific product state relative to the reactant state. Despite these differences, however, σ and ΔΔG both predict that 75 wt% THF will best facilitate the selective conversion of cyclohexanol into cyclohexene, whereas 75 wt% DMSO would be the worst-performing solvent mixture in this regard.

Finally, Figures 7.4.1e and 4f compare the predicted and experimentally determined values for the rate and selectivity of cyclohexanol conversion to cyclohexene, respectively, for 75 wt% DMSO, DIOX, GVL, ACE, and THF. As shown in Figure 7.4.1e, the experimental rates of cyclohexanol conversion agree, within error, with those predicted by SolventNet in only two out of five cases. This behavior is likely attributable to the fact that SolventNet was trained on datasets that did not include rate values below $\sigma = -0.5$, which is a limitation that could be addressed by training the machine learning network with an expanded set of reactants. Despite this quantitative disagreement, however, SolventNet *does* capture the consistently negative σ values confirmed by experiments. Futhermore, SovlentNet captures the realistic behaviors whereby 75 wt% DMSO supresses the reactivity of cyclohexanol to the greatest extent, while GVL and DIOX suppress the reactivity of cyclohexanol to a lesser extent.

Comparing the ΔΔG values in the selected solvent systems to the experimentally deterimined reaction selectivities, we see that the solvation free energy calculations are able to capture the key behaviors that 75 wt% THF produces cyclohexene with the greatest selectivity, while 75 wt% DMSO produces cyclohexene least seelctivity. In practice then, and without experimental data to guide the solvent design process, the compuational tools described herein

would have identified THF-water mixtures as “good” candidate solvent systems to facilitate the selective dehydration of cyclohexanol into cyclohexene, and this prediction would then be confirmed with a single experiment. Together, these results demonstrate how the MD- and machine-learning based tools can be used to guide the solvent selection process for acid-catalyzed dehdration reactions, reducing the experimental burden of emprically screening a large library possible cosolvent-water combinations.

7.4.2. Case study 2: fructose dehydration to 5-hydroxymethylfurfral

HMF is an important platform molecule derived from biomass, which is produced by the partial dehydration of FRU over Bronsted-acid catalysts (*23*). However, in a pure aqueous solvent, HMF yields are limited by: (1) the formation of humins (*24*), which are produced in parallel with HMF *via* the acid-catalyzed polymerization of FRU, and; (2) the hydrolysis of HMF to form stoichiometric amounts of formic and levulinic acids (LA) (*25*). As a result, the highest yields of HMF obtained from FRU in pure water are about 40% at 100% fructose conversion (*26*). Exhaustive experimental efforts over the past decade have identified a number of solvent systems that will facilitate this reaction in higher yields than pure water (*27-29*). Important among these alternative solvent systems are mixtures of water with DMSO (*30, 31*). Following the generalized solvent design procedure outlined above, we now demonstrate how DMSO-water mixtures could be readily identified from among a subset of other common industrial solvents as a promising candidate to facilitate this important biomass conversion reaction in high yield.

Figure 7.4.2a shows a simplified reaction scheme describing the acid-catalyzed conversion of FRU to HMF, and subsequent conversion of HMF into LA. FRU and HMF are both converted to humins, but for simplicity only the formations of HMF and LA are considered here. The rate constants associated with each step in the reaction scheme were obtained in pure water at 130°C,

and are in agreement with the values reported elsewhere (*32*). Water was chosen as a reference state solvent system for this example, because the reactant and products are all water-soluble. As mentioned above, however, FRU is only minimally soluble in most cosolvent-water mixtures when the organic phase is present in mass fractions above 90 wt%. Therefore, to facilitate the computational solvent screening process, we focus on 90 wt% of the organic phase for each candidate mixed solvent system, as an upper limit.

Following the procedure outlined in Figure 2, we first analyze MD simuations of FRU and HMF in the 90 wt% cosolvent-water systems using SolventNet. In this example, since the intermediate dehydration product (HMF) is desired, we are interested in cosolvent-water systems that will selectively enhance the reactivity of FRU *over* HMF. Accordingly, we sort the candidate cosolvent-water systems based on the *differences* in the predicted kinetic solvent parameters between HMF and FRU as a function of solvent system ($\sigma_{HMF}-\sigma_{FRU}$). Note that more negative values of $\sigma_{HMF}-\sigma_{FRU}$ indicate that the rate of HMF production is ehanced to a greater extent than the rate of LA formation, and selectivity to HMF is therefore promoted. Figure 7.4.2b shows these SolventNet-predicted $\sigma_{HMF}-\sigma_{FRU}$ values across the same ten water-cosolvent systems used in the previous case study (Figure 7.4.1b), though in this example the organic phase is fixed at 90 wt% instead of 75 wt%. These SolventNet-predicted $\sigma_{HMF}-\sigma_{FRU}$ values indicate that 90 wt% GVL, TMSO and DMSO might best facilitate the selective conversion of FRU into HMF, by supressing the formation of LA, while 90 wt% ACE and THF are predicted to be the worst performing solvet mixtures in this regard. 90 wt% DIOX is predicted to perform in a capacity intermediate between these extrema. Therefore, as in the previous example, we select GVL, DMSO, DIOX, ACE and THF for further analysis using the solvation free energy calculations. We note again that, in practice, only those sovlents which SovlentNet predicts to most enhance the reactiivty of desired

reaciton step would be intergoated further using solvation free energy calculations, whereas herein we select solvent systems spanning the full range of predictied reactivities for validation purposes only.

Given that this is a series reaction, we again focus on suppression of the LA product. Accordingly, Figure 7.4.2d shows the ΔΔG between FRU and LA for 90 wt% GVL, DMSO, DIO, ACE and THF. Note that positive ΔΔG values correspond to the suppression of the undesired product (LA). Of the five cosolvent-water systems, we find that 90 wt% DMSO is the only cosolvent-water mixture that is anticipated to suppress the formation of LA, whereas all other cosolvent-water mixtures stabilize the formation of LA from FRU. Therefore, between the SolventNet-predicted $\sigma_{HMF}-\sigma_{FRU}$ values *and* the MD-estimated ΔΔG values, 90 wt% THF is predicted to effect the lowest selectivity towards HMF, whereas 90 wt% DMSO is predicted to effects the greatest selectivity towards HMF.

Figures 7.4.2d,e compare the predicted and experimentally determined values for the rates of FRU conversion, and the selectivity of FRU conversion toward HMF, respectively, in 90 wt% DMSO, GVL, ACE, DIOX, and THF. As in the example of cyclohexanol dehydration, the SolventNet-predicted and experimentally determined σ_{FRU} values not generally in quanitative agreement, but SolventNet does capture the experimetnally confirmed behavior that FRU reactivity is universally enhanced in all mixed sovlent systems tested. Furthermore, SovlenNet again captures the realistic behavior that THF enhances FRU reactivity to the least extent, while 90 wt% GVL and DMSO enhance FRU reactivity to the greatest extent. Finally, Figure 7.4.2e shows that the MD-derived ΔΔG values correctly predict the experimetnally-determined behaviors that 90 wt% THF facilitates the acid-catalaysed dehydration fo FRU to HMF with the lowest

selectivity of those sovlent tested, while DMSO facilitates this reaciton with near quanitative seletivity.

Together, these results demosntrate howfollowing the generalized protocol in Figure 7.1.1 could predict that 90 wt% DMSO is the preferred mixed solvent system to afford HMF from FRU with nearly 100% selectivity, in agreement with experimetnal findings from multiple reports. Notibly, however, we have found that 90 wt% ACE performs poorly in selectively converting FRU to HMF. This finding (confirmed *via* expeirmentation), is in constrast to recent reports that ACE-water mixtures are able to produce HMF from FRU in high yield. We rationallize this descpacy by noting that these contradictory reports utilized different water contents, temperatures, and acid catalysts than those invesigataed herein.(*27*) Therefore, the temperature and catalyst used clearly play an critical role that is not investigated in this work, wherein only the role of the solvent is considered. Accordinly, future work will focus on incorporating the effects of temperature and catalyst selection to the design of efficient reaction systems for biomass conversion processes.

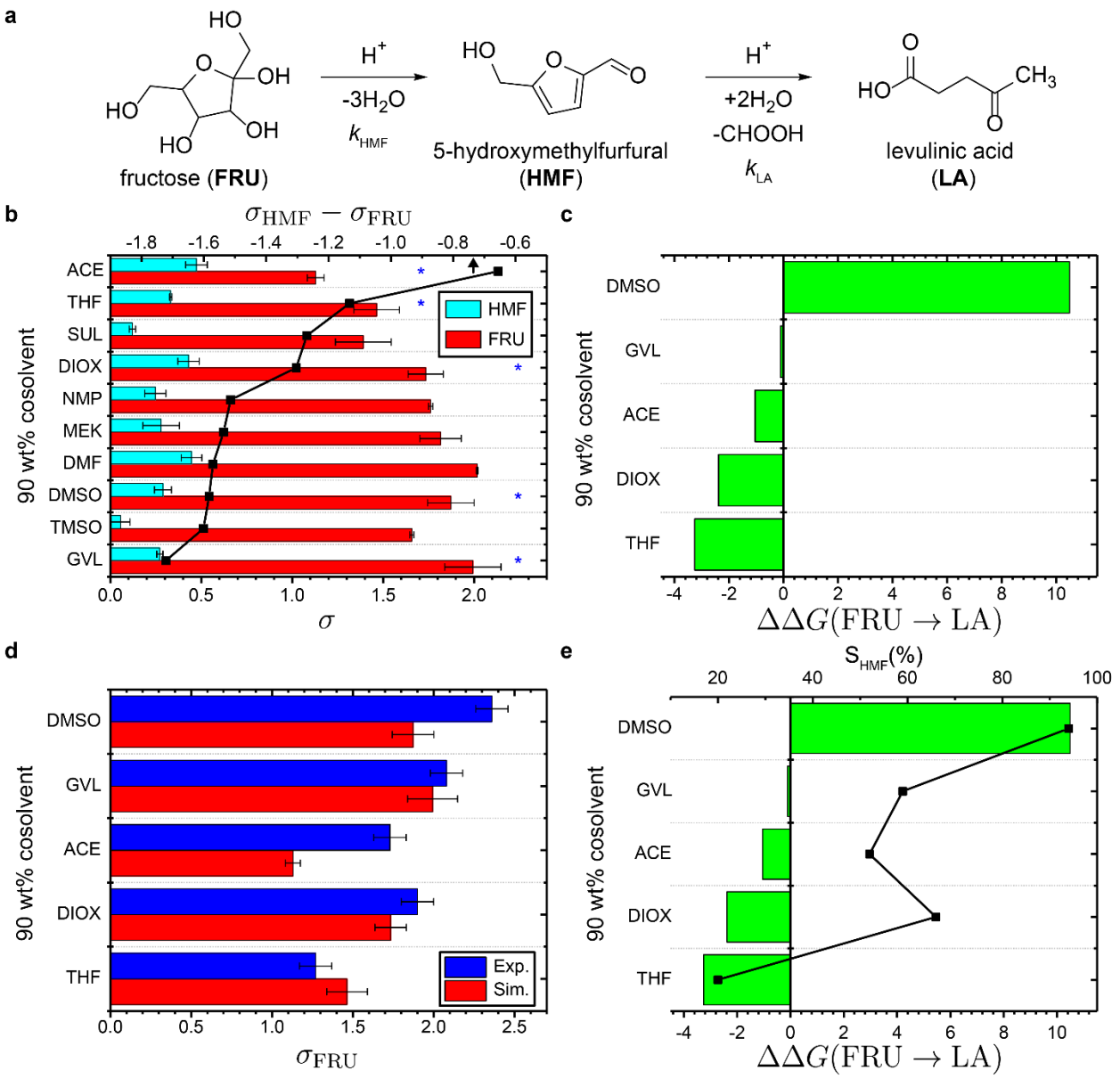


Figure 7.4.2. Case study of fructose conversion to HMF. (**a**) Acid-catalyzed dehydration of fructose (FRU) to afford 5-hydroxymethylfurfural (HMF), followed by an addition reaction to afford levulinic acid (LA). (**b**) Kinetic solvent parameters predicted by SolventNet for 90 wt% cosolvents. Blue asterisks indicate solvent systems that are representative of good cosolvents (THF, GVL, and ACE), and poor cosolvents (DIOX and DMSO). (**c**) Relative solvation free energies ($\Delta\Delta G$) between product (LA) and reactant state (FRU) in 90 wt% organic cosolvents

relative to pure water. (**d**) Comparison between kinetic solvent parameters predicted from SolventNet (red) and experimentally determined (blue). (**e**) Comparison between $\Delta\Delta G$ and percent selectivity towards HMF.

7.5. Conclusions

We have synthesized key, fundamental insights regarding the effects of water-cosolvent mixtures on the rates and selectivities of acid-catalyzed reactions into a workflow for the rational design of solvent systems for biomass conversion processes using computationally efficient methods and minimal experiments. To this end, MD simulations machine learning tools are used to estimate the influence of solvent composition on reactivity of biomass-derived molecules by analyzing the solvation environment around the reactant. Furthermore, MD simulations inform on the selectivity of biomass conversion reactions by quantifying the stability, or solvation free energies, of reactant and product states as a function of solvent composition. Together, these MD- and machine-learning-based tools are combined into a workflow for down selecting mixed solvent systems for biomass conversion applications with minimal experiments.

We have demonstrated the usage of this workflow by analyzing two case studies: the dehydration of cyclohexanol into cyclohexene and the acid-catalyzed partial dehydration of fructose to 5-hydroxymethylfurfural. In both case studies, the MD- and machine-learning based tools do not achieve quantitative agreement with experimentally determined reaction rates nor selectivities. However, they do predict key trends that allow for best-performing sovlent systems (as expressed by their ability to faciliate the desired reactions in high yield) to be down-selected from a large library of candiate sovlent systesm that would be laborous and difficult to investigate using experimetnal screening methods alone.

These results demonstrate that fundamental insights as to how solvent-solute interactions drive reactivity in liquid phase catalytic processes can be distilled into computationally efficient screening methods and predictive design tools that allow for the rational design of mixed solvent systems for biomass conversion applications. These efforts therefore represent a step toward a molecular-level understanding the role of mixed solvent systems in controlling reactivity of liquid phase reactions, and toward the rational design of mixed solvent systems in a framework that significantly reduces the experimental burden that typically accompanies the development of new chemical processes.

7.6. Acknowledgements

This work was supported in part by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Numbers DE-SC0018409. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1549562. This work also used the computing resources and assistance of the UW-Madison Center for High Throughput Computing (CHTC) in the Department of Computer Sciences. The CHTC is supported by UW-Madison, the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation, the Wisconsin Institutes for Discovery, and the National Science Foundation, and is an active member of the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science. The authors acknowledge support from the Department of Chemical and Biological Engineering at the University of Wisconsin-Madison and the Wisconsin Alumni Research Fund.

7.7. References

1. A. K. Chew, R. C. Van Lehn, Quantifying the Stability of the Hydronium Ion in Organic Solvents With Molecular Dynamics Simulations. *Frontiers in chemistry* **7**, 439 (2019).
2. T. W. Walker *et al.*, Universal kinetic solvent effects in acid-catalyzed reactions of biomass-derived oxygenates. *Energy & Environmental Science* **11**, 617-628 (2018).
3. R. C. V. L. A. K. Chew, Machine tools predict kinetic solvent effects in acid-catalyzed biomass conversion reacitons *Nature communications* **(in press)**, (2019).
4. T. W. W. A. K. Chew, B. Demir, L. Witteman, J. Euclide, G. W. Huber, J.A. Dumesic, R. C. Van Lehn, "Effect of mixed-solvent environments on the selectivity of acid-catalyzed dehydration reactions". *ACS Catalysis* **(in press)**, (2019).
5. S. Páll, M. J. Abraham, C. Kutzner, B. Hess, E. Lindahl, in *Solving Software Challenges for Exascale: International Conference on Exascale Applications and Software, EASC 2014, Stockholm, Sweden, April 2-3, 2014, Revised Selected Papers,* S. Markidis, E. Laure, Eds. (Springer International Publishing, Cham, 2015), pp. 3-27.
6. R. B. Best *et al.*, Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling from the Backbone and Side Chain and Dihedral Angles. *Journal of Chemical Theory and Computation* **8**, 3257-3273 (2013).
7. K. Vanommeslaeghe *et al.*, CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *Journal of computational chemistry* **31**, 671-690 (2009).
8. W. Yu, X. He, K. Vanommeslaeghe, A. D. MacKernell Jr, Extension of the CHARMM General Force Field to Sulfonyl- Containing Compounds and Its Utility in Biomolecular Simulations. *J. Comput. Chem.* **33**, 2451-2468 (2012).
9. H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma, The missing term in effective pair potentials. *The Journal of Physical Chemistry* **91**, 6269-6271 (1987).
10. D. Shivakumar *et al.*, Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *Journal of Chemical Theory and Computation* **6**, 1509-1519 (2010).
11. M. R. Shirts, J. D. Chodera, Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **129**, (2008).
12. P. V. Klimovich, M. R. Shirts, D. L. Mobley, Guidelines for the analysis of free energy calculations. *J Comput Aid Mol Des* **29**, 397-411 (2015).
13. P. Daoutidis, W. A. Marvin, S. Rangarajan, A. I. Torres, Engineering biomass conversion processes: a systems perspective. *AIChE Journal* **59**, 3-18 (2013).
14. R. J. Madon, E. Iglesia, Catalytic reaction rates in thermodynamically non-ideal systems. *Journal of Molecular Catalysis A: Chemical* **163**, 189-204 (2000).
15. M. A. Mellmer *et al.*, Solvent effects in acid-catalyzed biomass conversion reactions. *Angewandte chemie international edition* **53**, 11872-11875 (2014).
16. S. H. Krishna, T. W. Walker, J. A. Dumesic, G. W. Huber, Kinetics of levoglucosenone isomerization. *ChemSusChem* **10**, 129-138 (2017).
17. M. J. Lindsay, T. W. Walker, J. A. Dumesic, S. A. Rankin, G. W. Huber, Production of monosaccharides and whey protein from acid whey waste streams in the dairy industry. *Green chemistry* **20**, 1824-1834 (2018).
18. M. A. Mellmer *et al.*, Solvent-enabled control of reactivity for liquid-phase reactions of biomass-derived compounds. *Nature Catalysis* **1**, 199-207 (2018).

19. Y. Liu *et al.*, Enhancing the catalytic activity of hydronium ions through constrained environments. *Nature communications* **8**, 14113 (2017).
20. D. Mei, J. A. Lercher, Effects of Local Water Concentrations on Cyclohexanol Dehydration in H-BEA Zeolite. *The Journal of Physical Chemistry C*, (2019).
21. J. He, C. Zhao, J. A. Lercher, Impact of solvent for individual steps of phenol hydrodeoxygenation with Pd/C and HZSM-5 as catalysts. *Journal of catalysis* **309**, 362-375 (2014).
22. X. Wang, R. Rinaldi, A route for lignin and bio-oil conversion: dehydroxylation of phenols into arenes by catalytic tandem reactions. *Angewandte Chemie International Edition* **52**, 11499-11503 (2013).
23. R.-J. Van Putten *et al.*, Hydroxymethylfurfural, a versatile platform chemical made from renewable resources. *Chemical reviews* **113**, 1499-1597 (2013).
24. S. K. Patil, C. R. Lund, Formation and growth of humins via aldol addition and condensation during acid-catalyzed conversion of 5-hydroxymethylfurfural. *Energy & Fuels* **25**, 4745-4755 (2011).
25. F. S. Asghari, H. Yoshida, Kinetics of the decomposition of fructose catalyzed by hydrochloric acid in subcritical water: formation of 5-hydroxymethylfurfural, levulinic, and formic acids. *Industrial & Engineering Chemistry Research* **46**, 7703-7710 (2007).
26. S. P. Teong, G. Yi, Y. Zhang, Hydroxymethylfurfural production from bioresources: past, present and future. *Green Chemistry* **16**, 2015-2026 (2014).
27. A. H. Motagamwala, K. Huang, C. T. Maravelias, J. A. Dumesic, Solvent system for effective near-term production of hydroxymethylfurfural (HMF) with potential for long-term process improvement. *Energy & Environmental Science*, (2019).
28. Y. Román-Leshkov, J. N. Chheda, J. A. Dumesic, Phase modifiers promote efficient production of hydroxymethylfurfural from fructose. *Science* **312**, 1933-1937 (2006).
29. X. Qi, M. Watanabe, T. M. Aida, R. L. Smith Jr, Selective conversion of D-fructose to 5-hydroxymethylfurfural by ion-exchange resin in acetone/dimethyl sulfoxide solvent mixtures. *Industrial & Engineering Chemistry Research* **47**, 9234-9239 (2008).
30. G. Tsilomelekis, T. R. Josephson, V. Nikolakis, S. Caratzoulas, Origin of 5-hydroxymethylfurfural stability in water/dimethyl sulfoxide mixtures. *ChemSusChem* **7**, 117-126 (2014).
31. S. H. Mushrif, S. Caratzoulas, D. G. Vlachos, Understanding solvent effects in the selective conversion of fructose to 5-hydroxymethyl-furfural: a molecular dynamics investigation. *Physical Chemistry Chemical Physics* **14**, 2637-2644 (2012).
32. R. Weingarten, J. Cho, R. Xing, W. C. Conner Jr, G. W. Huber, Kinetics and reaction engineering of levulinic acid production from aqueous glucose solutions. *ChemSusChem* **5**, 1280-1290 (2012).

Chapter 8. Solid-State NMR Studies of solvent-mediated woody biomass decomposition over sulfuric acid and implications for enzymatic conversion of residual cellulose[6]

8.1. Introduction

Lignocellulosic plant matter (*biomass*) is a heterogeneous, polymeric material composed mainly of xylan (hemicellulose), glucan (cellulose) and lignin, along with entrained water, minerals, proteins and extractives.(*1-3*) Of these constituent fractions, cellulose is present in both amorphous phases, and relatively crystalline phases constituted by parallel glucan chains, which are arrayed into ordered domains (*4*) with characteristic lateral dimensions (microfibrils).(*5, 6*) The microfibrils are bound to the less-ordered (amorphous) cellulose, hemicellulose and lignin matrices *via* covalent and non-covalent interactions to form the cell wall.(*7, 8*) These structures are represented schematically in Figure 8.1.1.

As an abundant and naturally occurring source of carbon, deconstruction of biomass to yield platform chemicals is a promising strategy to partially displace global demand for fossil resources.(*9-11*) To this end, several technologies have been proposed, including enzymatic conversion(*12, 13*) and solvent-based methods.(*14-16*) Of these two examples, as shown in Chapters 3-7, solvent-assisted conversion of biomass over acid catalysts is characterized by rapid catalytic turnovers, but suffers from poor selectivity and often requires elevated temperatures (>= 160ºC). Hydrolytic enzymes will selectively hydrolyze cellulose and hemicellulose into

[6] This Chapter was adapted from manuscript in prep by the same name. T. W. Walker, J. A. Dumesic, B. G. Fox, N. Kuch and K. Vander Meulen conceived the work. T. Walker carried out the biomass pretreatment experiments, FTIR and XRD characterizations. T. W. Walker and C. Clewett designed and carried out the NMR experiments. N. Kuch and K. Vander Meulen designed and carried out the enzymatic conversion studies. T. W. Walker wrote the manuscript. All authors edited the manuscript.

corresponding sugars at mild temperatures, but these processes are limited by high operating costs and slow catalytic turnovers.(*17, 18*)

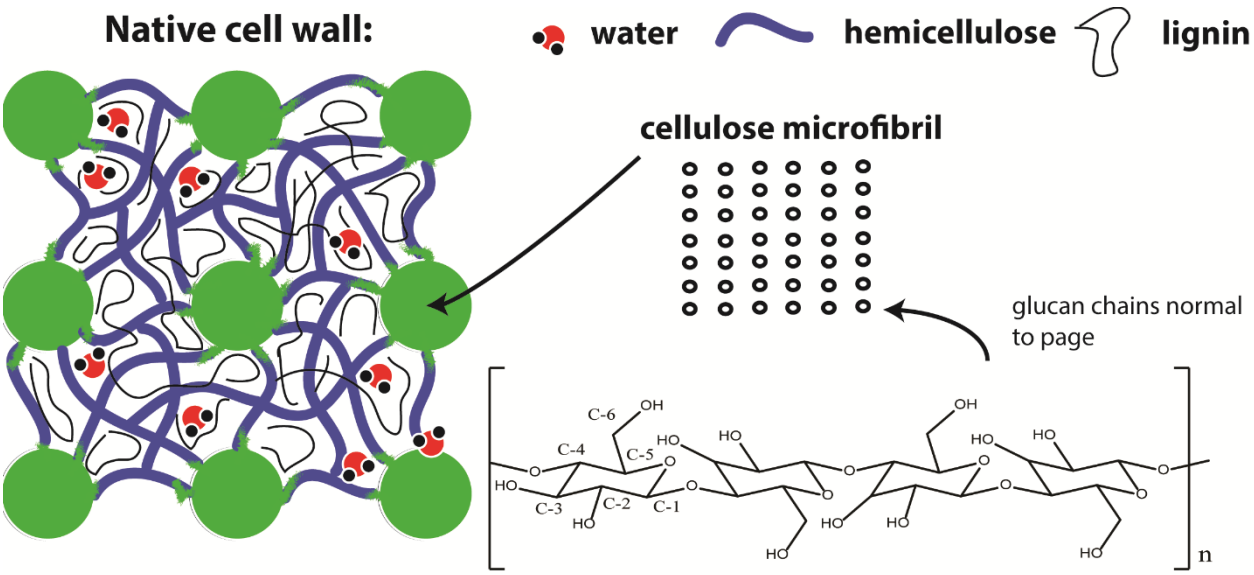


Figure 8.1.1. Schematic representation of the cellulose-, hemicellulose- and lignin containing structures present in the cell wall of woody biomass. Cellulose microfibrils are semi-crystalline in nature, and are constituted by parallel chains of glucan arrayed in ordered planes (represented here by an arbitrary *m* x *n* matrix, with ellipses denoting single chains whose long axes are oriented normal to the page). The six carbon centers present in the repeating units of cellulose are labeled according to IUPAC convention.

One strategy to address these challenges is to combine enzymatic hydrolysis of biomass with a solvent-mediated pretreatment step,(*19*) whereby the solvent (usually containing a dilute acid catalyst) first dissolves the hemicellulose and lignin fractions, leaving behind a residual, cellulose-rich solid that is more easily degraded by hydrolytic enzymes than the native cell wall.(*20*) However, the details connecting the physical and chemical changes effected by the

solvent to the improved yield of the enzymatic hydrolysis step are not fully elucidated. For example, it is proposed that the solvent alters the cellulose pore structure,(*21*) and/or the crystallinity of cellulose;(*22*) these changes hypothetically improve the accessibility of the glyosidic linkages to enzymes, or the affinity of enzymes for the surface accessible facets of the cellulose, respectively. However, attempts to verify these hypotheses have often proven inconclusive,(*23*) or yielded contradictory results.(*24, 25*) As such, few bases exist to compare biomass pretreatment strategies in a context that can predict their effectiveness (as determined by improved sugar yields in the enzymatic hydrolysis step), and optimizing these tandem pretreatment-and-enzymatic-conversion methods remains an empirical challenge.

Herein, we present a strategy to probe the physical and chemical properties of crystalline cellulose present in raw and solvent-pretreated biomass using a combination of high resolution magic-angle-spinning (HR/MAS) and cross-polarization (CP)/MAS nuclear magnetic resonance spectroscopy (NMR), infrared spectroscopy (FTIR) and X-ray diffraction (XRD) techniques. Proton spin-relaxation-edited (PSRE) CP/MAS NMR is used to differentiate ^{13}C signals originating from the crystalline and amorphous portions of cellulose, which overlap in non PSRE NMR spectra. We show that, as hemicellulose and lignin are removed by the solvent (observable by NMR and FTIR), the characteristic dimensions of the remaining crystalline cellulose microfibrils (observable by XRD) are unchanged. In contrast, as the lignin and hemicellulose are removed, the magnetic environment at the exterior of the *crystalline* cellulose microfibrils (observable by ^{13}C PSRE CP/MAS NMR) does change. The latter result indicates that the cohesive forces within the crystalline cellulose microfibrils are comparable to those bonding the microfibrils to the hemicellulose and lignin matrices in the plant wall, so that the presence of hemicellulose and lignin likely prevent cellulose-specific enzymes from binding to the surface-accessible

glycosidic linkages, inhibiting cellulose hydrolysis. ^{1}H HR/MAS NMR reveals that, as the residual cellulose from pretreated-biomass is dried, distinct hydration environments form within the pore structure, and water binds more strongly with the surface accessible facets of the pore walls, likely hindering the diffusion of cellulases within the pore structure, further hindering enzymatic conversion.

We combine these characterizations with enzymatic reactivity studies to show that the NMR-observable reductions in cellulose-xylan interactions in solvent-pretreated biomass correlate with the yield of sugars afforded by hydrolysis of the residual cellulose by engineered cellulases. Based on these observations, we present a model to predict the yield of enzymatic sugars from solvent-pretreated woody biomass by tracking changes in solid-state CP/MAS NMR-observable metrics. The techniques developed herein therefore serve as a diagnostic to monitor the solvent-mediated decomposition of biomass, and to compare and predict the effectiveness of different solvent-based biomass pretreatment strategies for enzymatic conversion of biomass to platform sugars.

8.2. Experimental

8.2.1. Materials

Corn stover, switchgrass, sorghum, and poplars (NM6 and P39) were obtained from the Great Lakes Bioenergy Research Center (GLBRC) in the form of debarked, 5 mm chips, and used as received. Water (Fisher, HPLC grade), acetone (Sigma, 98%), ethanol (Fisher, HPLC grade, 200 proof), γ-valerolactone (GVL; Sigma, 95%), tetrahydrofuran (THF; Acros, 99%+ anhydrous with 200 ppm BHT inhibitor), acetonitrile (MeCN; Sigma, 98%), dimethyl sulfoxide (DMSO; Sigma, 98%), and *N*-methyl pyrrolidinone (NMP; Sigma, 99%) were obtained from vendors and used as received. Sulfuric acid (98 wt%), NaOH and acetic acid were obtained from Sigma and

used as received. β-*D*-glucose, xylose, furfural, and 5-hydroxymethylfurfural (Sigma-Aldrich, ACS reagent grade) were obtained from vendors, and used as calibration standards.

CelR_cbm3a was created through fusion of the Hungateiclostridium thermocellum CelR enzyme domain (GenBank accession code CAE51308.1 residues 31 – 642, i.e., GH9 and CBM3c modules only) with an additional 153-residue CBM3a from the scaffolding protein CipA, connected via the 48-residue CipA-native linker sequence. The lac operon-controlled expression plasmid was shipped to a third party, where it was expressed and purified at scale.

8.2.2 Sample prep and solvent-mediated pretreatment of biomass over sulfuric acid

All reaction conditions can be found in Table 8.3.1; we will describe the sample prep using entry 9. For a typical experiment, 6 grams of as-received 5-mm biomass chips (P39 poplar) were loaded into a 60- mL pressure-sealed glass reactor equipped with a magnetic stir bar. 60 grams of an appropriate solvent system (80 wt% GVL with 20 wt% water) containing 100 mM sulfuric acid were then added, and the glass reactor was sealed with a screw-top cap equipped with a Teflon O-ring. The glass reactor was then submerged in a silicon oil bath at the desired temperature (140º*C*). The reactor contents were agitated by rotating the stir bar at 500 rpm. The reactor was held at the desired temperature for a period of time (one hour), and then quenched in an ice bath at 0º*C* for 15 minutes. The reactor contents were then removed, and the liquid and solid fractions were separated by vacuum filtration with a fine-mesh filter paper, and a Buchner funnel connected to a house vacuum system at -15 in. Hg.

The solid fraction from the reactor was washed with 3 volumes of HPLC grade water, and dried with vacuum filtration to produce a solid filter cake of about 70 wt% moisture content. The

bleaching step, when applied, first consisted of stirring the solids at room temperature for three hours in a solvent system consisting of water with 1 wt% dissolved $NaClO_2$ and 0.33 mL acetic acid per gram of $NaClO_2$. The resulting solids were then separated with vacuum filtration, and stirred for an additional 4 hours at 50º*C* in a solvent system consisting of water with pH adjusted to 11 using NaOH (1.0 mM NaOH). The solids/liquids ratio in this both bleaching steps was fixed at 3/97 by weight. The resulting solids from the second bleaching step were then dried using vacuum filtration for one hour to produce a solid filter cake of about 60 wt% moisture content. When applied, a drying step consisted of placing the filtered solids from the solvent pretreatment or bleaching steps in a vacuum oven at 85º*C* and -15 in. Hg overnight, and taking the difference in mass from the dried and never-dried material as the moisture content of the sample.

The liquid fraction from the reactor was collected, diluted with water ten times by weight to precipitate the GVL-soluble lignin, and then vacuum filtered again to remove the solid lignin. The lignin was vacuum treated overnight at 85º*C*, and weighed to determine the lignin yield. The diluted, lignin-free liquid fraction was then filtered with a PTFE 0.2-micron syringe filter, and analyzed for soluble sugar and furan content using a Shimadzu-1020 HPLC equipped with a refractive index detector and a UV-vis scanning photo diode array. The mobile phase used was 5 mM sulfuric acid, with a flow rate of 0.6 mL/min at 30º*C*. The glucose, xylose, furfural and 5-hydroxymethyl furfural content of the sub-micron filtered, lignin-free liquids was then determined using calibration curves and external standards. The glucan and xylan oligomer content of the filtered, lignin-free liquids was then determined by adding 3 wt% H2SO4, stirring at 120º*C* for one hour, and then re-analyzing the samples using HPLC. The difference in glucose and xylose content between the hydrolyzed and non-hydrolyzed samples was taken to determine the soluble glucan and xyaln oligomer content, respectively. Note that the xylan and glucan oligomer yields,

plus the small quantity of sugar degradation products (furfural and 5-hydroxymethylfurfural) detected were added to the corresponding sugar yields (xylose and glucose, respectively) to determine total soluble carbohydrate yield.

Separately, the moisture content of the untreated biomass was determined by drying the as-received samples in a vacuum oven overnight, as described above. The extractives content (waxes, proteins, minerals, etc.) of the untreated biomasses was determined by drying the as-received biomasses overnight, and then processing the dried samples under reflux in a Solex extractor using water, then ethanol, and then acetone as extraction solvents, with each step being carried out at the boiling point of the corresponding extraction solvent. After correcting for the original moisture content, the wt% of total extractives was determined by taking the mass difference between the dried and extracted biomass samples.

8.2.3. ^{13}C solid-state cross-polarization magic-angle-spinning (CP/MAS) NMR

About 200 mg of solid sample was packed into 4-mm thin-walled silicon nitride rotors and sealed with glass-filled Torlon caps. ^{1}H/^{13}C PSRE CP/MAS NMR spectra were acquired on a Bruker Avance III 500 MHz spectrometer with a proton radio frequency of 500.22 MHz and a ^{13}C radio frequency of 125.76 MHz. Each rotor was spun at 4 kHz in a 4 mm Doty Scientific MAS probe. ^{1}H and ^{13}C 90º pulse lengths of 2.30 and 4.55 μs were used, respectively. Spectra shown represent 2056 signal averages with a 0.3 second acquisition time and an 8 second recycle delay between scans. The proton radiofrequency strength was $\gamma B_H/(2\pi)$ = 46 kHz during spin-locking with a spin-lock time of 3-10000 μs, as described below. The proton radiofrequency was 99.53 kHz (maximum) during cross-polarization with a 70-100% ramp and a contact time of 2000 μs.

The proton decoupler radio frequency was 100 kHz during acquisition. ^{13}C spectra were referenced to the up field adamantine peak at 28.7 ppm in separate standard.

Crystalline cellulose produces relatively sharp, well-defined ^{13}C CP/MAS-NMR spectral features. However, as described above, native biomass also contains a preponderance of amorphous materials mainly corresponding to lignin and hemicellulose, but also a fraction of the cellulose. These amorphous materials produce broad, asymmetric features that distort the baseline and overlap the resonances corresponding to crystalline cellulose. In order to facilitate quantitative analysis of the crystalline cellulose fraction by NMR, is necessary to remove the overlapping signals produced by these amorphous materials. To this end, we followed the method of proton spin-relaxation editing (PSRE) developed by Newman *et al*.(*26*) In this method, differences are exploited in the rotating-frame proton relaxation behavior between crystalline cellulose and the non-crystalline cellulose, lignin and hemicellulose matrices to separate the NMR signals corresponding to these two fractions.

Briefly, ^{13}C CP/MAS NMR spectra are collected with a spin-lock delay applied prior to cross polarization. During spin-locking, the ^{1}H net magnetic vector is first oriented along the transverse plane by application of a standard 90º pulse, and is then held in the xy plane by the sustained application of a rotating magnetic field for a fixed time (the spin-lock delay). As longer spin-lock delays are applied, magnetization decays in the ^{1}H rotating frame, so that less magnetization is transferred to the carbon nuclei during cross-polarization. The result is a number of CP/MAS spectra which, in order of increasing spin-lock delays, are attenuated to an increasing extent. Importantly, however, the proton radio frequency during spin-locking is optimized so as to effect the greatest contrast in rotating-frame relaxation time constants ($T_{1\rho}$*'s*) between crystalline cellulose and the amorphous materials in the sample. That is, protons associated with crystalline

materials loose magnetization in the xy plane either faster or slower than amorphous materials during spin-locking. Newman and co-workers have shown this contrast is optimized at a spin-lock frequency of 46 kHz, which is used herein, so that crystalline cellulose generally has about a 2-3 times longer $T_{1\rho}$ value than amorphous cellulose, hemicellulose or lignin.

Figure 8.2.1 displays two CP/MAS spectra for a cellulose-rich material isolated from P39 biomass (Entry 9 in Table 8.3.1) with spin-lock delays of 0.003 and 10.0 ms applied prior to cross-polarization. Major resonances corresponding to the carbon centers in crystalline cellulose (as per the labeling convention described in Figure 1) are denoted.(*27*) Note that no spinning side bands were observed under these conditions.(*28*) Also note that amorphous cellulose signals (broader features not immediately obvious upon visual inspection) overlap the larger, sharper resonances associated with crystalline cellulose.(*26*) Finally, note that the C-4 region of the CP/MAS cellulose spectrum is split into upfield and downfield resonances. This behavior is discussed later.

Two key resonances are denoted in Figure 8.2.1: the peak at 104.4 ppm corresponding to the C-1 carbon in crystalline cellulose, and the peak at 63.1 ppm corresponding to the methoxy groups in amorphous lignin. Comparing the short (0.003 ms) spin lock delay to the longer (10 ms) spin-lock delay, note that the crystalline cellulose signal at 104.4 ppm has lost about 25% of its intensity. In contrast, the signal from the methoxy groups (amorphous lignin) has lost about 60% of its intensity.

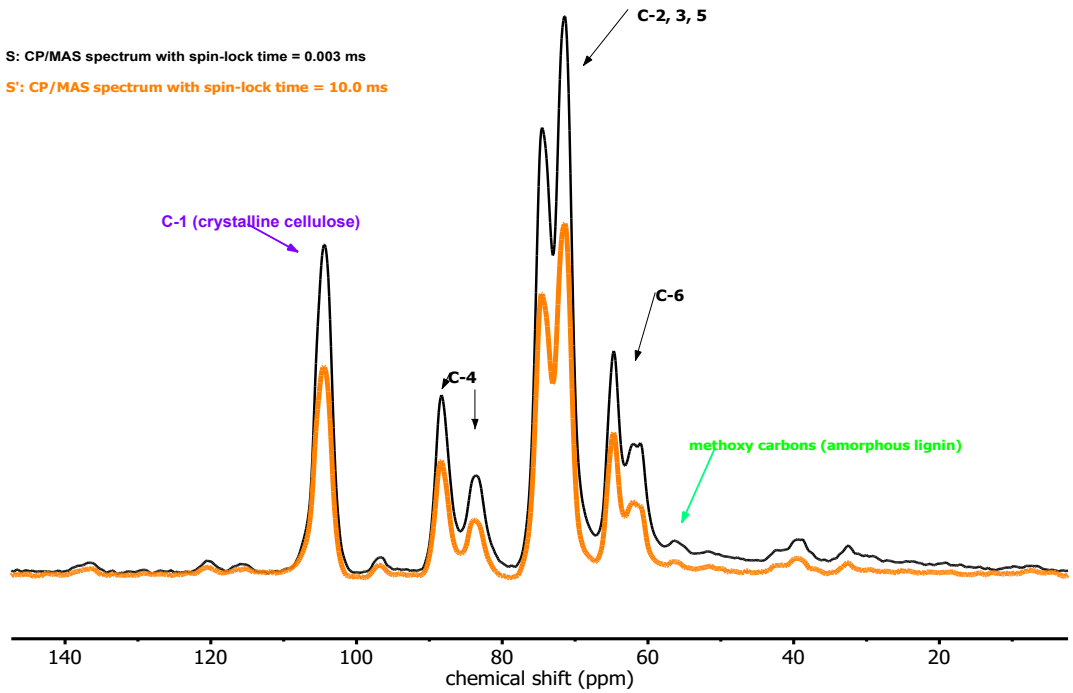


Figure 8.2.1. CP/MAS spectra corresponding to the cellulose-rich material obtained from P39 biomass as described in Entry 9 of Table 8.3.1. Spectra were collected with a spin-lock delay of 0.003 (black) and 10.0 ms (orange) prior to cross-polarization. Key resonances are denoted.

Figure 8.2.2 displays the integrated areas for the aforementioned resonances at 104.4 ppm (crystalline cellulose) and 63.1 ppm (amorphous lignin) in the cellulose-rich material described in Entry 9 of Table 1 as a function of the spin-lock delay prior to cross-polarization. Due to the fact that these resonances are reasonably isolated from hetero-materials in the chemical shift domain (*e.g.,* the crystalline cellulose C-1 peak at 104.4 has no resonances corresponding to amorphous material near it), these two integrals are described by a simple exponential decay with respect to spin-lock time. The time constant associated with each of these fitting functions is the rotating-frame relaxation constant ($T_{1,\rho}$) for the crystalline and amorphous phases in the solid biomass. As

expected, the $T_{1,\rho}$ value for crystalline cellulose (21.46 ms) is about twice as long as the $T_{1,\rho}$ value for the amorphous lignin (9.59 ms).

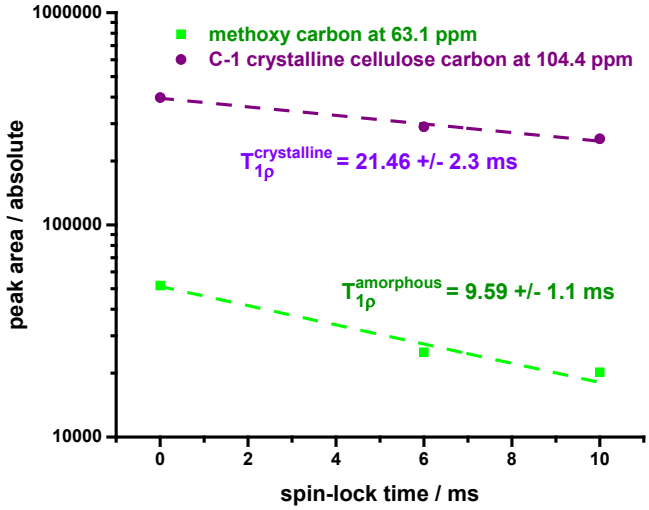


Figure 8.2.2. Absolute integrals for the peaks corresponding to the C-1 carbon in crystalline cellulose (104.4 ppm) and methoxy carbons in amorphous lignin (63.1 ppm) as a function of spin-lock time from the spectra displayed in Figure 8.2.1.

Finally, we produce a proton spin-relaxation-edited (PSRE) CP/MAS sub spectra that displays only ^{13}C signal corresponding to crystalline or amorphous materials in the biomass being analyzed. We do this by assuming that the total spectrum (***S***) obtained with a short (0.003) spin-lock delay is a linear combination of signal from the crystalline (***C***) and amorphous (***A***) materials in the biomass:

$$S = C + A \qquad (8.1)$$

Similarly, we assume that the attenuated spectrum (***S'***) obtained with a longer (*e.g.* 10 ms) spin-lock delay is a combination of signal from the crystalline and amorphous fractions in the biomass

sample, but that signal from each phase has been reduced by a fraction (f_C or f_A) by the application of the spin-lock delay:

$$S' = f_C C + f_A A \qquad (8.2)$$

f_C and f_A can be calculated as a function of spin-lock time (t_{SL}) using Figure 8.3.2, where:

$$f_C = e^{-\frac{t_{SL}}{T_{1\rho}^{crystalline}}} \qquad (8.3)$$

And

$$f_A = e^{-\frac{t_{SL}}{T_{1\rho}^{amorphous}}} \qquad (8.4)$$

These relationships can be expressed as a matrix operation:

$$\begin{bmatrix} 1 & 1 \\ f_C & f_A \end{bmatrix} \begin{bmatrix} C \\ A \end{bmatrix} = \begin{bmatrix} S \\ S' \end{bmatrix} \qquad (8.5)$$

And the vector (C, A) is solved for to reproduce the sub spectra corresponding to crystalline cellulose and the amorphous materials in the solid biomass sample. Figure 8.2.3 displays the results of this procedure, where the PSRE spectra shown are again associated with the cellulose-rich material described in Entry 9 of Table 8.3.1.

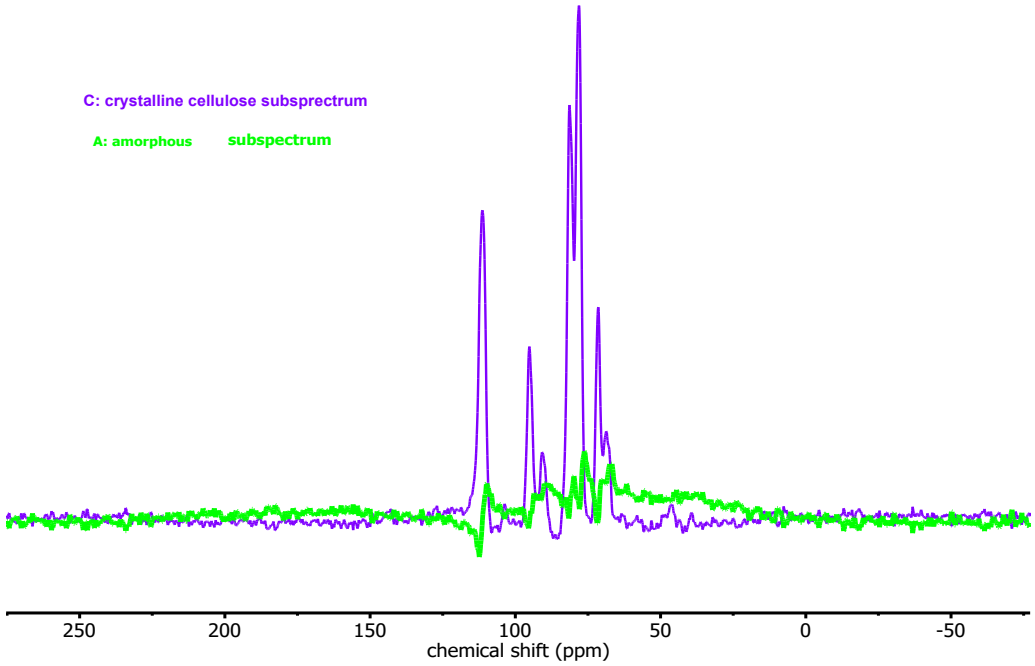


Figure 8.2.3. Proton spin-relaxation-edited (PSRE) CP/MAS sub spectra corresponding to crystalline cellulose (purple) and amorphous material (green) in the cellulose-rich material obtained from P39 biomass as described in Entry 9 of Table 8.31.

Figure 8.2.3 displays a chemical shift range from -75 to 275 ppm in order to show that, after the PSRE procedure described above, the resulting sub spectra each retain phasing and a baseline that goes to zero in the limits of the observed sweep width. Comparing the sub spectra for the crystalline cellulose (purple) and amorphous biomass material (green), note how the crystalline phase is characterized by sharp spectral features and a flat baseline. In contrast, the amorphous phase is characterized by broad features that are slightly offset compared to their crystalline counterparts. This behavior is characteristic is semi-crystalline polymeric materials. In particular, note the large amorphous signal that overlaps the C-4 region of the crystalline cellulose sub spectrum. This behavior, like aforementioned C-4 resonance spitting, will again be discussed later.

8.2.4. Attenuated total reflectance infrared (ATR-FTIR) spectroscopy

Attenuated total reflection–Fourier transform infrared spectroscopy (ATR–FTIR) was conducted using a Bruker Optics Vertex system with a diamond-germanium ATR single reflection crystal. Untreated and solvent-pretreated biomass samples were dried in a vacuum oven overnight to remove water content prior to analysis, and were pressed uniformly against the diamond surface using a spring-loaded anvil. Sample spectra were obtained in triplicates using an average of 128 scans over the range between 400 cm^{-1} and 4000 cm^{-1} with a spectral resolution of 2 cm^{-1}. Air was used as background.

8.2.5. X-ray diffraction characterizations of native and pretreated cellulose

Cellulose microfibrils scatter incident X-rays to produce a diffraction pattern consistent with a monoclinic unit cell.(*29*) Accordingly, X-ray diffractograms were collected for the native P39 and solvent-pretreated biomass samples, using a Bruker D8 Discovery diffractometer with a Cu K_a X-ray source operating at 1000 kV and 100 mA, with a 5-mm aperture and 600 sec exposure time. Biomass samples were analyzed without additional grinding to reduce particle size. Samples were self-adhered, when necessary, with vacuum grease. A separate sample of boron nitride was analyzed to assess the inherent line broadening of the instrument.

8.2.6. Enzymatic hydrolysis of residual cellulose

Hydrolysis reactions were prepared with 0.5 mg/mL pretreated cellulose and 0.05 mg/mL CelR in 0.1M phosphate buffer, pH 6.0 to a final volume of 1 mL. Reactions were incubated in a Heidolph Titramax 1000, with a Heidolph Inkubator 1000 used to control temperature. Reactions

were run for 24 hours at 50 ºC with 1050 rpm (0.0185 x g) shaking. After incubation, reactions were centrifuged for 5 minutes at 21,130 x g. The concentration of soluble sugar produced was determined via the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific). Briefly, 100 µL of working solution and 5 µL of supernatant were heated for 15 minutes at 80 ºC. Controls used were CelR only, cellulose only, and buffer only samples, and a glucose standard curve was used for converting absorbance units to mg/mL glucose in solution. Four reactions were performed for each condition, with supernatant being sampled in triplicate.

8.2.7. ^{1}H High-Resolution (HR)/MAS NMR spectroscopy

^{1}H HR/MAS NMR experiments were performed on a Bruker Avance spectrometer at 500 MHz using a 4 mm Doty Scientific MAS NMR probe. Solvent-pretreated biomass samples of about 70 wt% moisture content were sealed into 30 µL Kel-F HR/MAS rotor inserts (Bruker Biospin Inc.), with the MAS experiments performed at 4 kHz with a ±0.1 K temperature regulation. Dried, solvent-pretreated biomass samples were re-wetted by soaking in an amount of water corresponding to the moisture content of the never-dried samples (typically ~70 wt% water to 30 wt% solids) for 72 hours, and then loaded into 30 µL Kel-F HR/MAS rotor inserts. The sample temperatures under HR/MAS conditions (T_{MAS}) were calibrated using a neat ethylene glycol thermometer. The one-dimensional ^{1}H NMR spectra were obtained using a single pulse Bloch decay sequence with 16 scan averages and 10 seconds recycle delays. Transverse spin–spin relaxation times (T_2) were found using a rotor-synchronized CPMG (Carr–Purcell–Meiboom–Gill) pulse sequence with 16 scan averages, and inter-pulse delays varied between 4 and 20000 rotor cycles (corresponding to 0.002 to 10 seconds transverse spin-spin relaxation times). The ^{1}H NMR

chemical shifts were referenced to the internal standard of neat water, $\delta = +4.7$ ppm at 298 K with respect to the chemical shift of TMS, $\delta = 0$ ppm.

8.3. Results

8.3.1 Solvent-mediated pretreatment of biomass over sulfuric acid

Figure 8.3.1 displays a three-step pretreatment process based on prior work by our own group and others.(*10, 16, 30, 31*) Generally, this process consists of:

(1) A low temperature (60-120ºC) ***lignin removal step***, wherein raw biomass is treated in solvent system consisting of mostly of an organic solvent, with some water and a dilute mineral acid catalyst.
(2) A higher temperature (100-160ºC) ***sugar removal step,*** wherein the residual solids from Step (1) are treated in a solvent system, again consisting of an organic solvent and dilute acid catalyst, but with more water than Step (1), and;
(3) A ***bleaching step***, wherein trace impurities are removed from the remaining solids to yield a pristine cellulose product.

The purpose of the first step is to remove as large a portion of the lignin as possible in a form that closely resembles the native lignin in the cell wall; higher temperatures and solvents systems containing more water will alter the chemical nature of the extracted lignin.(*32-34*) The first step also removes part of the hemicellulose in the form of soluble five-carbon sugar monomers and oligomers (xylan), along the entrained water and extractives in the biomass. The purpose of the

second step is to remove the rest of the hemicellulose and lignin. A portion of the cellulose may also be removed in this step, in the form of soluble six-carbon sugar monomers and oligomers (glucan). The bleaching step, if used, removes trace lignin and degraded sugars (humins),(*35*) which have been shown to inhibit the activity of some hydrolytic enzymes.(*36*)

In the example displayed in Figure 8.3.1, the organic solvent used is γ-valerolactone (GVL), a non-toxic and bio-renewable solvent that has the demonstrated abilities to solubilize all fractions of woody biomass,(*16*) and to facilitate selective liquid-phase biomass conversion reactions over acid catalysts.(*30, 37*) Other organic solvents may be used, so long as they are miscible with water (which is required to solubilize the carbohydrate fractions of biomass),(*38*) and are stable at the desired pretreatment temperatures. Other solvents used in this study were: tetrahydrofuran (THF), acetonitrile (MeCN); *N*-methyl-pyrrolidinone (NMP); and dimethyl sulfoxide (DMSO).

The biomass type shown in Figure 8.3.1 is P39 poplar, an engineered deciduous energy crop optimize for growth on marginal lands (*i.e.*. lands that are not useable for growing food).(*39*) Three other high yield but non-food-competitive biomass feedstocks were studied in this work: corn stover, switchgrass, sorghum and an additional engineered poplar (NM6).(*40*) The catalyst (H_2SO_4) used in Figure 8.3.1 is a standard reagent used in demonstrated biomass pretreatment/conversion strategies.(*41*) The bleaching step is based on industry-standard protocols, using water a solvent.(*42*)

In a real process, the solvent composition, temperature, acid concentration and reaction time used in each pretreatment step would be optimized for a particular biomass feedstock. In Figure 8.3.1, representative reaction conditions are shown, with corresponding yields of soluble products in each liquid fraction. Note that the mass balance for the overall process is less than

100%, which results from the formation of unaccountable soluble and insoluble polymeric species (humins), which are ubiquitous in biomass conversion processes.(*43*) The yields shown in Figure 8.3.1 are typical for those reported under similar conditions.(*16, 30*)

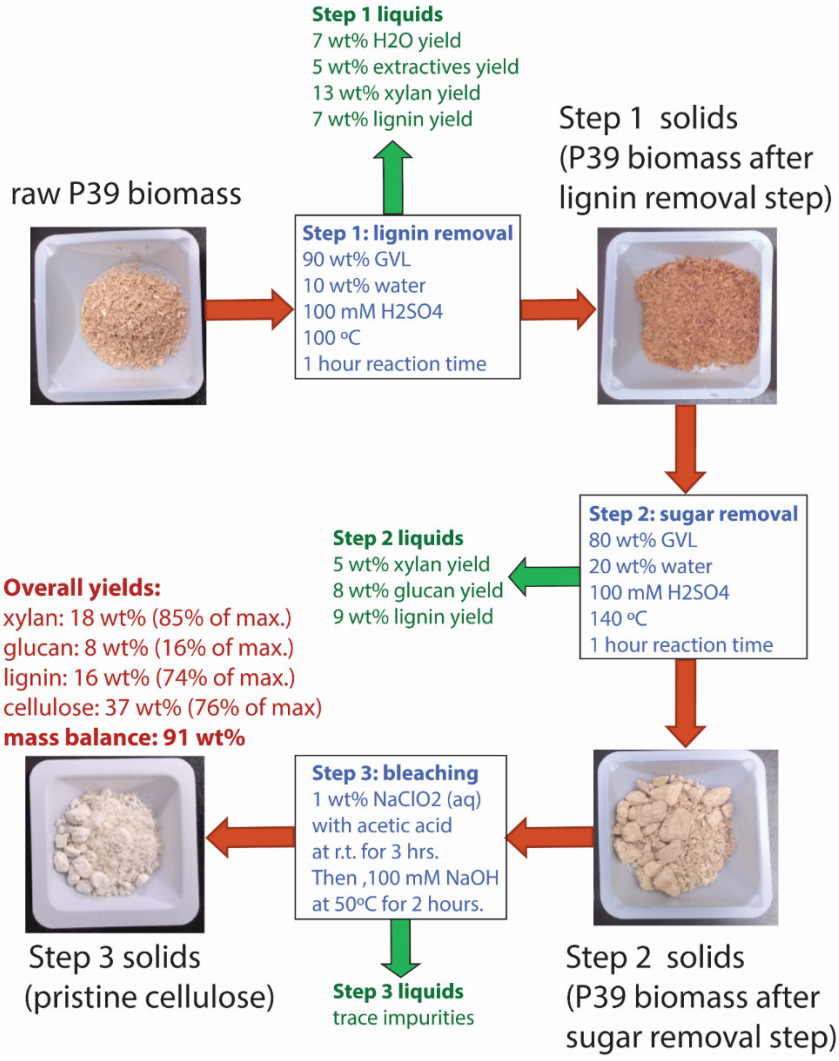


Figure 8.3.1. An exemplary solvent-assisted biomass pretreatment sequence based on the general two-step lignin- and then sugar-removal process described above, followed by an additional bleaching step. Reaction conditions: γ-valerolactone (GVL) solvent with HPLC-grade water as a cosolvent. Sulfuric acid (H_2SO_4) is used as catalyst. Each step was carried out at 9/1 wt/wt liquids to solids in pressure-sealed glass reactors, with mixing facilitated by a magnetic stir bar agitated at 500 rpm.

8.3.2 ^{13}C solid-state cross-polarization magic-angle-spinning (CP/MAS) NMR

Figure 8.3.2 displays the ^{13}C CP/MAS NMR spectra for three representative samples: raw P39 biomass, the same biomass pretreated in GVL-water mixtures to remove lignin (Figure 2, Step 1 solids), and then hemicellulose (Figure 2, Step 2 solids). Key resonances are delineated in Figure 3, corresponding to the six carbon centers in cellulose,(*44*) as well as the acetate groups in hemicellulose, and the methoxy groups present in lignin.(*45*) These latter two resonances decrease in intensity in the pretreated samples as compared to the raw P39, indicating a reduction in the lignin and hemicellulose content, as expected.

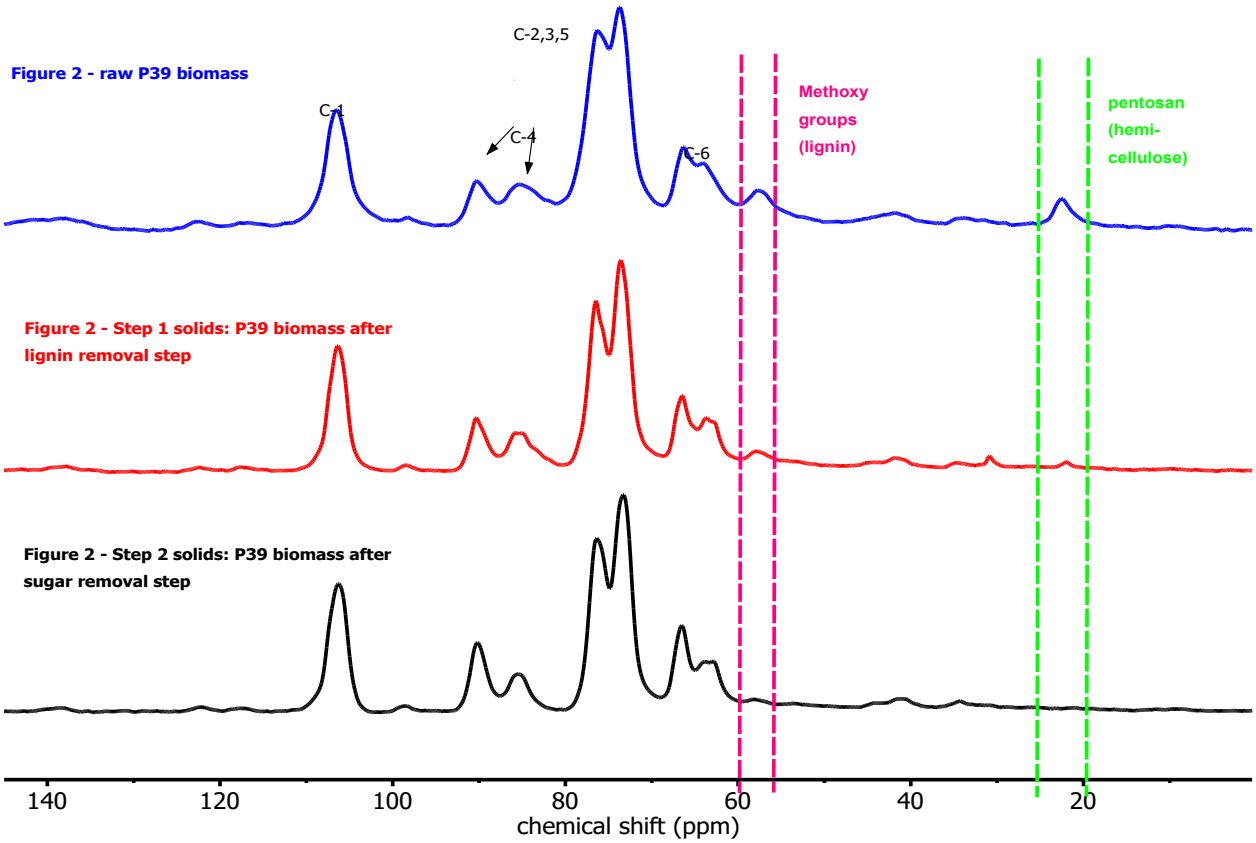


Figure 8.3.2. ^{13}C MAS-NMR spectra corresponding to native P39 biomass, and P39 biomass pretreated to remove lignin and hemicellulose. Conditions for the pretreatment steps associated with each spectrum are noted in the Figure 8.3.1.

8.3.3 attenuated total reflectance infrared (ATR-FTIR) spectroscopy

The removal of lignin and hemicellulose from P39 biomass effected by the solvent-mediated pretreatment is confirmed by attenuated-total-reflectance infrared spectroscopy (ATR-FTIR). These results are shown in Figure 8.3.3, wherein the spectral features corresponding to lignin and hemicellulose vanish going from native P39 biomass to the pretreated material.(*46*) Note that the spectral bands corresponding to hemicellulose and cellulose at about 1000 cm^{-1}

overlap, so that the relative height of the bands at 500 and 1000 cm^{-1} change as hemicellulose is removed from the surface of the cellulose microfibrils.

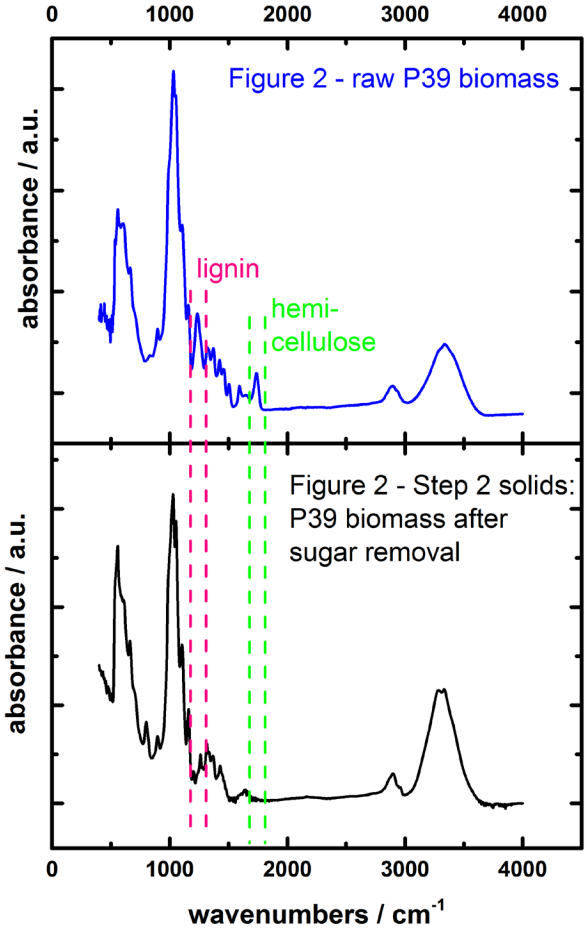


Figure 8.3.3. ATR-FTIR spectra for native P39 biomass (top) and P39 pretreated in GVL-water solvent systems to remove hemicellulose and lignin (bottom). Note that signals corresponding to cellulose and hemicellulose overlap at ~1000 wavenumbers.

8.3.4 Proton spin-relaxation edited (PSRE) CP/MAS NMR

Figure 8.3.4 displays the total ^{13}C CP/MAS NMR spectra for raw P39 biomass, along with the PSRE sub spectrum corresponding to the cellulose fraction from the same sample. The aforementioned broad spectral features are mostly removed in the PSRE spectrum, resulting in better-separated peaks and a flat baseline.(*26*) Note that the resonances corresponding to the C-4 and C-6 carbon centers in the PSRE and the unedited CP/MAS NMR spectra are split into two non-equivalent features. This behavior has been attributed to differences in the chemical environment between cellulose chains located on the outside *vs.* the inside of the microfibrils.(*47*) In particular, the C-4 carbon resonances that correspond to these two environments are well separated in the PSRE spectrum.

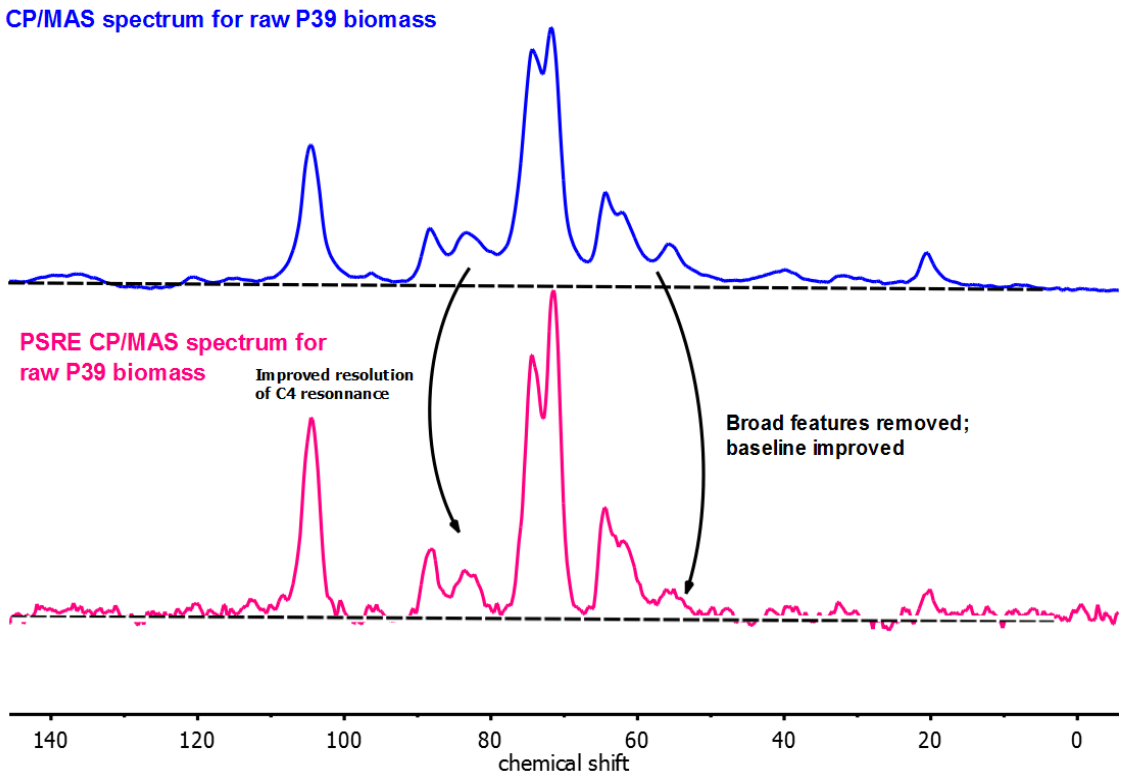


Figure 8.3.4. CP/MAS-NMR spectra for native P39 biomass. The top display is the unedited spectrum produced by averaging 2056 scans with a 10 sec recycle delay time and 0.03 ms spin-lock mixing time. The bottom display is the proton spin-lock edited spectrum corresponding to the crystalline cellulose fraction of native P39.

PSRE cellulose spectra corresponding to the native and pretreated biomasses tested in this study were collected, and compared on the basis of the relative peak areas for the downfield C-4 resonance at ~90 ppm and the upfield C-4 resonance at ~86 ppm, denoted C_{90} and C_{86}, respectively. Figure 8.3.5 displays this procedure. When comparing the PSRE cellulose spectra corresponding to raw and pretreated P39 biomass (with peak areas normalized to the C-1 resonance), note that the combined area for the upfield and downfield C-4 peaks remains constant. In contrast, the *relative* C-4 peak areas, as expressed by the relationship:

$$X_{NMR} = \frac{C_{90}}{C_{90}+C_{86}} \qquad (8.6)$$

Reveal an enrichment of the downfield C-4 peak area going from the raw to the pretreated P39 cellulose. Together, these results indicate a *redistribution* of the C-4 carbons from one chemical environment to the other, rather than a selective depletion of one phase during the GVL pretreatment process. This behavior has been attributed to a reduction in xylan-cellulose interactions at the surface of the cellulose microfibrils.(*48-50*) Accordingly, we attribute increases in X_{NMR} resulting from solvent-mediated pretreatment of biomass to the destruction of cellulose-hemicellulose linkages at the surface of the cellulose microfibrils.

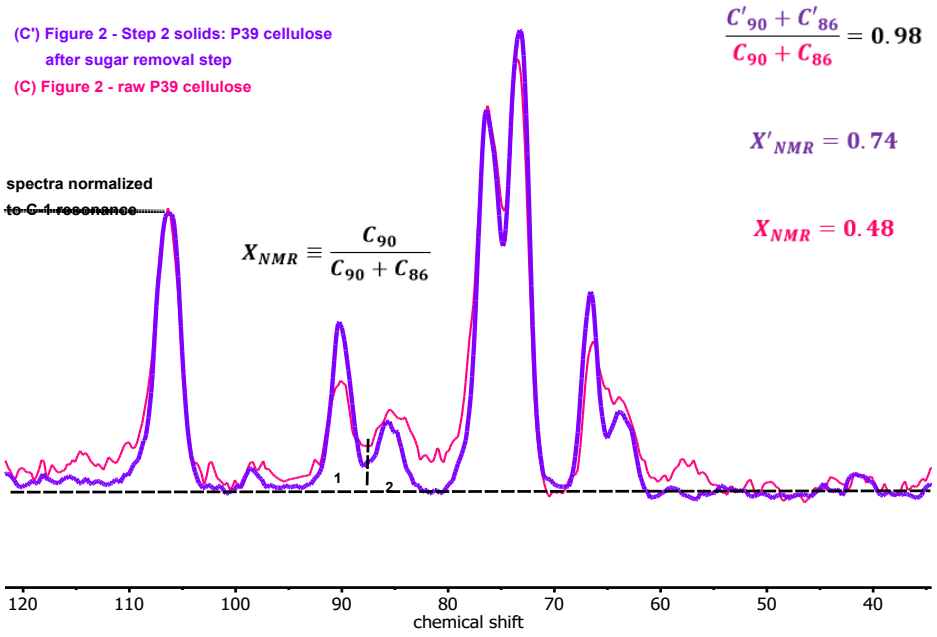


Figure 8.3.5. Proton spin-lock edited (PSRE) CP/MAS NMR spectra corresponding to the crystalline cellulose fraction of native P39 biomass (pink display), and P39 biomass pretreated in GVL/water systems to remove lignin and hemicellulose (purple display).

8.3.5 X-ray diffraction characterizations of native and pretreated cellulose

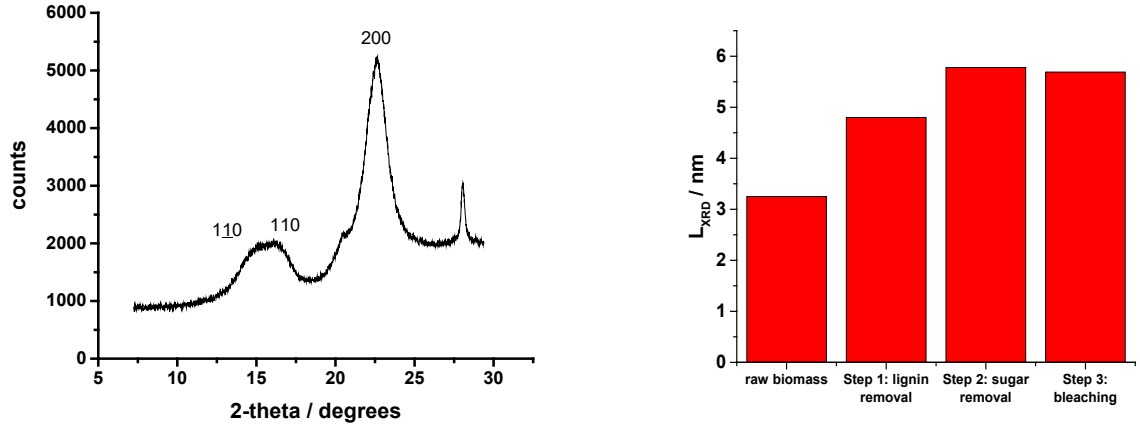


Figure 8.3.6. Representative X-ray diffractogram for native P39 biomass (left) and XRD-derived estimates of the diameter of the cellulose microfibrils in native and solvent-pretreated P39 biomasses.

Figure 8.3.6 displays a representative diffractogram corresponding to cellulose in raw P39 biomass. The diffraction peaks associated with the 1$\underline{1}$0, 110 and 200 facets of the monoclinic unit cell are indicated; their positions on the 2θ scale correspond to inter-planar spacings of 5.91, 5.43 and 3.92 Angstroms, respectively. After the appropriate baseline corrections and accounting for the inherent line broadening of the instrument, the full width at half-maximum (FWHM) intensity for the 200 peak corresponding to each sample are collected. These data are used to estimate the lateral dimensions of the ordered cellulose domains (microfibrils), by invoking the standard relationship:(*51*)

$$L_{XRD} = \frac{K\lambda}{\beta cos\theta} \qquad (8.7)$$

Where L_{XRD} is the average cross-sectional length of a cellulose microfibril, □ is the full width at half maximum intensity of the 200 diffraction peak (in radians), □ is the Bragg angle (again in radians), □ is the wavelength of the scattered X-ray beam, and K is a factor on the order of 0.9 (which is the value used here). Based on the line broadening of the 200 diffraction peak, the apparent size of the microfibrils ranges from 3.25 for native P39 biomass, to about 5.5 nm for GVL-pretreated P39, which is within the range of microfibril dimensions reported for similar cellulosic materials.(*26, 52*) The greatest variance in the apparent value of L_{XRD} corresponds to the native and minimally processed P39 samples, while L_{XRD} approaches a near-constant value as lignin and hemicellulose are removed. It should be noted that phenomena other than changes in the characteristic dimensions of an ordered phase can result in broadening of X-ray diffraction features. These additional phenomena include epitaxial strain, lattice defects, and the presence of amorphous material in the crystal matrix.(*53*) We therefore suggest that the observed variations in the value of L_{XRD} corresponding to native or minimally-pretreated P39 cellulose are attributable to these additional imperfections and amorphous materials, which are removed by the GVL-pretreatment and bleaching processes, causing the apparent dimension L_{XRD} to approach a constant value.

8.3.6 Enzymatic hydrolysis of residual cellulose

The GVL-pretreated solids derived from each step in Figure 8.3.1 were collected, washed with water, filtered to produce a solid product of about 70 wt% moisture content, and hydrolyzed in the presence of engineered cellulases (see ESI for details). Similar tests were also conducted for the different biomass types (corn stover, switchgrass, sorghum and NM6-poplar), and solvent

systems (*54*)(THF, DMSO, NMP and MECN) used in this study. Bleaching steps were applied to some of the GVL-pretreated samples prior to enzymatic hydrolysis to test the effects of this procedure on the digestibility of the residual cellulose. Finally, some of the bleached and unbleached GVL-pretreated celluloses were dried in a vacuum oven at 85º*C* overnight, and then subjected to enzymatic hydrolysis as well. The yields of soluble sugars derived from this enzymatic hydrolysis step were then compared to assess the effect of each pretreatment step on the enzymatic reactivity of the resulting cellulose, whereby higher yields of sugars from the enzymatic hydrolysis step indicate a more effective pretreatment process. These results are displayed in Table 8.3.1, along with X_{NMR}, which was defined previously.

Table 8.3.1. Reaction conditions for solvent-assisted pretreatment of various biomass, and corresponding enzymatic sugar yields and the NMR-derived observable X_{NMR}. The temperature in the lignin removal step was 100ºC for each entry. The catalyst was 100 mM H_2SO_4 for all lignin- and sugar-removal steps. Bleaching was carried out according to standard protocols. Enzymatic hydrolysis was performed with 10 wt% CelR and 0.5 mg/mL pretreated material in a total volume of 1 mL for 24 hours with shaking.

Entry	Biomass type	Organic solvent (wt%) – lignin removal step	Organic solvent (wt%) – sugar removal step	Temp. – sugar removal step / ºC	Bleached ? (Y/N)	Dried? (Y/N)	X_{NMR}	Enzymatic sugar yield from residual cellulose (wt% of theor. maximum)
1	P39	GVL (90)	--	--	N	N	0.504	0.0
2	P39	GVL (90)	--	--	N	Y	.0.529	0.0
3	P39	GVL (90)	--	--	Y	N	0.523	0.0
4	P39	GVL (90)	--	--	Y	Y	0.556	0.0
5	P39	GVL (90)	GVL (80)	120	N	N	0.545	0.2
6	P39	GVL (90)	GVL (80)	120	N	Y	0.573	0.0
7	P39	GVL (90)	GVL (80)	120	Y	N	0.608	3.1
8	P39	GVL (90)	GVL (80)	120	Y	Y	0.639	1.5
9	P39	GVL (90)	GVL (80)	140	N	N	0.745	23.5
10	P39	GVL (90)	GVL (80)	140	N	Y	0.782	5.4
11	P39	GVL (90)	GVL (80)	140	Y	N	0.775	29.4

12	P39	GVL (90)	GVL (80)	140	Y	Y	0.814	9.0
13	P39	THF (90)	THF (80)	140	N	N	0.779	20.9
14	P39	MeCN (90)	MeCN (80)	140	N	N	0.650	9.3
15	P39	NMP (90)	NMP (80)	140	N	N	0.609	4.6
16	P39	DMSO (90)	DMSO (80)	140	N	N	0.633	4.8
17	NM6	GVL (90)	GVL (80)	140	N	N	0.711	19.3
18	Sorghum	GVL (90)	GVL (80)	140	N	N	0.582	5.5
19	switchgrass	GVL (90)	GVL (80)	140	N	N	0.545	7.5
20	Corn stover	GVL (90)	GVL (80)	140	N	N	0.698	13.5

8.3.7 ^{1}H High-Resolution (HR)/MAS NMR spectroscopy

Figure 8.3.7 displays representative HR/MAS spectra corresponding to P39 biomass pretreated in GVL-water solvents to remove hemicellulose and lignin for never-dried and dried materials (Entries 9 and 10 in Table 8.3.1), but the dried material has been re-wetted by soaking in water (7/3 water to dry solids) for 72 hours (see experimental section). HR/MAS spectra were collected with 100 single-rotor-cycle interpulse delays in a Carr–Purcell–Meiboom–Gill (CPMG) pulse sequence for a total of 50 ms transverse relaxation time. Each of the spectra within Figures 8.3.7 display resonances corresponding to the solution-phase protons in the wetted cellulose; *i.e.*, to liquid water only. Without any magnification, only one resonance is visible: a large, complicated multiplet at about 9.5 ppm corresponding to water wetting the surface-accessible facets of the cellulose. The inset of Figure 8.3.7 shows the same spectra magnified 100 times. In these magnified images, we see the singlet at 4.7 ppm corresponding to bulk water. Notably, there is a

small feature present in the spectra corresponding to the re-wetted material that is not present in the never-dried material. The appearance of this feature indicates the formation of distinct hydration environment upon drying and re-wetting, which we attribute to a change in the cellulose pore structure upon drying. This behavior has been attributed to water becoming trapped in segregated hydration environments upon rapid changes in the pore structure of semi-crystalline polymeric materials.(*55*) Accordingly, assign the minor feature at 8.9 ppm to a partially "collapsed" cellulose pore structure that forms upon drying.

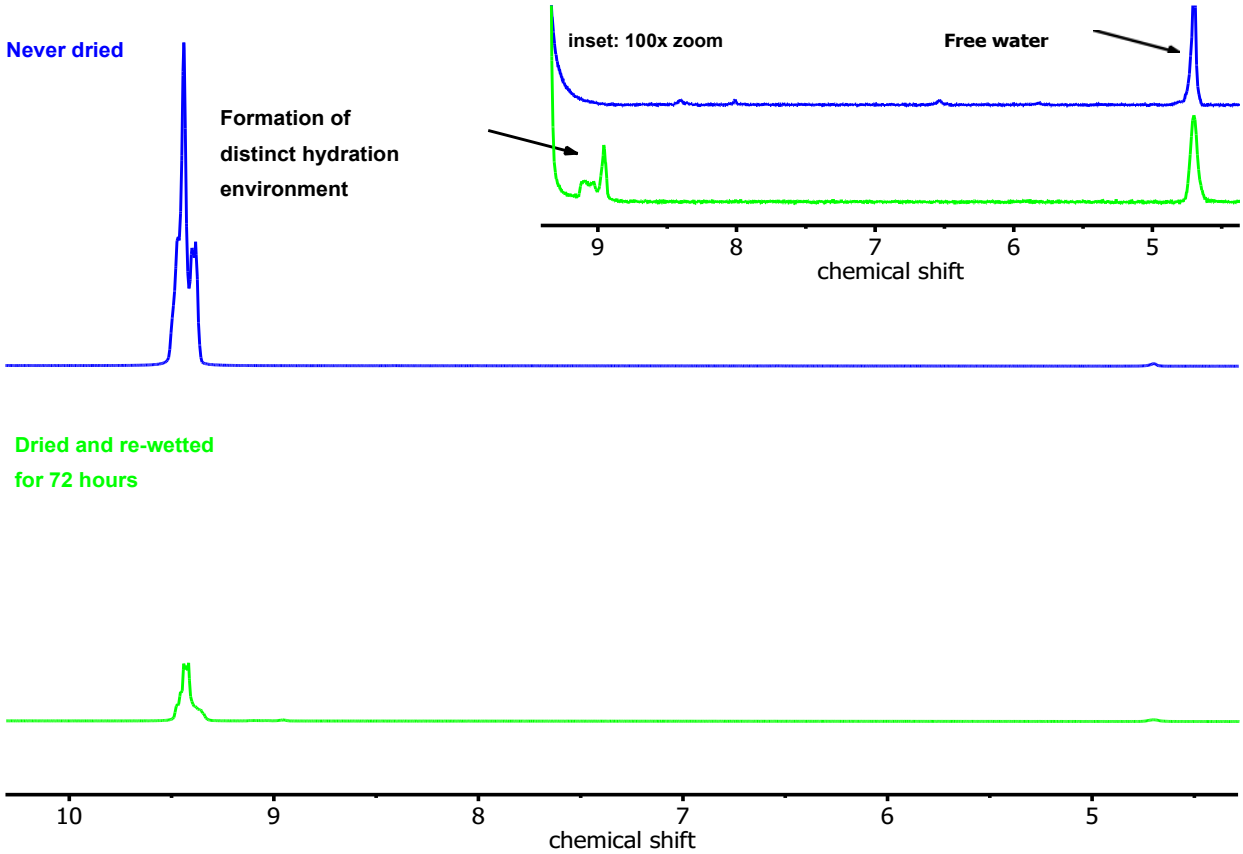


Figure 8.3.7. HR/MAS NMR spectra with a 50 ms Hahn-echo delay applied (100 single-rotor-cycle interpulse delays in a Carr–Purcell–Meiboom–Gill (CPMG) pulse sequence) for cellulose from P39 biomass pretreated in 90 wt% GVL at 100ºC for one hour, then 80 wt% GVL at 140ºC for one hour (Table 8.3.1, Entries 9 and 10). Top spectrum corresponds to cellulose that was never

dried, and analyzed directly after collection from the pretreatment steps. Bottom spectrum corresponds to the same material after drying in a vacuum oven overnight, then re-wetting by soaking in water for 72 hours.

8.4. Discussion

To understand the effects of different solvent-mediated biomass pretreatment strategies on the enzymatic digestibility of the residual cellulose, we first analyze the GVL-pretreated P39 biomasses and corresponding enzymatic sugar yields only. In this context, we probe the effects of the lignin removal step, the sugar removal step, and the bleaching and drying steps. In doing so, we find that the enzymatic sugar yields from all GVL-pretreated P39 celluloses generalize on the basis of X_{NMR}, whereby increases in one quantity correspond universally to increases in the other. We then demonstrate that this relationship extends across all pretreated celluloses tested in this study, regardless of the native biomass type or pretreatment solvent used.

8.4.1 Effects of GVL-pretreatment, bleaching and drying on the enzymatic digestibility of cellulose derived from P39 biomass

If increases in X_{NMR} correspond to a reduction in cellulosic material that is bonded to hemicellulose and lignin, and these extracellular materials occlude hydrolytic enzymes' access to the cellulose microfibrils[25], it is suggested that X_{NMR} might serve as a predictor of the enzymatic digestibility of the residual cellulose derived from GVL-pretreated biomass. Following this hypothesis, Figure 8.4.1 displays the yields of enzymatic sugars from the hydrolysis of GVL-pretreated P39 (both with and without additional bleaching) as function of X_{NMR}. Two curves are

shown: one corresponding to the never-dried samples, and the same samples after drying in a vacuum oven at -15 in. Hg for 24 hours (open symbols). In general, for both the never-dried and dried GVL-pretreated P39 celluloses, the yield of enzymatic sugars after a 24 reaction time increases monotonically with X_{NMR}.

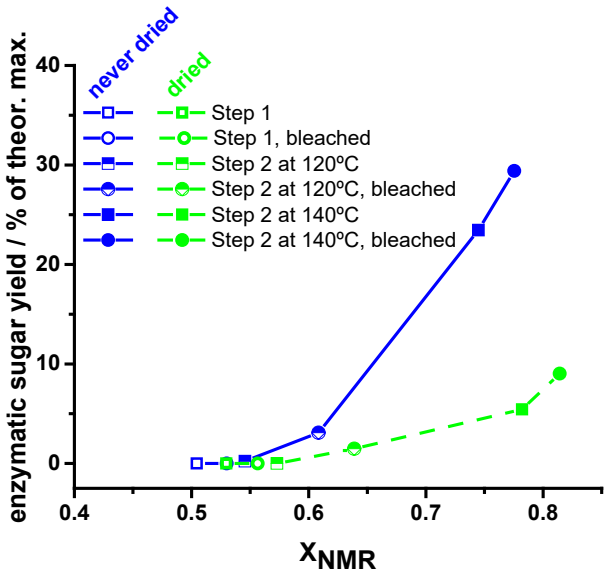


Figure 8.4.1. Enzymatic sugar yields from P39 biomass pretreated with GVL-water solvent systems under various conditions, both with and without bleaching steps (entries 1 through 12 in Table 1).

To further probe the hypothesis that X_{NMR} serves are a predictor of the enzymatic digestibility of solvent-pretreated cellulose, we now compare enzymatic sugar yields as a function of X_{NMR} for the different biomass types and solvent systems tested, where there were no bleaching steps or drying, and the sugar removal step was fixed at 140ºC. This analysis allows for the effects of different biomass types and organic pretreatment solvents on enzymatic sugar yields to be assessed at a self-consistent set of conditions with respect to pretreatment times, temperatures, and water content of the solvent systems. Figure 8.4.2 displays these results. Data points represent:

(left) enzymatic sugar yields as a function of X_{NMR} for P39 biomass pretreated in GVL-, THF-, MeCN-, NMP- and DMSO-water mixtures, and; (right) corn stover, switchgrass, sorghum, NM6 poplar and P39 poplar pretreated in GVL-water solvents. As demonstrated by the linear regression analyses displayed in Figure 8.4.2, enzymatic reactivity generalizes universally with X_{NMR} across all tested solvent systems and biomass types at a fixed set of reaction times and temperatures for the lignin removal and sugar removal steps.

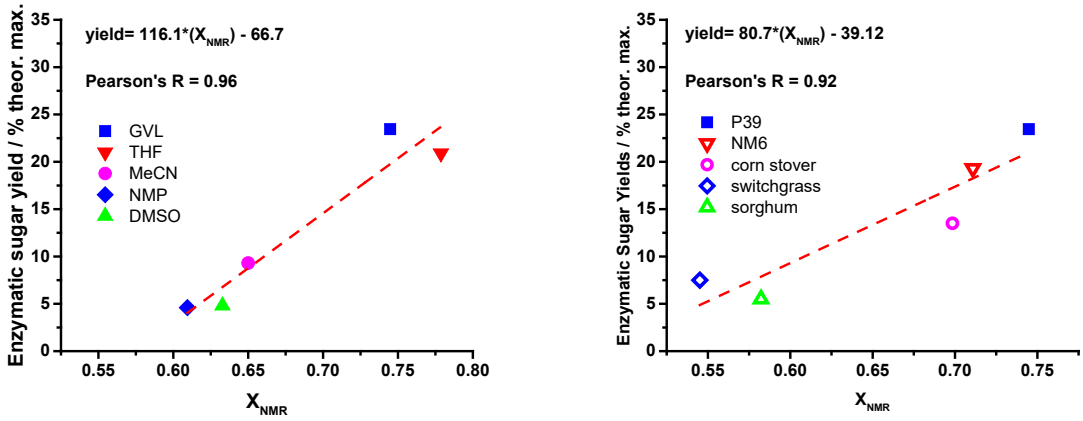


Figure 8.4.2. Enzymatic sugar yields versus X_{NMR} for: (left) P39 biomass pretreated in various water/organic-cosolvent mixtures to remove lignin and hemicellulose (entries 11, 15-18 in Table 1), and; (right) different biomass types pretreated in GVL-water solvent mixtures to remove lignin and hemicellulose (entries 11, 19-22 in Table 1). Least squares linear fit to the data are shown, demonstrating a quantiative corrlation between X_{NMR} and enzyamtic sugar yields across various biomass types and pretreatment solvents.

Finally, Figure 8.4.3 displays enzyamtic sugar yields as a function of X_{NMR} for all entires in Table 1 corresponding to *never-dried* celluoses. As demonstrated by the linear regression in

Figure 11, there is a qunatiative, linear and *universal correlation* between X_{NMR} and enzyamtic reactivity (as expressed by sugar yields) for all biomass types, solvent systems, and pretreatement temperatures tested in this study. Therefore, the results displayed in Figure 8.4.3 indicate that for a *fixed set of enzymatic hydroylsis conditions*, X_{NMR} serves as a *universal predictor* of the enzyamtic digestibility of solvent-pretreated celluloses, whereby increases in the former translate systematically into increases in the latter.

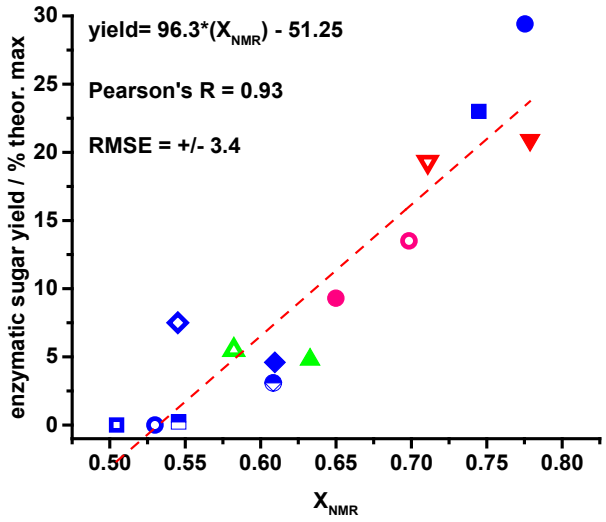


Figure 8.4.3. enzymatic sugar yields versus X_{NMR} for all *never-dried* entries in Table 1. Symbols correspond to those datapoints indicated in the legends in Figures 9 and 10.

8.4.2 The effects of drying on the hydration behavior of GVL-pretreated P39 cellulose

In contrast to the never-dried celluloses, the dried celluloses in Figure 8.4.1 exhibit a slight increase in X_{NMR}, but systematically lower yields of enzymatic sugars. This result indicates that decreasing the moisture content of pretreated cellulose alters its physical or chemical properties,

other than the fraction of microfibril surface chains liberated from the lignin-hemicellulose matrix, and that these changes correspond to a decrease in enzymatic digestibility. Furthermore, the results displayed in Figure 8.3.7 indicate that drying GVL-pretreated celluloses alters the pore structure, creating distinct hydration environments as compared to the never-dried material. It has been reported that polymeric materials composed of identical monomer units but synthesized using different methods exhibit similar behavior, whereby different preparation methods produce distinct pore structures, and therefore different hydration environments between materials. These differences in hydration environments can be characterized in terms of the binding strengths of water with the surface accessible facets of the pore structure, the diffusivity of water within the pores, and the extent to which water exchanges readily between the pore structure and the bulk fluid. Changes in these characteristics would suggest changes in the ease with which macromolecules such as cellulases diffuse from the bulk solution to the interior pores of the cellulose suprastructure, explaining differences in reactivity between dried and never-dried cellulose materials, and these changes can be probed using ^{1}H HR/MAS NMR.(*56, 57*)

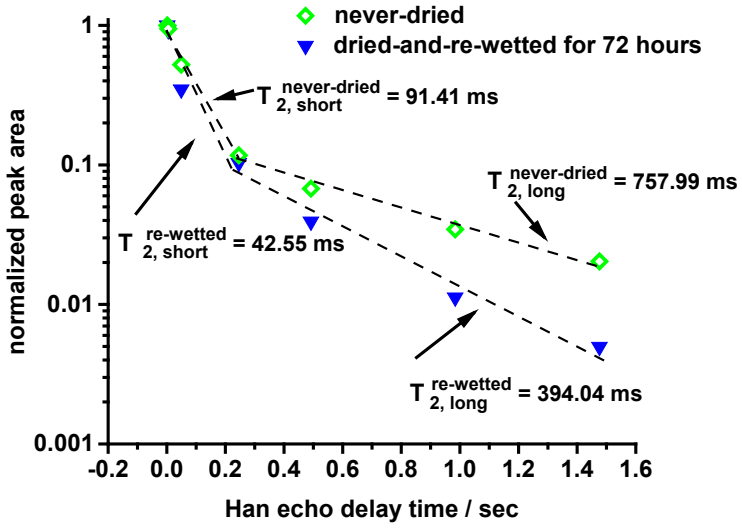


Figure 8.4.4. Normalized, integrated areas of main hydration resonance at ~9.45 ppm in the ^{1}H HR-MAS spectra as a function of the number of CPMG rotor cycle delays or, equivalently, the Hahn echo delay time. Integrals shown are for never-dried, and dried-then-re-wetted P39 biomass that was treated with GVL-water solvents to remove hemicellulsoe and lignin (entries 9 and 10 in Table 1, respectively).

Figure 8.4.4 displays the total integrated area of the main water resonance at ~9.5 ppm (hydrated cellulose; Figure 8), as a function of the the Han Echo delay time in a rotor-synchronized CPMG ^{1}H HR-MAS pulse sequence for GVL-pretreated P39 cellulose. Closed symbols correspond to samples that were never dried, while open symbols corresond to the same samples after being dried in a vacuum oven overnight, and then re-wetted by soaking in DI water at room temperature for 72 hours. Transverse spin-spin relaxation times (T_2) were estimated by fitting the data in Figure 8.4.4 to a bi-exponential curve (shown in solid lines). Accordingly, two time constants were assigned to the T_2 relaxation process for the never-dried and the dried-and-re-wetted GVL-pretreated cellulose samples: one long relaxation time on the order of 10 seconds ($T_{2,\,long}$), and one short relaxation time on the order of 1 second ($T_{2,\,short}$). These time constants are also displayed in Figure 8.4.4. Going from the never-dried to the dried-and-re-wetted samples, the time constansts associated with the spin-spin-relaxation processes for water within the cellulose pore strucutre roughly double. Short spin-spon relaxation times have been attributed to water that binds more strongly within a polymer matrix.(*58*)

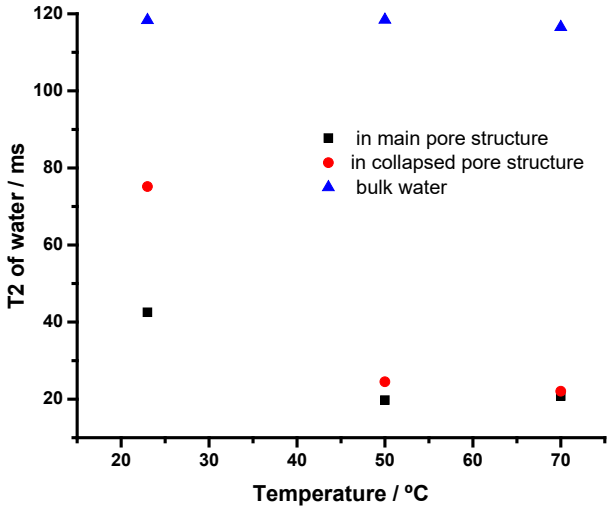


Figure 8.4.5. T_2 relaxation time contants for water in the main pore structure of cellulose (9..4 ppm in Figure 11) and the "collapsed" pore strucutre of cellulose that forms upon drying (8.9 ppm in Figure 11) as a function of temperature. Spectra were collected with a CPMG filter (Hahn echo delay time) of 2 to 250 ms, and the absolute integrals of the correpsonding resonances were analyzed as a function of Hahn echo delay time using a single exponential decay to estimate spin-spin relaxation time constants (T_2's).

Finally, we examined the temperature-dependent spin-spin relaxation behavior of the ^{1}H HR/MAS resonances in Figure 8.4.4 that we attribute bulk water (4.7 ppm), and main pore sturcutre (9.4 ppm) and "collapsed" pore stucture (8.9 ppm) of dried-and-re-wetted cellulose. These data are displayed in Figure 8.4.5. Briefly, ^{1}H HR-MAS NMR spectra were collected using a CPMG pulse sequence with a Hahn-echo delay between 4 and 250 ms, with the temperature varied between 23 and 70 ºC. The absolute integrals of the aforementioned resonances were analyzed as a function of Hahn echo delay time using a single exponential decay to estimate spin-spin relaxation time constants (T_2's) .

At room temperature, the spin-spin relaxation times (T_2's) of bulk water, and water in the main and collapsed pore stucture of previous dried celluose are clearly differentiated. If water were able to exchange readily between cellulose pore strcuture and the bulk water, albiet at a slow rate on the NMR timescale at room temperature, then it would be expected that the spin-spin relaxation rates (T_2's) of the corresponding resonsnaces would approach one another with increasing temperature, as the rate of exchange increases. However, we find that the spin-spin relaxation time constant of bulk water is contant with tempearture, whereas the T_2's for water in the main and collapsed cellulose pore strucutres are initially differentiated at 23ºC, but decrease and approach one another with increasing temperature. Accordingly, this behavior indicates that, while water may exchange (diffuse) between the main or secondary hydration environmetns (main pore stucutre and collapsed pore strucutres) at these temperatures, water does not exchange readily between the cellulose pore strucutres and the bulk.

All things considered, we interpert these results to indicate that, upon drying GVL-pretreated cellulose, the cellulose pore strucutre partially collapses, trapping water in a number of segregated hydration enviroments wherein water binds more stongly with the surface-acessible facets of cellulose, and does not exchange readily with the bulk water outside the cellulose pores. We expect that these changes in hydration behavior, as compared to the never-dried cellulose, inhibits the diffusion of cellulases into the pores strucutres, as well as the binding of the cellulases to the surface accessible facets of the cellulose, limiting enzymatic sugar yields.

8.4.3. Physical interpretations and predicting the effectiveness of solvent-assisted biomass pretreatment strategies using NMR

Following the discussions above, we propose a physical interpretation for the key changes effected by solvent-mediated pretreatment of woody biomass over acid catalysts. These changes are illustrated schematically in Figure 8.4.6, below. As native lignocellulosic biomass is pretreated in semi-aqueous organic solvent systems (for example GVL-water solvents) over acid catalysts, and/or bleached, lignin, hemicellulose and other extracellular material is removed from the cell wall. At sufficiently low temperatures and short reaction times, however, the cellulose microfibrils are largely untouched, so that their characteristic dimension do not change. Furthermore, without a drying step, water occupies the interstitial areas between the remaining microfibrils, so that the pore structure formed by these rigid features in the native cell wall is preserved. The result is a mostly cellulose-containing solid through which water can easily diffuse and exchange with the bulk liquid, so that hydrolytic enzymes (cellulases) can readily migrate to the pore interior.

In order to facilitate the hydrolysis of the cellulosic material, however, the enzymes must also bind to the glyosidic linkages exposed on the surface-accessible facets of the cellulose microfibrils, and so the proportion of these linkages that are exposed to the cellulases ultimately determines the enzymatic reactivity of the cellulose. This fraction of exposed linkages, in turn, is determined by the extent to which the solvent is able to remove not just the bulk extracellular material, but to destroy the intimate linkages between the cellulose and the hemicellulose/lignin matrices. As discussed above, this characteristic feature can be probed using PSRE CP/MAS NMR, though the observable X_{NMR}. Furthermore, as demonstrated by the analysis displayed in the Figure 8.4.3, the extent to which these linkages are destroyed (improving enzymatic reactivity) at a fixed set of pretreatment conditions (time, temperature and acid concentration) depend both on the organic solvent used for a fixed biomass type, and upon the biomass type that is pretreated in

a fixed organic solvent system. Upon drying, the microfibril ultrastructure collapses, trapping remaining water into a number of segregated hydration environments. The water in these hydration environments binds more strongly than the water present in the never-dried material, so that it does not diffuse as easily, nor exchange as readily with the bulk liquid. This hinders the ability of macromolecules such as enzymes to diffuse into the cellulose ultrastructure interior, reducing the reactivity of hydrolytic enzymes, and limiting the yields of enzymatic sugars.

Reasons as to why a particular solvent system is better able to liberate cellulose surface chains from the surrounding hemicellulose/lignin matrices in some biomass types than in others likely relate to structural differences between biomasses, so that the cellulose microfibrils are more tightly bound by hemicellulose or lignin and therefore more difficult to liberate with solvent mediated-pretreatments. Reasons as to why, for a fixed biomass type, one solvent system is better able to liberate cellulose from the surrounding extracellular material than another likely relate to mechanistic details underlying the hemicellulose and lignin depolymerization process. For example, THF/water systems have been shown to locally phase separate on hydrophilic and hydrophobic facets of cellulose, increasing the amount of cellulose solubilized compared to pure water.(*59*) This preferential solvation behavior has been shown to drive reactivity in biomass conversion processes, and may be relevant here as well.(*60-62*) Probing these effects and corresponding hypotheses in further detail will require more detailed biomass characterizations, rigorous theoretical treatments (such a atomistic simulations), and will form the basis for future work.

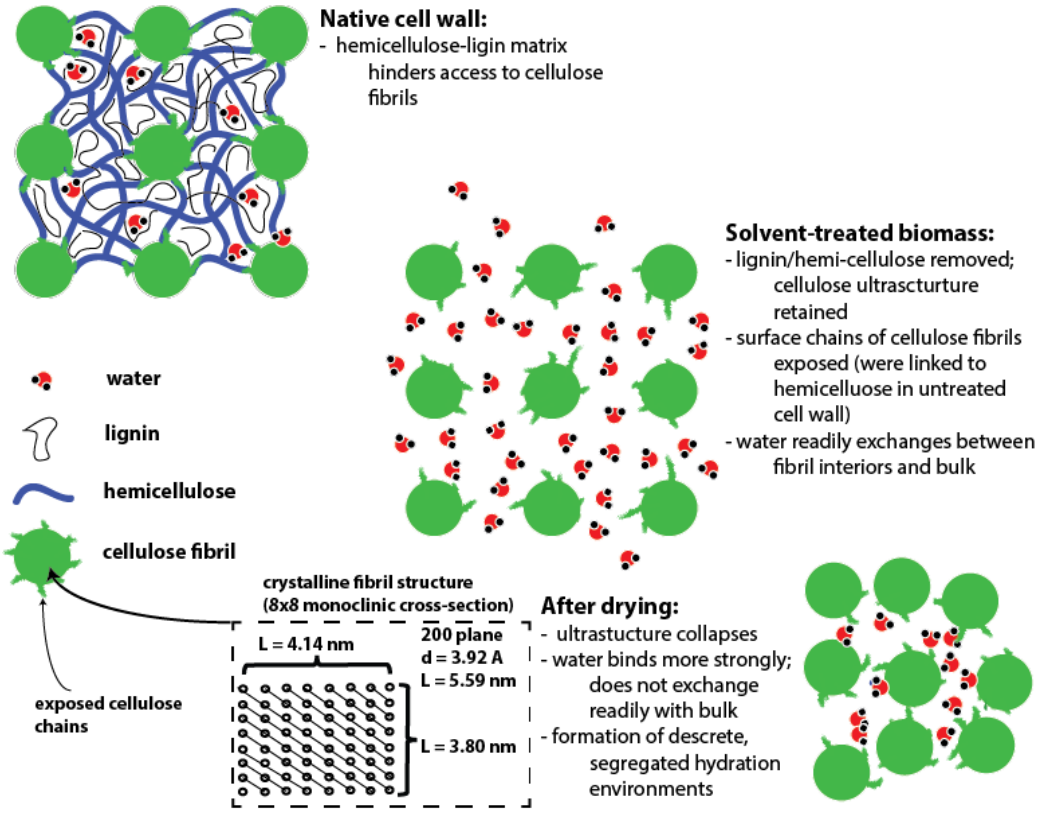


Figure 8.4.6. Schematic representation of the structural features of lignocellulosic biomass at various stages of the pretreatment process. Characteristic dimensions of the cellulose microfibrils are for native P39 biomass; values are estimated based on the combined NMR/XRD analysis described in the main text.

8.5. Conclusions

At short reaction times and lower temperatures, pretreatment of lignocellulosic biomass in aqueous mixtures of organic solvents with dilute acid catalysts removes lignin, hemicellulose and extracellular material from the bulk interstitial areas between crystalline cellulose microfibrils in the cell wall. At higher temperatures and longer reaction times, the characteristic dimensions of

the crystalline cellulose domains remain unchanged, but the intimate linkages between the exterior of the cellulose microfibrils and the lignin/hemicellulose matrix are destroyed, exposing surface-accessible cellulose chains on the outsides of the microfibrils. The extent to which these surface-accessible cellulose chains are liberated from the lignin and hemicellulose is observable by solid state NMR, and herein we express these changes in terms a single, quantitative metric. Changes in this NMR-observable metric as function of biomass pretreatment conditions (reaction time and temperature) universally predict the yields of enzymatic sugars afforded by hydrolysis of the residual cellulose, regardless of biomass type or the organic pretreatment solvent used.

Drying the residual cellulose collected from solvent-pretreated biomass alters the pore structure constituted by the cellulose microfibrils, so that water binds more strongly within the pores, and does not exchange readily with the bulk. This limits the diffusion of enzymes into the cellulose ultrastructure, limiting enzymatic reactivity of the cellulose accordingly. Therefore, toward designing improved biomass conversion strategies, solvent-mediated pretreatments that are most able to liberate the surface-accessible facets of the cellulose microfibrils from the surrounding lignin and hemicellulose render biomass most readily hydrolysable by enzymes, and likewise best improve the yields of enzymatic sugars from biomass. The moisture content of the pretreated cellulose should be carefully controlled prior to enzymatic conversion, as changes in moisture content alter the structure of the residual cellulose and limit enzymatic reactivity.

The NMR-based analysis prescribed herein serves as a screening tool to compare solvent-mediated biomass pretreatment techniques, differentiating the most effective strategies on the basis a simple, quantitative, NMR-observable metric. In this context, the effectiveness of solvent-mediated pretreatments can be compared on the basis of different biomass types, and different solvent compositions. With respect to the former, detailed structure-property relationships that

connect the composition and structural features of different biomasses with their susceptibility to solvent-assisted pretreatments should then be developed. With respect to the latter, molecular-level understandings of how semi-aqueous organic solvent mixtures facilitate the decomposition of cellulose-extracellular linkages should be probed using reactivity measurements and detailed, atomistic simulations. Together, these new insights would provide a framework to effect the rational design of new solvent systems for pretreating lignocellulosic biomass, where the properties of the solvent system are deliberately tuned to target the cellulose-extracellular linkages in a particular biomass type.

8.6. Acknowledgements

This work was supported in part by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Numbers DE-SC0018409.

8.7. References

1. O. Bobleter, Hydrothermal degradation of polymers derived from plants. *Progress in Polymer Science* **19**, 797-841 (1994).
2. N. Mosier *et al.*, Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresource Technology* **96**, 673-686 (2005).
3. B. M. Jenkins, L. L. Baxter, T. R. Miles, T. R. Miles, Combustion properties of biomass. *Fuel Processing Technology* **54**, 17-46 (1998).
4. A. C. O'SULLIVAN, Cellulose: the structure slowly unravels. *Cellulose* **4**, 173-207 (1997).
5. Y. Nishiyama, Structure and properties of the cellulose microfibril. *Journal of Wood Science* **55**, 241-249 (2009).
6. A. Frey-Wyssling, The Fine Structure of Cellulose Microfibrils. *Science* **119**, 80-82 (1954).
7. L. Salmén, Micromechanical understanding of the cell-wall structure. *Comptes Rendus Biologies* **327**, 873-880 (2004).
8. P. J. Harris, B. A. Stone. (Wiley-Blackwell, Chichester, 2009), pp. 61-93.

9. J. N. Chheda, G. W. Huber, J. A. Dumesic, Liquid-Phase Catalytic Processing of Biomass-Derived Oxygenated Hydrocarbons to Fuels and Chemicals. *Angewandte Chemie International Edition* **46**, 7164-7183 (2007).
10. S. G. Wettstein, D. M. Alonso, E. I. Gürbüz, J. A. Dumesic, A roadmap for conversion of lignocellulosic biomass to chemicals and fuels. *Current Opinion in Chemical Engineering* **1**, 218-224 (2012).
11. B. E. Gaddy, V. Sivaram, T. B. Jones, L. Wayman, Venture Capital and Cleantech: The wrong model for energy innovation. *Energy Policy* **102**, 385-395 (2017).
12. H. Jørgensen, J. B. Kristensen, C. Felby, Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. *Biofuels, Bioproducts and Biorefining* **1**, 119-134 (2007).
13. Q. A. Nguyen, J. N. Saddler, An integrated model for the technical and economic evaluation of an enzymatic biomass conversion process. *Bioresource Technology* **35**, 275-282 (1991).
14. L. Shuai, J. Luterbacher, Organic solvent effects in biomass conversion reactions. *ChemSusChem* **9**, 133-155 (2016).
15. C. M. Cai, T. Zhang, R. Kumar, C. E. Wyman, THF co-solvent enhances hydrocarbon fuel precursor yields from lignocellulosic biomass. *Green Chemistry* **15**, 3140-3145 (2013).
16. J. S. Luterbacher *et al.*, Nonenzymatic Sugar Production from Biomass Using Biomass-Derived γ-Valerolactone. *Science* **343**, 277-280 (2014).
17. D. Klein-Marcuschamer, P. Oleskowicz-Popiel, B. A. Simmons, H. W. Blanch, The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnology and Bioengineering* **109**, 1083-1087 (2012).
18. D. Humbird *et al.*, "Process Design and Economics for Biochemical Conversion of Lignocellulosic Biomass to Ethanol: Dilute-Acid Pretreatment and Enzymatic Hydrolysis of Corn Stover," (; National Renewable Energy Lab. (NREL), Golden, CO (United States), 2011).
19. X. Zhao, K. Cheng, D. Liu, Organosolv pretreatment of lignocellulosic biomass for enzymatic hydrolysis. *Applied Microbiology and Biotechnology* **82**, 815 (2009).
20. A. T. W. M. Hendriks, G. Zeeman, Pretreatments to enhance the digestibility of lignocellulosic biomass. *Bioresource Technology* **100**, 10-18 (2009).
21. C. A. Mooney, S. D. Mansfield, R. P. Beatson, J. N. Saddler, The effect of fiber characteristics on hydrolysis and cellulase accessibility to softwood substrates. *Enzyme and Microbial Technology* **25**, 644-650 (1999).
22. S. P. S. Chundawat *et al.*, Restructuring the Crystalline Cellulose Hydrogen Bond Network Enhances Its Depolymerization Rate. *Journal of the American Chemical Society* **133**, 11163-11174 (2011).
23. V. S. Chang, B. Burr, M. T. Holtzapple, in *Biotechnology for Fuels and Chemicals: Proceedings of the Eighteenth Symposium on Biotechnology for Fuels and Chemicals Held May 5–9, 1996, at Gatlinburg, Tennessee,* B. H. Davison, C. E. Wyman, M. Finkelstein, Eds. (Humana Press, Totowa, NJ, 1997), pp. 3-19.
24. L. T. Fan, Y.-H. Lee, D. R. Beardmore, The influence of major structural features of cellulose on rate of enzymatic hydrolysis. *Biotechnology and Bioengineering* **23**, 419-424 (1981).

25. H. E. Grethlein, The Effect of Pore Size Distribution on the Rate of Enzymatic Hydrolysis of Cellulosic Substrates. *Bio/Technology* **3**, 155-160 (1985).
26. R. H. Newman, Estimation of the lateral dimensions of cellulose crystallites using 13C NMR signal strengths. *Solid State Nuclear Magnetic Resonance* **15**, 21-29 (1999).
27. R. H. Atalla, J. Gast, D. Sindorf, V. Bartuska, G. Maciel, Carbon-13 NMR spectra of cellulose polymorphs. *Journal of the American Chemical Society* **102**, 3249-3251 (1980).
28. R. Newman, J. Hemmingson, I. Suckling, Carbon-13 nuclear magnetic resonance studies of kraft pulping. *Holzforschung-International Journal of the Biology, Chemistry, Physics and Technology of Wood* **47**, 234-238 (1993).
29. J. Sugiyama, R. Vuong, H. Chanzy, Electron diffraction study on the two crystalline phases occurring in native cellulose from an algal cell wall. *Macromolecules* **24**, 4168-4175 (1991).
30. D. M. Alonso *et al.*, Increasing the revenue from lignocellulosic biomass: Maximizing feedstock utilization. *Science Advances* **3**, e1603301 (2017).
31. L. Shuai, Y. M. Questell-Santiago, J. S. Luterbacher, A mild biomass pretreatment using γ-valerolactone for concentrated sugar production. *Green Chemistry* **18**, 937-943 (2016).
32. J. S. Luterbacher *et al.*, Lignin monomer production integrated into the γ-valerolactone sugar platform. *Energy & Environmental Science* **8**, 2657-2663 (2015).
33. T. Yokoyama, Revisiting the Mechanism of β-O-4 Bond Cleavage During Acidolysis of Lignin. Part 6: A Review. *Journal of Wood Chemistry and Technology* **35**, 27-42 (2015).
34. P. J. Deuss *et al.*, Aromatic Monomers by in Situ Conversion of Reactive Intermediates in the Acid-Catalyzed Depolymerization of Lignin. *Journal of the American Chemical Society* **137**, 7456-7467 (2015).
35. N. Johar, I. Ahmad, A. Dufresne, Extraction, preparation and characterization of cellulose fibres and nanocrystals from rice husk. *Industrial Crops and Products* **37**, 93-99 (2012).
36. Q. Yang, X. Pan, Correlation between lignin physicochemical properties and inhibition to enzymatic hydrolysis of cellulose. *Biotechnology and Bioengineering* **113**, 1213-1224 (2016).
37. M. A. Mellmer *et al.*, Solvent Effects in Acid-Catalyzed Biomass Conversion Reactions. *Angewandte Chemie International Edition* **53**, 11872-11875 (2014).
38. M. A. Mellmer, D. Martin Alonso, J. S. Luterbacher, J. M. R. Gallo, J. A. Dumesic, Effects of γ-valerolactone in hydrolysis of lignocellulosic biomass to monosaccharides. *Green Chemistry* **16**, 4659-4662 (2014).
39. F. Unda, H. Kim, C. Hefer, J. Ralph, S. D. Mansfield, Altering carbon allocation in hybrid poplar (Populus alba × grandidentata) impacts cell wall growth and development. *Plant Biotechnology Journal* **15**, 865-878 (2017).
40. S. K. Hamilton, M. Z. Hussain, A. K. Bhardwaj, B. Basso, G. P. Robertson, Comparative water use by maize, perennial crops, restored prairie, and poplar trees in the US Midwest. *Environmental Research Letters* **10**, 064015 (2015).
41. T. A. Lloyd, C. E. Wyman, Combined sugar yields for dilute sulfuric acid pretreatment of corn stover followed by enzymatic hydrolysis of the remaining solids. *Bioresource Technology* **96**, 1967-1977 (2005).
42. J. I. Morán, V. A. Alvarez, V. P. Cyras, A. Vázquez, Extraction of cellulose and preparation of nanocellulose from sisal fibers. *Cellulose* **15**, 149-159 (2008).

43. S. K. R. Patil, C. R. F. Lund, Formation and Growth of Humins via Aldol Addition and Condensation during Acid-Catalyzed Conversion of 5-Hydroxymethylfurfural. *Energy & Fuels* **25**, 4745-4755 (2011).
44. R. H. Atalla, J. C. Gast, D. W. Sindorf, V. J. Bartuska, G. E. Maciel, Carbon-13 NMR spectra of cellulose polymorphs. *Journal of the American Chemical Society* **102**, 3249-3251 (1980).
45. G. Gilardi, L. Abis, A. E. G. Cass, Carbon-13 CP/MAS solid-state NMR and FT-IR spectroscopy of wood cell wall biodegradation. *Enzyme and Microbial Technology* **17**, 268-275 (1995).
46. C. Zhou, W. Jiang, B. K. Via, O. Fasina, G. Han, Prediction of mixed hardwood lignin and carbohydrate content using ATR-FTIR and FT-NIR. *Carbohydrate Polymers* **121**, 336-341 (2015).
47. W. L. Earl, D. L. VanderHart, Observations by high-resolution carbon-13 nuclear magnetic resonance of cellulose I related to morphology and crystal structure. *Macromolecules* **14**, 570-574 (1981).
48. P. T. Larsson, E.-L. Hult, K. Wickholm, E. Pettersson, T. Iversen, CP/MAS 13C-NMR spectroscopy applied to structure and interaction studies on cellulose I. *Solid state nuclear magnetic resonance* **15**, 31-40 (1999).
49. R. H. Newman, J. A. Hemmingson, I. D. Suckling, in *Holzforschung - International Journal of the Biology, Chemistry, Physics and Technology of Wood*. (1993), vol. 47, pp. 234.
50. E.-L. Hult, P. T. Larsson, T. Iversen, A comparative CP/MAS 13C-NMR study of cellulose structure in spruce wood and kraft pulp. *Cellulose* **7**, 35-55 (2000).
51. B. Post, X-ray diffraction procedures for polycrystalline and amorphous materials. Harold P. Klug and Leroy E. Alexander, John Wiley & Sons, New York, 1974, pp. 960. $37.50. *X-Ray Spectrometry* **4**, A18-A18 (1975).
52. K. Leppänen *et al.*, Structure of cellulose and microcrystalline cellulose from various wood species, cotton and flax studied by X-ray scattering. *Cellulose* **16**, 999-1015 (2009).
53. T. Ungár, Microstructural parameters from X-ray diffraction peak broadening. *Scripta Materialia* **51**, 777-781 (2004).
54. A. M. Hindeleh, D. J. Johnson, Peak resolution and X-ray crystallinity determination in heat-treated cellulose triacetate. *Polymer* **13**, 27-32 (1972).
55. T. M. Alam, K. K. Childress, K. Pastoor, C. V. Rice, Characterization of free, restricted, and entrapped water environments in poly (N-isopropyl acrylamide) hydrogels via 1 H HRMAS PFG NMR spectroscopy. *Journal of Polymer Science Part B: Polymer Physics* **52**, 1521-1527 (2014).
56. C. Zhang *et al.*, Hierarchical porous structures in cellulose: NMR relaxometry approach. *Polymer* **98**, 237-243 (2016).
57. C. Corsaro *et al.*, The role of water in the degradation process of paper using 1H HR-MAS NMR spectroscopy. *Physical Chemistry Chemical Physics* **18**, 33335-33343 (2016).
58. M. J. Selig, L. G. Thygesen, C. Felby, Correlating the ability of lignocellulosic polymers to constrain water with the potential to inhibit cellulose saccharification. *Biotechnology for Biofuels* **7**, 159 (2014).

59. B. Mostofian *et al.*, Local Phase Separation of Co-solvents Enhances Pretreatment of Biomass for Bioenergy Applications. *Journal of the American Chemical Society* **138**, 10869-10878 (2016).
60. T. W. Walker *et al.*, Universal kinetic solvent effects in acid-catalyzed reactions of biomass-derived oxygenates. *Energy & Environmental Science* **11**, 617-628 (2018).
61. M. A. Mellmer *et al.*, Solvent-enabled control of reactivity for liquid-phase reactions of biomass-derived compounds. *Nature Catalysis* **1**, 199-207 (2018).
62. S. H. Mushrif, S. Caratzoulas, D. G. Vlachos, Understanding solvent effects in the selective conversion of fructose to 5-hydroxymethyl-furfural: a molecular dynamics investigation. *Physical Chemistry Chemical Physics* **14**, 2637-2644 (2012).

Chapter 9. Conclusions, future work and outlook

In this dissertation, we have investigated solvent effects in acid-catalyzed biomass conversion reactions in mixtures of water with organic cosolvents (*i.e.,* mixed solvent environments). In this context, we have focused on two important, acid-catalyzed reaction classes: dehydrations, which are used to remove oxygen from biomass-derived molecules in the form of water, and; hydrolyses, whereby etherified linkages in biomass-derived polymers (cellulose and hemicellulose) are cleaved to yield constituent monomer units (sugars). To this end, we have employed a combination of experimental reaction kinetics measurements, spectroscopic characterizations and molecular simulations. In doing so, we have established important molecular-level insights as to how water controls chemical reaction rates and selectivities in some of these acid-catalyzed processes. Furthermore, we have shown how these effects can be understood in terms of broadly generalizable principles; namely, preferential solvation and solvation free energies. We have probed these effects using molecular dynamics simulations, and developed model-predictive tools to aid in the rational design of mixed solvent environments for biomass conversion reactions. Finally, we have shown how solid-state NMR can be used to monitor the solvent-mediated decomposition of whole biomass over acid catalysts by quantifying the extent to which linkages between crystalline cellulose microfibrils and extracellular material (hemicellulose and lignin) are destroyed as a function of solvent composition, reaction time and temperature. Finally, we have explored how the destruction of these cellulose-extracellular linkages improves the enzymatic digestibility of cellulose.

In Chapter 3, we explored the role of water in the acid-catalyzed isomerization of LGO to HMF.(*1*) The purpose of this investigation was to establish whether this isomerization was partially responsible for the observed behavior that, upon the addition of a small amount of water

to a THF solvent, the major product afforded from the acid-catalyzed decomposition of cellulose changes from LGO to HMF. We showed that, in an ~1 wt% solution with pure water at room temperature, LGO undergoes a double hydration to afford a geminal diol (DH) in near quantitative yields. We demonstrated that this interconversion between LGO and DH is reversible, so that at elevated temperatures the conversion of LGO to DH does not proceed to completion. We established that, in the presence of a sulfuric acid catalyst at 125º*C*, LGO and DH are converted into HMF, which in turn is converted into levulinc acid and heterogenous, water-insoluble polymeric materials (humins). At this temperature, the rate at which LGO and DH are interconverted is fast compared to the rate of HMF formation, so that the kinetics of HMF formation can be described as first-order with respect to the combined concentration of LGO and DH in solution. Accordingly, we lowered the reaction temperature and added 50 wt% THF cosolvent to the reaction mixture, slowing the rate of LGO and DH interconversion, so that the rate of HMF formation could be measured when the former reaction is far from equilibrium. We showed that, under these condition, the rate of HMF formation is proportional to the concertation of LGO in solution, but not the concentration of DH. However, in a pure THF solvent without additional water, no HMF is formed from LGO. Together, these two results indicated LGO is converted to HMF, not DH, but that water is required to facilitate the reaction. We proposed a mechanism consistent with these observations, whereby water hydrolyzes the anhydro bridge of LGO, allowing for the resulting hemiacetal to undergo a ring-opening tautomerization to a key reaction intermediate which. Upon the removal of a water, this proposed intermediate would then yield HMF.

Following these insights, we concluded that the isomerization of LGO to HMF is slow in the presence of small amounts of water (~ 1-5 wt%), and therefore cannot explain the observation

that HMF is the primary product of cellulose dehydration in water/THF mixtures, while LGO is the major product in pure THF. Accordingly, in Chapter 4, we interrogated the effects of water content, organic cosolvent properties and different acid catalysts on the conversion of cellulose to LGO and HMF in more detail.(*2*) We confirmed that, as water is titrated into a pure THF solvent, the selectivity of cellulose conversion over acid catalysts gradually shifts from LGO to HMF. Based on reactivity studies of model compounds, we established the full reaction network underlying the acid-catalyzed cellulose conversion. In doing so, we showed that LGO and HMF are both formed from cellulose, passing through LGA and glucose as intermediates, respectively. Importantly, as water is added to a THF solvent, LGA is reversibly hydrolyzed to glucose, which partially explains the increased rate of HMF formation from cellulose in the presence of water. However, based on an analysis of initial rates, we showed that the presence of water enhances the intrinsic reactivity of glucose towards HMF to a greater extent that the reactivity of LGA towards LGO, which fully explains the aforementioned behaviors. Using technoeconomic analysis, we demonstrated that these water-mediated behaviors have important implications for process design purposes, whereby the cost of water/THF separations and the combined yields of LGO and HMF as a function of solvent composition must be considered together to identify minimal operating costs.

Having demonstrated that the composition of the solvent system and the performance of biomass conversion processes are linked via the relative rates the underlying chemical reactions, in Chapter 5 we explored the molecular bases by which mixed solvent environments control the rates of acid-catalyzed reaction of biomass-derived model compounds.(*3*) To this end, we measured the forward rates of acid-catalyzed dehydrations and hydrolyses of biomass-derived oxygenates, for example, xylitol dehydration to afford 1,4-anhydroxylitol, and ethyl-tert-butyl

ether hydrolysis to afford tert-butanol and ethanol. We collected apparent rate constants describing the kinetics of these reactions in pure water, and in mixtures of water with THF, GVL and dioxane. We found that, as the water content of the solvent system decreases, those reactants with fewer vicinal hydroxyl groups become universally more reactive. We explain these solvent effects by using classical molecular dynamics simulations to quantify the size and properties of water-enriched local solvent domains that nucleate in the vicinity of hydrophilic reactants which, owing to the acidic proton's affinity for water relative to the organic phase, effectively co-locates reactants and protons in solution, enhancing acid-catalyzed reaction rates. We developed a model, constituted by linear combinations of MD-derived observables, which describes the rates of acid-catalyzed reactions of all seven model compounds as a function of the water content of the solvent system. In doing so, we establish that solvent effects in acid-catalyzed biomass conversion reactions are at least partially encoded in the positions and momenta of the water and organic cosolvent molecules alone. When combined with minimal experimental measurements, classical MD therefore provides a basis to anticipate these solvent effects without detailed knowledge of the reaction mechanism or computationally intensive methods. While facilitating rapid comparison of solvent systems in a computationally efficient framework, however, an important limitation of this approach is the inability to explicitly consider the selectivity each reaction step to a single product.

In Chapter 6, we address the aforementioned limitation by exploring the effects of mixed solvent environments on the selectivity of reactions occurring in parallel. (*4*)To this end, we measured the rates of 1,2-propanediol (PDO) conversion in water, and in mixtures of water with GVL, dioxane, NMP, THF, TMSO and DMSO. We found that in all solvent system but DMSO, propanal was the major dehydration product afford from this reaction. In DMSO, however, we

found that acetone was formed as the major dehydration product. We explained these solvent effects by demonstrating that, as DMSO is added to a pure water solvent, the rate of propanal formation is suppressed by nearly two orders of magnitude, while the rate of acetone formation is promoted to a smaller extent. Accordingly, we used MD-enabled perturbation free energy calculations to quantify the effect of mixed solvent environments on the reaction free energy change doing from PDO to propanal, finding that increases in this reaction free energy change correspond universally to decreases in the rate and selectivity of propanal formation across all solvent systems tested. To explain the unique behavior that acetone in formed from PDO in DMSO/water mixtures, we used three-dimensional spatial distribution maps to show how DMSO is the only cosolvent amongst those tested that competes for hydrogen bond donating facets of the hydroxyl groups on PDO. Following this insight, we proposed a mechanism whereby DMSO participates directly in the reaction mechanism that affords acetone from PDO. Together, these insights demonstrate how MD-enabled characterization of the solvation free energy changes can capture trends in selectivity for biomass conversion processes that afford multiple products in parallel. This approach therefore provides a method to compare large numbers of possible water/cosolvent systems for biomass conversion process, where the selectivity to a desired product is considered explicitly. Importantly, as evidenced by the unique behavior of DMSO/water solvent systems in the acid-catalyzed dehydration of PDO, the efficacy of this technique is limited to those solvent systems wherein solvent molecules do not participate directly in the reaction mechanism. This limitation is due to the fact that classical MD tools cannot capture the covalent bond forming or breaking events that constitute these interactions, nor model reaction kinetics directly.

In Chapter 7, we combined the aforementioned experimental and computational tools into a workflow for screening mixed solvent environments for biomass conversion processes using a

minimal combined amount of experimental and computational resources. We have demonstrated the usage of this workflow by analyzing two case studies: the dehydration of cyclohexanol into cyclohexene and the acid-catalyzed partial dehydration of fructose to 5-hydroxymethylfurfural. From both case studies, the computational tools. These results demonstrate the general bases by which solvent-solute interactions drive reactivity in mixed solvent environments, and how these understandings can be distilled into computationally efficient screening methods and predictive design tools that allow for the rational design of mixed solvent systems for biomass conversion applications. Together then, the results presented in Chapters 5, 6 and 7 represent a step toward a molecular-level understanding the role of mixed solvent systems in controlling reactivity of liquid phase reactions, and toward the rational design of mixed solvent systems in a framework that significantly reduces the experimental burden that typically accompanies the development of new chemical processes.

In Chapter 8, we demonstrated how mixtures of water and organic cosolvents can be used to pretreat lignocellulosic biomass, so that the characteristic dimensions of the crystalline cellulose domains remain unchanged, but the intimate linkages between the exterior of the cellulose microfibrils and the lignin/hemicellulose matrix are destroyed, exposing surface-accessible cellulose chains on the outsides of the microfibrils. The extent to which these surface-accessible cellulose chains are liberated from the lignin and hemicellulose is observable by solid state NMR, and universally predicts the yields of enzymatic sugars afforded by hydrolysis of the residual cellulose, regardless of biomass type or the organic pretreatment solvent used. These NMR-based analyses therefore serve as a screening tool to compare solvent-mediated biomass pretreatment techniques, differentiating the most effective strategies on the basis NMR-observable metrics.

9.1. Future Work

The key insights established in this dissertation are: (1) that experimental reactivity measurements can be combined with computationally efficient molecular dynamics simulations to probe solvent effects in biomass conversion processes, providing broad insight as to how solvents control acid-catalyzed processes, and; (2) that spectroscopic methods can be used to characterize solvent effects in the conversion of real biomass-derived materials. Going forward, we recommend two corresponding areas for future work.

9.1.1. Extending the solvent design tools developed in this dissertation to processing realistic biomass-derived feedstocks

First, we have shown that experiments and MD-enabled, model-predictive tools can be combined to design solvent systems to maximize selectivity to desired products for simple, model reaction sequences. In future work, these tools should be brought to bear to design improved solvent systems for processing actual, biomass-derived materials under realistic process conditions. For example, lignin and pyrolysis oils are two aromatic-containing materials derived from biomass, but their high oxygen content and acidity make them unusable as a fuel. Furthermore, lignin and pyrolysis oils are complex mixtures composed of thousands of distinct compounds. As these materials consists primarily of phenolic residues, it has been proposed that identifying conditions under which phenols are selectively reduced to yield alkanes and aromatics should generate insight as to what reaction conditions are best suited for lignin and bio-oil conversion into liquid fuel precursors. As we have discussed elsewhere, this can be a tacit and

problematic assumption in general, given the complex chemical nature of biomass derived feedstocks.(*5*)

However, Rinaldi and coworkers *were* able to use phenol reduction as a probe reaction to calibrate a new process for reducing the oxygen content in both lignin and bio-oil feedstocks.(*6*) In a hexadecane solvent, over Raney Ni and H-BEA with isopropanol as a hydrogen source, the yields of benzene, cyclohexene and cyclohexane from phenol were tuned by modulating the catalyst properties, reactant to hydrogen source mass ratios, and other conditions. It was found that the choice of acid catalyst in part determined the product distribution derived from this process, indicating that selectivity might be controlled by the acid-catalyzed step. In other work, Sels and co-workers demonstrated the polarity of the solvent system exhibited a significant effect on the fraction of soluble lignin-derived products in the catalytic fractionation of whole biomass, indicating that the presence of water or other polar co-solvents might likewise effect the stability of desired products derived from these processes.(*7*) Finally, it has been shown that a combination of solution-state NMR, GPC, and Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR MS) can be used to propose a model structure to describe the composition of native lignin in woody biomass. It is then possible to characterize the lignin structure after pyrolysis, and after catalytic hydrogenation of the pyrolysis products derived therefrom. Such analyses show that there are both catalytic *and homogeneous reactions* that are occurring during the lignin hydrogenation process, the former of which includes hydrogenation of aromatic, alkene and carbonyl functionalities that are preserved during the pyrolytic depolymerization process.(*8*) The conclusions from this work are in contrast to the more limited insight that is garnered when simple, lignin-representative model compounds like phenols are studied.

Following these points, in future work, we suggest that a combination of reactivity measurements and high-resolution spectrometric and spectroscopic methods (such has FT-ICR MS and heteronuclear correlational NMR) should be used to propose model reaction networks underlying the acid-catalyzed conversion of real biomass-derived materials such as cellulose, lignin and pyrolysis oil. Thereafter, molecular-dynamics enabled perturbation free energy calculations (or other computational tools) could be employed to quantify the effect of mixed solvent environments on the stability of key species lying intermediate between the raw materials and desired products in the reaction network, or representative structures approximating the same. In doing so, we suggest that novel solvent systems could be developed for processing real biomass derived materials into value-added products using computationally-enabled, rational design approaches similar to those developed in this dissertation.

A related detail that differentiates model studies of biomass-derived compounds and the processes better representing real biomass conversion conditions is the concentration of the reactant species in solution. It should be noted that, Chapter 5, 6 and 7 of this dissertation, the concentration of the reactant was typically less than ~1 wt%. This allowed for key assumptions to be made in analyzing the molecular dynamics results, namely that interactions of dissolved solutes with themselves are dominated by their interactions with solvent molecules. In real processes, the concentration of reactant species in the liquid phase is typically ~10 wt% or greater. Accordingly, in future work, the principles established in this dissertation should be extended to include the effect of increased reactant loadings on the rates and selectivities of acid-catalyzed biomass conversion reactions. In this context, we expect that the same principles of solvation free energies and preferential solvation discussed herein could be more or less directly applied to solvent systems containing higher concentrations of reactant molecules. For example, inclusion of reactant

molecules in a classical MD simulation environment to approximate a 10 wt% solution and calculating the solvation free energy change required to add a single, additional reactant molecule would be straightforward. From an experimental perspective, care should be taken to account for changes in apparent reaction orders and transport limitations to ensure that measured rates can be analyzed in a context that allows for intrinsic reactivity to be characterized.

9.1.2. combining molecular simulations, reactivity measurements and advanced spectroscopic characterizations to study the fundamental details underlying the liquid-phase decomposition of real, solid biomasses.

We showed in Chapter 8 of this dissertation that solid state NMR can be used to quantify the extent to which key structures in solid biomass are converted in mixed solvent environments, namely the linkages between cellulose microfibrils and extracellular lignin and hemicellulose. In this context, the effectiveness of solvent-mediated pretreatments can be compared on the basis of different biomass types, and different solvent compositions.

Accordingly, in future work, detailed structure-property relationships that connect the composition and structural features of different biomasses with their susceptibility to solvent-assisted pretreatments at a fixed solvent composition should then be developed. To this end, advanced spectroscopic and X-ray scattering techniques should be used to better resolve the structural details that describe real biomass both before and after solvent-assisted pretreatment. For example, Alam and co-workers showed that pulsed-field gradient (PFG) HR/MAS NMR can be used to probe the dynamics of water (as expressed by self-diffusion coefficients) within the pore structure of semi-crystalline polymeric materials prepared by different methods.(*9*) These

methods could be applied to analyzing the dynamics of water in raw and partially-decomposed biomass, providing insight as to the structural changes that occur during the solvent-mediated pretreatment of biomass, and therefore in the early stages of biomass decomposition over acid catalysts. Additionally, the acid-catalyzed depolymerization of cellulose a can be studied using double-quantum relaxometery, which has been used to measure the strength of non-covalent interactions (hydrogen bonds) between adjacent polymer chains in semi-crystalline porous materials.(*10*) Also, HSQC and HMBC NMR can be used to identify the key chemical linkages between cellulose and lignin or hemicellulose in the cell wall.(*11*) Finally, grazing-incidence X-ray scattering can be used to probe fine structural details related to the crystallinity of cellulose,(*12*) and test whether these structures (as a function of different biomass types) can be differentiated with respect to their susceptibility to solvent-mediated disruption of their constituent hydrogen bonding networks, or depolymerization by hydrolytic enzymes.

Regarding the role of organic cosolvents in mediating solid biomass decomposition, molecular-level understandings of how semi-aqueous organic solvent mixtures facilitate the decomposition of cellulose-extracellular linkages or disrupt the hydrogen bonding networks which constitute crystalline cellulose domains should be probed using reactivity measurements and detailed, atomistic simulations. For example, MD simulations have been used to characterize local phase separations of water and organic cosolvents at biomass-relevant interfaces, providing insight into the role of THF/water mixtures in improving the solubility of cellulose over pure THF or pure water alone.(*13*) Additionally, the reactivity of model compounds representing cellulose-extracellular moieties should be probed over acid catalysts in mixed solvent environments using experimental reaction kinetics studies and MD investigations, as described in elsewhere in this dissertation. Together, these new insights would provide a framework to effect the rational design

of new solvent systems for pretreating lignocellulosic biomass, where the properties of the solvent system are deliberately tuned to target the cellulose-extracellular linkages in a particular biomass type.

9.2. Outlook

A central theme of this dissertation is that mixtures of water with organic cosolvents can preferentially enhance the rates of desirable reaction steps in acid-catalyzed biomass conversion processes, and that these solvent effects can be understood in terms of broadly-applicable bases (*e.g.,* preferential solvation) that can be leveraged for design purposes, using few experiments and computationally efficient molecular simulations. However, due to their high cost, organic solvents must typically be recycled with near 100% efficiency for industrial-scale processes to be economically viable, which is seldom realistic. Therefore, despite their potential to improve the yields of biomass conversion reactions, it should be noted that not a single existing industrial process for converting biomass into commodity chemicals uses non-aqueous solvents. Accordingly, a long term research goal to consider in the near future is how desirable aspects of solvent effects in biomass conversion processes can be engineered into chemically reacting systems without the need for large volumes of organic cosolvents. For example, recent work by Sun *et al.* has demonstrated that organic cosolvent moieties (such as dimethyl sulfoxide) can be integrated into heterogeneous solid frameworks alongside Brønsted-acidic species, resulting in immobilized solvation environments around the acid sites that effect high yields of HMF from the dehydration of fructose.(*14*) It should be noted that the reaction solvent in this study was tetrahydrofuran (THF), rather than water. In future work, solid materials that mimic solvation

effects for biomass conversion should be engineered to be hydrothermally stable, so that water can be used as a reaction solvent. Alternatively, one could consider the work of Scott and co-workers, who used operando MAS NMR to demonstrate that water and glucose partition from a bulk mixture with GVL into the pores of hydrophilic zeolites, enhancing the rate of Lewis-acid catalyzed glucose conversion within the pores relative to the same reaction over the same zeolites in pure water.(*15*) Following this insight, one can envision a family of Brønsted-acid, hydrophobic materials into which organic cosolvents would partition from a bulk, semi-aqueous solution, creating local solvation environments that are enriched in the organic phase (similar to 90+ wt% cosovlent mixtures studied herein) while only requiring a small amount of organic solvent to be present in the bulk. This would be accomplished, in principle, by grafting hydrophobic moieties into the pore structure of zeolitic materials, or synthesizing new porous materials from hydrophobic scaffoldings. These efforts would represent a rich research opportunity, whereby the properties of solid acidic materials could be tuned to deliberate modulate their wetting properties, and therefore control the local solvent composition (and therefore reactivity of acid sites) within the pores. These behaviors could be readily probed using computationally efficient MD methods, and their implications for the reactivity of dissolved reactant molecules should be tested using the reaction kinetics measurement techniques described in this dissertation.

A more general point to consider is that, throughout this dissertation, we have sought to establish the molecular bases by which solvents control the reactivity of biomass-derived compounds in the most broadly generalizable terms possible, so as to empower the rational design of new processes. This was accomplished by connecting the experimentally measured reactivity of model compounds with MD-observable properties of the solvated reactants in mixed solvent environments. Importantly, however, we are able to rationalize, or provide a physical interpretation

for these correlations between MD-observables and experimental reactivity measurements by drawing on fundamental, first-principles-based insights generated elsewhere. So, while the theme of this dissertation has been quantifying kinetic solvent effects in biomass conversion processes using minimal experiments and classical MD simulations, we in no way suggest that the development of the aforementioned design tools precludes the need to understand these same solvent effects in a fundamental, detail-oriented framework. Rather, we suggest that the detailed, fundamental insights garnered from experimental, mechanistic investigations and quantum chemical calculations should be built upon by considering how they might generalize beyond the isolated case studies wherein they are generated. In fact, we would like to stress that it is only by engaging with a first-principles level of theory that we have any frame of reference from which to understand the general bases underlying physical processes. Herein, we demonstrated this concept by analyzing solvent effects in biomass conversion reactions.

A final point to consider is the fledgling state of liquid-phase biomass conversion technologies, and the future of research efforts in this area. As we have described in Chapter 1 of this dissertation, 20+ years of concerted research efforts and large sums of capital investment have yielded few commercially viable technologies to convert biomass into usable products that are competitive with those currently derived from fossil resources. In some forums, these outcomes have brought the future of biomass conversion technologies and the associated, fundamental research activities into question. To contextualize this problem, however, we ask the reader to consider George Mitchel, who spent *decades* developing the technology to produce natural gas using hydraulic fracturing (or fracking) technologies.(*16*) His team knew that the natural gas was abundant in tight underground formations, but did not have the technology to extract it. Skeptics vocally and publicly criticized him for wasting time and resources on this complex and difficult

problem. In the end, however, this technology turned into the most important energy advancement in the 21st century. Today, we find ourselves in a similar situation as George Mitchell was in during the 1990s: we know that biomass is available, abundant, and that it has energy content. What we are missing are the catalytic technologies to convert the biomass into fungible fuels and commodity chemicals. However, our society has no other short-term, realistic or economically viable options to produce sustainable liquid fuels and chemicals than from plant biomass. It is therefore our belief that, as an international research community, we can (and must) work together to overcome the long-term scientific challenges outlined herein, and develop technologies that will reduce our society's dependence on fossil resources. As demonstrated in this dissertation, we expect that the role of solvents will be a key facet of solving the fundamental problems underlying these challenges, especially with respect to their ability to control the rates and selectivities of acid-catalyzed biomass conversion processes.

9.3. References

1. S. H. Krishna, T. W. Walker, J. A. Dumesic, G. W. Huber, Kinetics of levoglucosenone isomerization. *ChemSusChem* **10**, 129-138 (2017).
2. D. M. Alonso *et al.*, Increasing the revenue from lignocellulosic biomass: Maximizing feedstock utilization. *Science advances* **3**, e1603301 (2017).
3. T. W. Walker *et al.*, Universal kinetic solvent effects in acid-catalyzed reactions of biomass-derived oxygenates. *Energy & Environmental Science* **11**, 617-628 (2018).
4. T. W. W. A. K. Chew, B. Demir, L. Witteman, J. Euclide, G. W. Huber, J.A. Dumesic, R. C. Van Lehn, "Effect of mixed-solvent environments on the selectivity of acid-catalyzed dehydration reactions". *ACS Catalysis* **(in press)**, (2019).
5. T. W. Walker, A. H. Motagamwala, J. A. Dumesic, G. W. Huber, Fundamental catalytic challenges to design improved biomass conversion technologies. *Journal of Catalysis* **369**, 518-525 (2019).
6. X. Wang, R. Rinaldi, A route for lignin and bio-oil conversion: dehydroxylation of phenols into arenes by catalytic tandem reactions. *Angewandte Chemie International Edition* **52**, 11499-11503 (2013).

7. T. Renders *et al.*, Synergetic effects of alcohol/water mixing on the catalytic reductive fractionation of poplar wood. *ACS Sustainable Chemistry & Engineering* **4**, 6894-6904 (2016).
8. D. J. McClelland *et al.*, Functionality and molecular weight distribution of red oak lignin before and after pyrolysis and hydrogenation. *Green Chemistry* **19**, 1378-1389 (2017).
9. T. M. Alam, K. K. Childress, K. Pastoor, C. V. Rice, Characterization of free, restricted, and entrapped water environments in poly(N-isopropyl acrylamide) hydrogels via 1H HRMAS PFG NMR spectroscopy. *Journal of Polymer Science Part B: Polymer Physics* **52**, 1521-1527 (2014).
10. T. J. Simmons *et al.*, Folding of xylan onto cellulose fibrils in plant cell walls revealed by solid-state NMR. *Nature Communications* **7**, 13902 (2016).
11. T.-Q. Yuan, S.-N. Sun, F. Xu, R.-C. Sun, Characterization of lignin structures and lignin–carbohydrate complex (LCC) linkages by quantitative 13C and 2D HSQC NMR spectroscopy. *Journal of agricultural and food chemistry* **59**, 10604-10614 (2011).
12. H. M. Ehmann *et al.*, Surface-sensitive approach to interpreting supramolecular rearrangements in cellulose by synchrotron grazing incidence small-angle X-ray scattering. *Acs Macro Letters* **4**, 713-716 (2015).
13. B. Mostofian *et al.*, Local Phase Separation of Co-solvents Enhances Pretreatment of Biomass for Bioenergy Applications. *Journal of the American Chemical Society* **138**, 10869-10878 (2016).
14. Q. Sun *et al.*, Creating solvation environments in heterogeneous catalysts for efficient biomass conversion. *Nature Communications* **9**, 3236 (2018).
15. L. Qi *et al.*, Operando solid-state NMR observation of solvent-mediated adsorption-reaction of carbohydrates in zeolites. *ACS Catalysis* **7**, 3489-3500 (2017).
16. T. Fitzgerald. (JSTOR, 2015).