NORTHWESTERN UNIVERSITY

Computational Modeling for Biomass Pyrolysis Applications

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSPHY

Field of Chemical and Biological Engineering

By

Lauren D. Dellon

EVANSTON, ILLINOIS

September 2019

Computational Modeling for Biomass Pyrolysis Applications

Lauren D. Dellon

Biomass has the potential to be our country's leading renewable source of energy. Specifically, fast pyrolysis is a promising method for the conversion of biomass to valuable fuels and chemicals. Given that fast pyrolysis has a residence time of about two seconds, computational methods are particularly useful in obtaining product distributions and characterizing important reaction pathways. This dissertation presents the development and application of multiple computational methods used to investigate two major roadblocks in the production of acceptable fuel sources.

The first part of this work was the enhancement of a structure generation algorithm for producing libraries of representative structures of lignin for any biomass source. The added complexity allowed for the investigation of areas of feasible lignin space, that which includes all possible structures satisfying the experimental characteristics of monomer distribution, bond distribution, molecular weight distribution, and branching coefficient simultaneously. Additionally, these lignin libraries can subsequently be used in kinetic modeling studies and molecular simulations.

The second part of this work was the development of a detailed microkinetic model for the zeolitic upgrading of biomass pyrolysis vapors. An automated network generator was used to construct a reaction network, and kinetic and thermodynamic parameters were estimated from group additivity, transition state theory, and density-functional theory. The framework established can serve as a platform to investigate different model compounds, zeolites, and operating conditions.

Acknowledgements

To my advisor, Professor Linda Broadbelt: Thank you for making my decision to attend Northwestern University the best decision of my life. I could not have asked for a more understanding, patient, and knowledgeable mentor during my PhD study. You truly are an inspiration in every way possible, and I hope to make you proud in my future endeavors.

To my esteemed committee members, Professor Justin Notestein and Professor Eric Weitz: Thank you for your support and encouragement during my PhD study. I would not have succeeded without your thoughtful concerns and suggestions.

To my committee member and laboratory advisor, Dr. David Robichaud: Thank you for making my time at NREL fun and beneficial. I always appreciated and enjoyed our in-depth conversations. In addition, I am grateful for our monthly teleconferences that kept me on my toes and propelled my research forward.

To my experimental collaborators, Anne Starace and Calvin Mukarakate: Thank you for demonstrating and explaining your experimental procedures to me. The information was vital to my understanding of catalytic fast pyrolysis.

To the Broadbelt Research Group: I have been so lucky to have worked with all of you. From thoughtful discussions to lab lunches and outings, I couldn't have asked for a more supportive group of people to have by my side during this journey.

To my friends, especially Rob Brydon and Andres Martinez: Thank you for being there for me throughout this adventure. Your support has meant so much to me, and I can only hope that we will keep in touch in the future.

To my family, Mom, Dad, Rachel, and Michael: Words cannot express how grateful I am to have had your support during my PhD study. I love you with all my heart, and I hope I have made you proud. Also, I would feel remiss if I didn't mention my pug, Minnie, for always being there for me during the ups and downs of research.

Financial support: The author would like to acknowledge financial support provided by ExxonMobil Research and Engineering Company. Funding from the National Science Foundation (CBET-1435228) is also gratefully acknowledged. Additionally, Chapters 3 and 4 were conducted under subcontract AEV-6-62063-01 as part of the Consortium for Computational Physics and Chemistry supported by the U.S. Department of Energy's Bioenergy Technologies Office (DOE-BETO) Contract No. DE-AC36-08GO28308 with the National Renewable Energy Laboratory. Early support from the Institute for Atom-Efficient Chemical Transformations (IACT), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, and Office of Basic Energy Sciences, is also gratefully acknowledged. The author is also grateful for financial support from PPG for a fellowship.

To my beloved parents,

Steven and Patricia Dellon,

For your unconditional love and support,

Without whom none of my success would be possible.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction and Background

1.1. Introduction to Biomass Conversion

The non-renewable nature and adverse environmental impacts of petroleum, coal, and natural gas have motivated the development of renewable sources of energy, particularly biofuels, which can be produced from many different sources such as manure, crop waste, and woody biomass.[1-4] In addition to reducing greenhouse gas emissions and lowering levels of climate pollution, an increased utilization of biofuels provides the U.S. with economic security, as it reduces our dependence on foreign oil.[5-7] Unfortunately, the usage of biofuels is not without its disadvantages. Most notably, the production of biofuels has yet to be optimized, so as to compete with traditional fuels in terms of cost-effectiveness and efficiency. As such, significant research is being performed to understand and enhance the processes that produce biofuels.

While syngas from gasification[8-10] and ethanol from hydrolysis[11,12] are possible alternatives to fossil fuels, pyrolysis, specifically fast pyrolysis, is an especially promising strategy for the efficient conversion of biomass to liquid bio-oil. Fast pyrolysis, a thermochemical technique characterized by rapid heating, low residence times and the absence of oxygen, produces gases, char, and liquid bio-oil, a complex mixture of water and hundreds of organic compounds belonging to the classes of alkanes, aromatic hydrocarbons, phenol derivatives, and many more.[2,13-21] In addition to the easy transportability and storage of a liquid product, bio-oil is obtained in high yields, up to 75%,[14,22,23] and can be used in many different applications. Naturally, the optimization of biomass pyrolysis could be crucial in our transition to renewable sources of energy. However, given the vapor residence time of approximately two seconds, it has been extremely difficult to identify products experimentally. Through the use of computational modeling, product yields are

readily calculated, temporal behavior is described, and process variables are easily changed to determine the effects of pressure, temperature, and acidity.

Unfortunately, there are several obstacles standing in the way of biomass pyrolysis revolutionizing the fuel industry. In order to compete with traditional fuels, all components of the feedstock must be utilized. While cellulose, the most abundant component, has been researched extensively and utilized to its full extent, lignin, mostly used as process heat, continues to hamper the economic potential of biomass pyrolysis. One primary challenge lies in the structure of lignin, an amorphous, complex polymer with no regular repeating patterns.[24-28] Although experimental methods, such as NMR,[29] can characterize lignin, representative structures are required in order to use them as reactants in models of lignin fast pyrolysis, as shown in Figure 1.1a. As such, one goal of this work is to computationally develop representative structures of lignin.

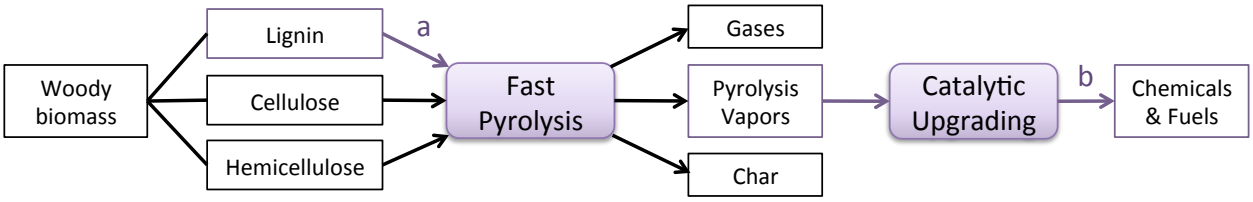


Figure 1.1. Route of woody biomass to chemicals and fuels with challenges labeled: (a) Generation of lignin structures in preparation for fast pyrolysis and (b) ex situ catalytic upgrading of pyrolysis vapors.

Another issue is the quality of the liquid bio-oil produced from the condensed pyrolysis vapors. Not only is it immiscible with fossil fuels due to its high water and oxygen content, it is also characterized by a high viscosity, instability, and corrosiveness, leading to storage issues.[1,2,13-16,30-32] Thus, it is necessary, through the use of carefully designed catalysts, to upgrade, primarily by deoxygenation, the pyrolysis product vapors before condensation into specialty chemicals and

fuels, as seen in Figure 1.1b. A major class of catalysts used for vapor phase upgrading is zeolites, which offer Brønsted acidic sites promoting acid-catalyzed chemistry.[2,14,15,33] In the case of catalytic fast pyrolysis (CFP) of biomass, this chemistry involves oxygen-containing carbenium ions (C^+) and oxonium ions (O^+) as chemisorbed intermediates.[2,34,35] Computational models, such as microkinetic models, can simulate the rapid, complex chemistry imparted by these chemisorbed intermediates at the individual species' level, as well as give insight into catalyst deactivation. Not only can these models be used to predict product distributions and investigate specific reaction pathways, but, most importantly, they can enable the design and synthesis of more effective catalysts as well as facilitate optimization of experimental process conditions. The second goal of this work is to develop a microkinetic model for the catalytic upgrading of acetone and acetic acid, two model compounds representing the product slate of biomass fast pyrolysis.

1.2. Background for Structural Models of Biomass

While the ultimate goal is to develop a detailed microkinetic model for whole biomass pyrolysis, current efforts have focused on generating models for fast pyrolysis of each of the three components of biomass – cellulose, hemicellulose, and lignin – separately. Thus, a representative structure for each of the components is required in computational form, so as to be the input for a kinetic model.

1.2.1. Cellulose

Cellulose, the most abundant component of biomass, accounting for 40-50 wt%,[18] has the most well-defined structure of all three components. It is a linear polymer comprised entirely of β-D glucopyranose units covalently linked with 1➔4 glycosidic bonds.[36] The abundance of hydroxyl groups leads to extensive hydrogen bonding, giving rise to the crystallinity and resistant

nature of cellulose. The structure of cellulose includes a non-reducing end, a reducing end, and repeating units ranging from 140 or less to 10000 or more depending on the source.[36] Fortunately, this allows cellulose chains to be classified and modeled as a combination of a left group, middle groups, and a right group.[18] However, such polydispersity has proven difficult to model and has prompted the use of continuous distribution kinetics, which explicitly incorporates the effect of chain length on the reaction rates and allows a molecular weight distribution of the polymer to be constructed at each time step.[37] Moment operations are necessary to transform the resulting reactor design integro-differential equations to ordinary differential equations (ODE's), which can then be coupled with low molecular weight product (LMWP) ODE's. The combination of ODE's and the relevant algebraic equations forms the basis for a continuum model.

1.2.2. Hemicellulose

Hemicellulose, the second most abundant component of biomass, accounting for 20-35 wt%, is a mixture of complex, branched and heterogeneous polysaccharides that are not characterized as being either cellulose or pectin.[38] Different assemblies of the functional groups, which include pentoses, hexoses, hexuronic acids and acetyl groups, lead to a range of diverse structures such as xylans, mannans, xyloglucans, β-1,3;1,4-glucans, and galactans.[39,40] Many initial kinetic models were ineffective in capturing this diversity by simply approximating hemicellulose as xylan, the predominant hemicellulose in hardwoods and straw.[41,42] However, hemicellulose can be more accurately represented by utilizing the same strategy used for cellulose; namely, defining a left group, middle groups, and a right group. This allows for the β-1,4-linked backbone exhibited by most hemicelluloses, as well as an adornment of functional groups to

encompass the wide variety of possible structures.[43] A continuum model can be constructed and solved in the same way as that for cellulose.

1.2.3. Lignin

One major concern in the utilization of biomass as a feedstock is the lack of cost-effective uses for lignin, which accounts for about 15-30 wt% of biomass.[44] To date, lignin has mainly been utilized as a low value fuel for process heat.[45] However, the aromatic nature and low oxygen-to-carbon ratio of lignin afford the possible production of specialty chemicals and fuels.[44] One primary challenge to lignin valorization lies in the severe uncertainty in its structure. As a complex polymer, it is composed of three monolignols, syringyl (S), p-hydroxyphenyl (P), and guaiacyl (G), connected by a variety of linkage types, giving rise to a hyper-branched topology.[24-26,28] As such, the abundance of these monomers and bonds are crucial determinants of the structure and reactivity of lignin. Adding to the difficulty in visualizing lignin is the diversity due to three main sources: biomass feedstock, isolation method, and characterization method. For instance, not only are the lignin molecules within a given biomass source not the same, there also exist notable differences between the three types of biomass: herbaceous, softwood, and hardwood. In addition, there is variation in the bond distribution, molecular weight distribution, and branching coefficient.

The diversity and breadth of experimental procedures have prompted the development of computational models to create representative lignin molecules, which could subsequently be utilized in kinetic models and molecular simulation. Although representations of a single lignin molecule have been put forth,[24,25,46,47] given lignin's diversity and variability, it is advantageous to develop a "library" of representative lignin molecules whose average properties match those of native lignin. Train and Klein were able to create libraries such as these that exhibited a linear

topology,[48] but it is well known that lignin exhibits a hyper-branched topology. Recently, Yanez et al. developed a stochastic method capable of generating libraries of structural representations of lignin with a hyper-branched topology.[49] However, for the purpose of method development, they focused their efforts on wheatstraw lignin and three bond types. This motivates the extension of the method to allow the creation of libraries of structural representations of lignin for any type of biomass, including softwood, hardwood, and herbaceous.

1.3. Background for Microkinetic Modeling

As mentioned, microkinetic modeling is used to simulate the kinetics for the catalytic upgrading of various oxygenates. The development of a microkinetic model is a complex process, outlined in Figure 1.2, that involves the automated generation of a reaction network, the estimation of key kinetic parameters, and the construction of reactor design ordinary differential equations.

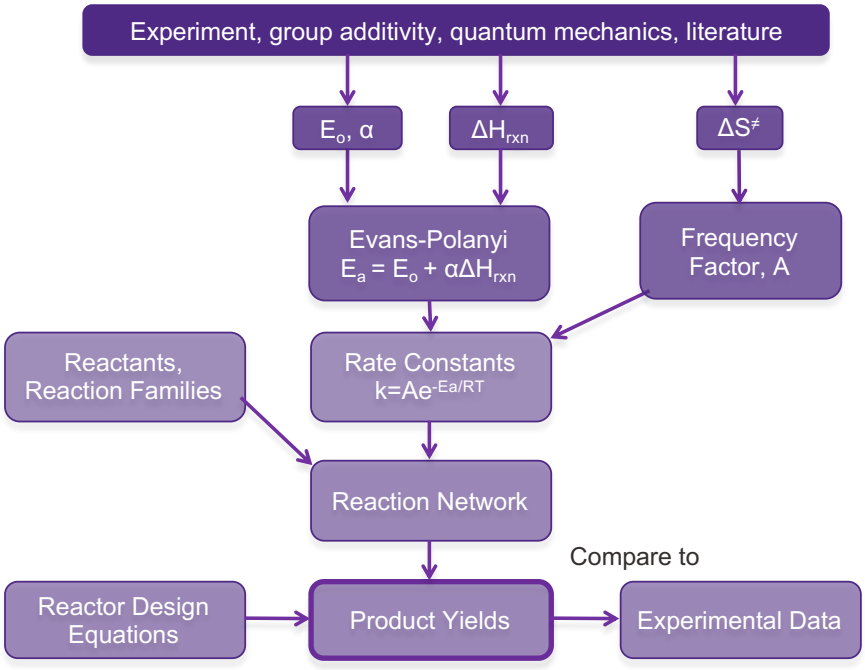


Figure 1.2. Flowchart for the development of a microkinetic model

Complex reaction networks can often contain thousands of different reactions, seemingly eliminating microkinetic modeling as a possibility to explore these networks. However, though the network is large, the types of reactions with similar chemistry that occur are usually small in number. This allows for the establishment of reaction families, drastically reducing the number of parameters required to fully define the system. In this work, NetGen, an automated network generator, will use the concept of reaction families to construct a reaction network.[50] A brief overview is as follows.

The modeler must input three components to the NetGen algorithm: the structure of the reactants, the reaction families and their rules of implementation, and kinetic correlations to provide estimates of rate constants. First, such a program requires the development of the representation of a chemical species, so as to be easily searched, modified, and stored. By implementation of graph theory concepts, a unique bond and electron (BE) matrix can be created, which describes not only the atomic connectivity of a molecule but also its electronic state.[50] As only a small number of atoms are actually participating in the reaction, a reduced BE matrix for the reaction is constructed. These same graph theory and BE concepts can be implemented to form a reaction transformation matrix, such that when it is added to the reduced BE matrix, a product matrix is formed. Figure 1.3 displays a hydride transfer reaction of a reduced BE matrix added to a reaction transformation matrix, one of which is defined for each reaction family.

Kinetic parameters, such as heats of reaction and Arrhenius parameters, are required in order to be able to calculate rate constants and thermodynamics for the reactions. NetGen outputs a description of each reaction by providing kinetic and thermodynamic parameters obtained from

the kinetic correlations input by the modeler. With a reaction network and the rate constants in hand, they can be interfaced with automatic, commercial solvers of choice.

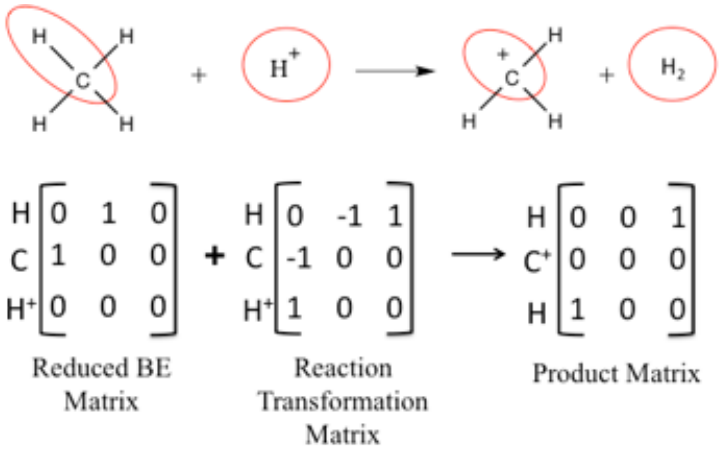


Figure 1.3. Hydride transfer as a bimolecular reaction example of NetGen. Atoms circled in red are those participating in the reaction and used to create the reduced BE matrix.

1.4. Outline of Research

Chapter 2 tackles the first goal of this work, which was to extend a stochastic method for the structure generation of lignin to accommodate more complexity and any type of biomass. The unique mechanistic details for several of the new lignin bond types were essential in deciding rules for bond formation in the algorithm. Apart from generating libraries of lignin structures, the added complexity allowed for the exploration of "lignin space", which is coined to represent all possible structures of lignin given the experimental characteristics of monomer distribution, bond distribution, molecular weight distribution, and branching coefficient. Using the overall approach, lignin libraries for any biomass source with reliable and consistent experimental data can be generated for future kinetic modeling studies or molecular simulations, and guidance can be provided to experimentalist to design and characterize lignin.

The ensuing chapters detail the second goal of this work, namely the development of a microkinetic model for the catalytic upgrading of various oxygenates. This process requires rate constants for the elementary reactions. The parameters in the Arrhenius equation can be related to thermodynamic properties through structure-reactivity relationships, such as the Evans-Polanyi relationship. For this relationship, enthalpies of formation of each species are required, which can be reasonably estimated using group additivity. However, the literature previously lacked group additivity values for oxygenates, oxonium ions, and oxygen-containing carbenium ions. In Chapter 3, 71 group additivity values for these types of groups were regressed, 65 of which had not been reported previously and five of which were newly estimated based on regression in the context of the 65 new groups. Heats of formation based on atomization enthalpy calculations for a set of reference molecules and isodesmic reactions for a small set of larger species for which experimental data was available were used to demonstrate the accuracy of the Gaussian-4 quantum mechanical method in estimating enthalpies of formation for species involving the moieties of interest. Isodesmic reactions for a total of 195 species were constructed from the reference molecules to calculate enthalpies of formation that were used to regress the group additivity values.

In Chapter 4, the details of the microkinetic model of the catalytic upgrading of vapor-phase oxygenates are elucidated. An automated network generator was utilized to apply relevant reaction families to the reactants, constructing a kinetic network comprising 580 unique species and 2,160 unique reactions. The kinetic parameters for the network were estimated using transition state theory, the Evans-Polanyi relationship, and thermodynamic data. The resulting mechanistic model is able to describe the experimental data presented in the literature for the transformation of acetic acid and acetone on HZSM-5 in a fixed-bed reactor. Additionally, the model solutions reveal

vital information regarding the mechanism by which acetic acid and acetone are upgraded to valuable fuels and chemicals.

Finally, Chapter 5 summarizes the relevant results and conclusions from this work. In addition, future directions are proposed. The structure generation program has the potential to be coupled with kinetic models or molecular dynamics simulations. The microkinetic model for catalytic upgrading can be further enhanced and applied to different catalysts.

Chapter 2: Computational Generation of Lignin Libraries

Material in this chapter is reproduced from the publication "Computational Generation of Lignin Libraries from Diverse Biomass Sources" by Lauren D. Dellon, Abraham J. Yanez, Wenjun Li, Ross Mabon, and Linda J. Broadbelt; *Energy & Fuels* **2017**, *31*(8), 8263-8274.

2.1. Introduction

Lignocellulosic biomass has the potential to serve as the renewable resource that propels the US toward its renewable and sustainable energy goals. One major stipulation in the use of biomass as a feedstock is a cost-effective use for lignin, which accounts for about 15-30 wt% of biomass. To date, lignin has mainly been utilized as a low value fuel for process heat. However, the aromatic nature and low oxygen-to-carbon ratio of lignin provide the potential for production of specialty fuels and chemicals.[51-53]

One primary obstacle to lignin valorization is the severe uncertainty in its structure. Lignin is an amorphous, complex polymer with no regular repeating patterns. It is primarily composed of three monomers, syringyl, *p*-hydroxyphenyl, and guaiacyl, connected by a variety of linkage types, giving rise to a hyper-branched topology, meaning branches within branches.[24-26,28] The abundance of these monomers and linkage types are crucial determinants of the structure and reactivity of lignin. Adding to the difficulty in understanding lignin structure and reactivity is the diversity due to three main sources: biomass feedstock, isolation method, and characterization method.

There are three main types of biomass: herbaceous, softwood, and hardwood, each having its own key structural characteristics. Softwood is primarily made up of the guaiacyl monomer, hardwood is made up of approximately even amounts of syringyl and guaiacyl, and herbaceous lignin has all three monomers.[29,54] In addition, there is variation in the bond distribution, molecular

weight distribution, and branching coefficient. Thus, not only are the lignin molecules within a given biomass source not the same, notable differences also exist between the biomass sources.

The isolation procedure can also have a significant effect on the structure of lignin. The most common procedure was developed by Björkman,[55] in which wood is suspended in toluene, milled, and extracted with 96% dioxane. The resulting lignin is known as Björkman lignin or milled-wood lignin (MWL). While other isolated lignins exist, such as cellulolytic enzyme lignin (CEL) and enzymatic mild acidolysis lignin (EMAL), MWL is considered to be representative of whole wood lignin and contains fewer carbohydrates than the other two types. However, the milling in all the procedures significantly degrades the lignin, leading to a decrease in the molecular weight and an increase in the content of carbonyl and phenolic hydroxyl groups.[56,57]

Finally, there has been a wide variety of procedures for lignin degradation and subsequent quantification of monomer and bond types. Throughout several exploratory experiments, Erickson and co-workers estimated the relative abundance of monomer and bond types in lignin by performing permanganate oxidative degradation.[58-61] In this procedure, major carboxylic aromatic acids are obtained by oxidation of the aliphatic side chains of methylated MWL. Another wet chemistry technique that has been utilized is acidolysis, i.e., degradation of lignin by refluxing with hydrochloric acid.[62] A combination of these wet chemistry procedures is needed to target specific lignin moieties and completely characterize lignin. On the other hand, spectroscopic methods, such as NMR, are efficient in achieving a detailed characterization of lignin moieties.[29,63-65]

This diversity and breadth of experimental procedures have prompted the development of computational models to create representative lignin molecules. Further, these lignin models could

subsequently be utilized in kinetic models and molecular simulation. Initial representations of lignin constituted one average lignin molecule.[25,47,66,67] However, given lignin's diversity and variability, it is advantageous to develop a "library" of representative lignin molecules, whose average properties match those of native lignin. Train and Klein were able to create libraries such as these that exhibited a linear topology,[48] but it is well known that lignin exhibits a hyper-branched topology. Recently, Yanez and co-workers developed a stochastic method capable of generating libraries of structural representations of lignin with a hyper-branched topology.[49] However, for the purpose of method development, they focused their efforts on wheatstraw lignin with only three bond types. This motivates the extension of the method to allow the creation of libraries of structural representations of lignin for all three types of biomass. The goal of this chapter is not to detail the main computational methods of developing libraries of lignin structures, as this was accomplished in the work by Yanez and co-workers,[49] but instead to acknowledge the underlying features of lignin and uncover intriguing relationships among physical characteristics. However, to aid the reader, brief summaries of the main concepts from previous work are available in Appendix A.

In this chapter, we extend the structure generation method developed by Yanez and co-workers[49] to include all sources of lignin, including softwood, hardwood, and herbaceous biomass. This extension requires a brief review of lignin biosynthesis to motivate the constraints that will be imposed when certain new bond types are added. This method also allows for an exploration of "lignin space", which includes all possible lignin structures given the four characteristics of biomass. Knowledge of lignin space has the power to guide scientists in future experiments and efforts to create "designer lignin" that has optimal structural and degradation characteristics.

2.2. Methodology

The method developed by Yanez and co-workers[49] and extended here seeks to recapitulate four crucial characteristics of native lignin: monomer distribution, bond distribution, molecular weight distribution, and branching coefficient. It has been demonstrated through several experimental methods, including acidolysis,[62] oxidative degradation,[61] and 2D-NMR,[29] that lignin is composed of three main phenylpropanoid units, or monomers. These are called syringyl, *p*-hydroxyphenyl, and guaiacyl, conveniently referred to as S, P, and G, respectively, and can be seen in Figure 2.1. However, this convention of lignin being composed primarily of three monomers is being questioned, due to the detection of *p*-coumarate and ferulate in plant lignins, predominantly grasses.[68,69] Ralph proposed to treat these hydroxycinnamates, seen in Figure 2.1, as additional monomers, since ferulates are incorporated into the lignin chain in the same way as the three traditional monomers, and *p*-coumarates act as terminal pendant groups on the lignin chain.[70] These monomers can easily be integrated into our structure generation program for feedstocks with significant amounts of these monomers, such as maize and elephant grass stems.[68,71] Seven linkages, or bond types, are predominant: β-O-4, β-5, 5-5, 4-O-5, β-1, α-O-4, and β-β, seen in Figure 2.2.[26,28,69,72] Note that possible methoxy substituents have been omitted from Figure 2.2 and the following Figures and Schemes in this chapter for simplicity. More recently, other cyclic bond types have been suggested to be present in lignin: dibenzodioxocin, spirodienone, and arylisochroman, seen in Figure 2.3. Although not explicitly included as bond types in our structure generation program, these bonds consist of several of the bond types seen in Figure 2.2 and must be considered when assigning bond distribution percentages, as will become apparent later. Dibenzodioxocin is an 8-membered ring consisting of a 5-5 bond, an α-O-4 bond,

and a β-O-4 bond that is easily detected by NMR. A majority of the 5-5 bonds in lignin are thought to occur in the form of dibenzodioxocin.[73-75] Arylisochroman is a structure consisting of a β-1 bond, an α-6 bond, and an α-O-α bond that was discovered as the result of a new β-1 coupling pathway.[73,76] Finally, it has been suggested that spirodienone, which consists of a β-1 bond, an α-1 bond, and an α-O-α bond, is present in native lignin but degrades to β-1 moieties in an acidic environment.[29,77]

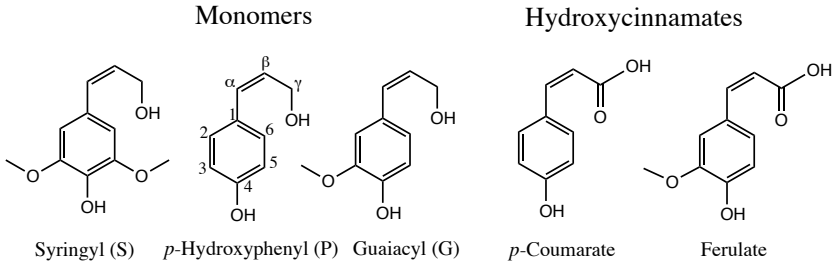


Figure 2.1. Three traditional monomers of lignin: syringyl (S), *p*-hydroxyphenyl (P), and guaiacyl (G) and two hydroxycinnamates, *p*-coumarate and ferulate. Note that some literature sources will abbreviate *p*-hydroxyphenyl as H, rather than P. The conventional labeling in lignin chemistry is displayed on *p*-hydroxyphenyl but can be applied to all three monomers. The monomers differ based on their degree of methoxylation *ortho* to the phenol substituent (the 3- and 5-carbon positions). Similarly, *p*-coumarate differs from ferulate by the lack of a methoxy group *ortho* to the phenol substituent.

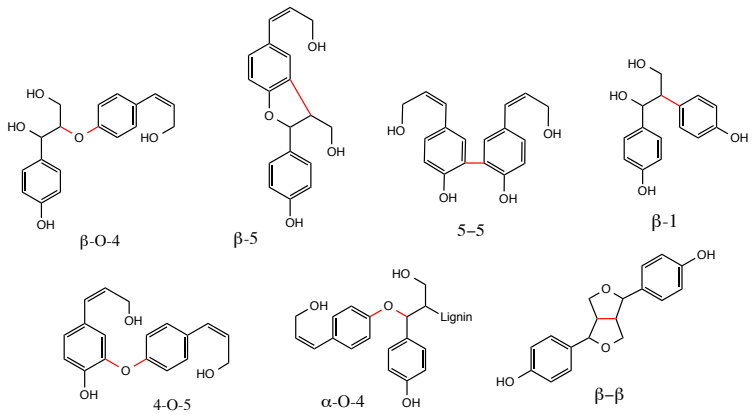


Figure 2.2. Seven major bond types used in this work constituting lignin: β-O-4, β-5, 5-5, β-1, 4-O-5, α-O-4, and β-β. The bond types are defined according to the conventional labeling presented in Figure 2.1.

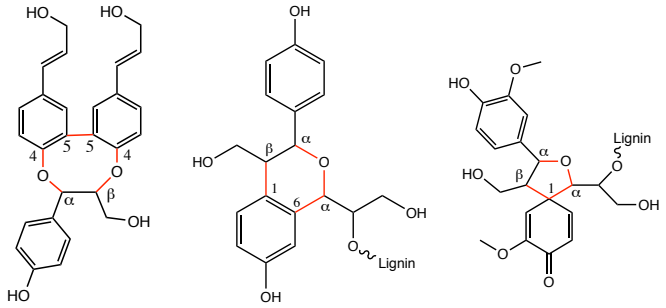


Figure 2.3. Special bond types detected in lignin. The bond types are defined according to the conventional labeling presented in Figure 2.1. These bond types are not explicitly integrated into our structure generation program. Rather, they have consequences for the abundance of 5-5 and β-1 bonds.

The molecular weight distribution is described by a truncated Schultz distribution, as outlined by Yanez et al.[49] For more information regarding the Schultz distribution, see Appendix A. Finally, correlations for the experimental branching coefficient, α, which reproduces the hyper-branched topology of native lignin, are available for western hemlock,[78] a softwood, and black cottonwood,[79] a hardwood. Figure 2.4 shows the five possible bonding locations on a monomer: β-C, α-C, 1-C, 5-C, and 4-O. When three or more of these sites are occupied, the monomer is considered branched. In this work, the simulated branching coefficient is defined as the number of monomers that are branched divided by the total number of monomers in the structure. For a brief discussion of hyper-branched lignin and the branching coefficient, refer to Appendix A. The method by which these four experimental characteristics are reproduced can be discretized into three different levels: formation of a single bond, formation of a single molecule, and formation of a lignin library. Each of these levels will be described in turn.

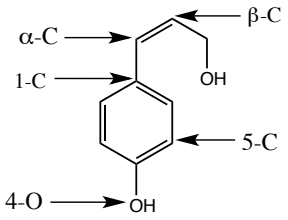


Figure 2.4. Bonding locations on a monomer: β-C, α-C, 1-C, 5-C, and 4-O. A monomer is considered branched when three or more of these sites are occupied. Note that *p*-hydroxyphenyl is shown here. Guaiacyl and syringl would have the same bonding locations as *p*-hydroxyphenyl, except that the 5-C position would not be a bonding location for syringyl because it is already occupied by a methoxy group.

2.2.1. Formation of a Single Bond

While the algorithm for lignin structure generation synthesizes lignin molecules, it does not directly simulate the process of lignin biosynthesis. However, the overarching goal is to create chemically sound structures that are representative of native lignin. Thus, knowledge of lignin biosynthesis is important in specifying allowed bonding patterns and prohibiting structures that cannot occur in nature.

A commonly accepted hypothesis in the literature is that lignin synthesis occurs by way of oxidative radical coupling.[24,29,80] Oxidation of the phenolic hydroxyl groups present on the three monomers results in a phenolic radical at the 4-O position, which is resonance-stabilized by radicals at the 5-C position, 1-C position, and β-C position. Coupling reactions are the basis of formation for the seven bond types in Figure 2.2, with the exception of the α-O-4 bond. For a majority of the bond types, the biosynthesis mechanism, while interesting, has no consequences for how the structure generation algorithm is formulated. However, the formation mechanisms for both the α-O-4 and β-1 bonds are quite unique and have important implications for the selection of allowed bonding types.

Given that an α-C radical does not exist, the α-O-4 bond must form via a route other than oxidative radical coupling. This pathway is the addition of a 4-O radical to a quinone methide, a common intermediate in several of the bond formation mechanisms. For example, Figure 2.5a depicts the formation of a quinone methide through the formation of a β-O-4 bond. An intermediate quinone methide is also produced in the β-β and β-1 mechanisms in a similar fashion. However, while intramolecular hydroxyl groups out-compete the 4-O radical to attack the quinone methide in the β-β and β-1 mechanisms, the intermediate quinone methide in the β-O-4 mechanism is susceptible to attack by a 4-O radical. Such an attack forms an α-O-4 bond,[24,53,73,81] as seen in Figure 2.5b. As previously noted, radical coupling cannot create the α-O-4 bond. Instead, the α-O-4 bond can only be created through the quinone methide intermediate of the formation of a β-O-4 bond, the sequence of the first reaction in Figure 2.5a followed by the reaction in Figure 2.5b. In other words, a β-O-4 bond must always precede an α-O-4 bond. The decision trees (see next section) are created to reflect this detail that α-O-4 could not be formed as an independent bond.

Another bond with a unique characteristic in the formation mechanism is the β-1 bond, seen in Scheme 2.5c. A β-O-4 structure with a radical at the 1-C position couples with a monolignol radical at the β-C position. In order to restore aromaticity, a group is required at the α-C position, usually a hydroxyl group. This allows for stabilization of the detached side chain. The products are the lignin chain with a glyceraldehyde-2-aryl ether end structure and a β-1 dimer, which can continue chain growth. However, due to the removal of a side chain, one of the monomers will lose its propanoid region.[24,73,82] It was important to reflect this information in the structure generation algorithm to ensure the balance of molecular weight and to prevent unrealistic structures with a non-detached side chain.

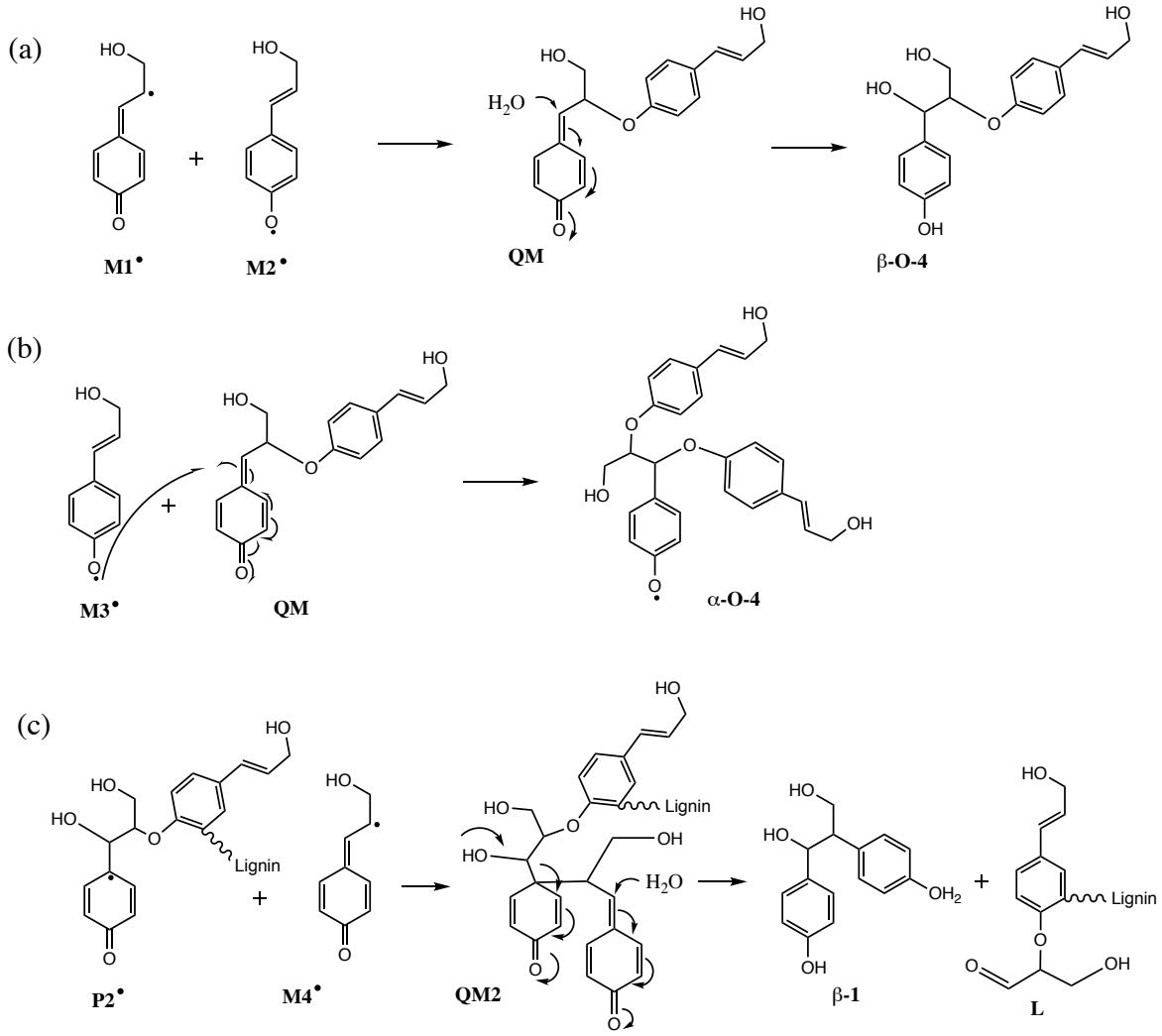


Figure 2.5. Bond formation mechanisms for (a) β-O-4, (b) α-O-4, and (c) β-1. Only P monomers are shown, but the same mechanisms apply to the S and G monomers. In (a), a monomeric β-C radical **M1•** couples with a monomeric 4-O radical **M2•** to form a quinone methide intermediate, **QM**. This intermediate is then re-aromatized to form the **β-O-4** linkage. In (b), a quinone methide **QM**, which can be produced as in (a), is aromatized via nucleophilic addition of a monomeric 4-O radical **M3•** to form an **α-O-4** bond as part of a polymeric radical. In (c), a polymeric 1-C radical **P2•** couples with a monomeric β-C radical **M4•** to form a quinone methide intermediate **QM2**. **QM2** is subsequently rearomatized by nucleophilic addition of water and the detachment of a side chain to form a **β-1** dimer, which can continue growth, and the original lignin chain **L** with a glyceraldehyde-2-aryl ether end structure. A wavy line is used to represent the continuation of the lignin chain.

2.2.2. Formation of a Single Molecule

An individual molecule (see Appendix A for a computational representation) is obtained by repeatedly sampling from decision trees encoding all monomers and allowable bonding patterns. The edges on the trees represent conditional probabilities of selecting a certain feature given the existence of a previous feature. More details are provided elsewhere.[49] The first decision trees developed will be presented, followed by a discussion of the two methods utilized to reduce the parameter space constituting the edge weights.

2.2.2.1. Initial Decision Trees

The decision trees for a general biomass source can be seen in Figure 2.6. The method by which these decision trees are sampled, branching is considered, and growth is terminated is explained by Yanez and coworkers.[49] The edge weights may take values between 0 and 1 and are subject to normalization constraints. For this set of decision trees, the number of edge weights requiring specification is 37. As there are 10 constraints (the experimental values of the monomer and bond distributions), the total number of degrees of freedom for this system is 27. These 27 edge weights, plus an additional adjustable parameter, the “branching propensity”, were optimized according to the stochastic optimization approach detailed by Yanez and coworkers[49] and summarized in Appendix A.

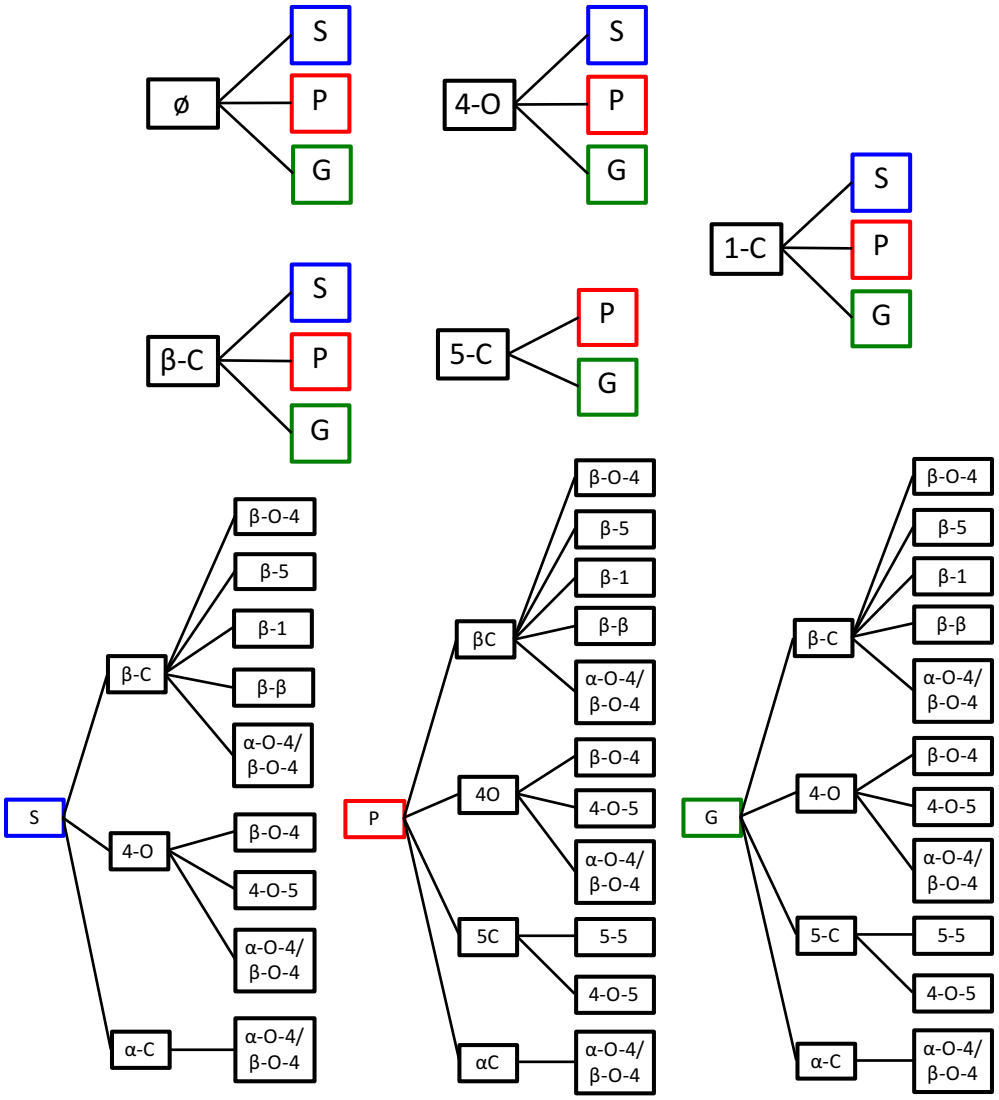


Figure 2.6. Original decision trees with 27 degrees of freedom. The nodes β-C, 4-O, α-C, 5-C, and 1-C are the bonding positions illustrated in Figure 2.4. The nodes S (syringyl), P (*p*-hydroxyphenyl), and G (guaiacyl) are the monomers illustrated in Figure 2.1. The nodes β-O-4, β-5, 5-5, β-1, 4-O-5, α-O-4/β-O-4, and β-β are the bonding patterns illustrated in Figure 2.2. The top five decision trees are monomer selection trees, i.e. a monomer is chosen given that a position on the monomer has been selected for bonding. The node containing ϕ represents the null condition, or the start of a lignin chain. Note that the 5-C decision tree does not include S as a possibility due to its pre-existing methoxy group at the 5-C position. The bottom three decision trees are position and bond selection trees, i.e. a position is selected on a given monomer and a subsequent bond is selected from that position. Again, note that the P and G trees are the same, while the S tree was amended to account for its 5-C methoxy group. Also, note that the α-O-4 bond is given as a combination of the α-O-4 and β-O-4 bonds, a direct result from the bond formation mechanism discussed earlier.

2.2.2.2. Reducing a Large Parameter Space

The abundance of bond types for the lignin space of interest leads to an immense parameter space. In addition, the random sampling technique utilized by Yanez and co-workers[49] was applied without any need to consider enhanced computational efficiency, given that only three bond types were explored previously. We present here two different methods by which the parameter space may be reduced and the efficiency of the edge weight optimization may be enhanced: reducing the number of degrees of freedom and biasing the edge weights.

In order to increase the efficiency of the edge weight optimization, we reduced the number of degrees of freedom by combining equivalent decision trees and removing redundant nodes, as can be seen in Figure 2.7. Whereas previously the conditional probability of selecting an S monomer given the 4-O position was different from the conditional probability of selecting an S monomer given the 1-C position, these two probabilities were set to be equal in the reduced decision tree. This reduction comes with no significant loss of information, as it has been shown that the edge weights are a vehicle for obtaining lignin structures that are consistent with experiment but do not have any fundamental link to the relative kinetics governing the chemistry of lignin synthesis. Additionally, while the α-O-4/β-O-4 combination can technically form from the β-C, α-C, and 4-O positions, these all lead to the same chemical structure. This assures that removal of the α-O-4/β-O-4 from the α-C and 4-O nodes does not limit the chemical structural space. For this reduced set of decision trees, the number of edge weights requiring specification is 17. Thus, the total number of degrees of freedom for this optimization problem is seven.

The random generation of 37 edge weights, all of which have the range (0,1), for the original decision trees makes obtaining optimum edge weight values within a reasonable

number of iterations very challenging. However, given the experimental data for the monomer and bond distributions, it is possible to bias the attempted edge weight values to be within a certain range. For instance, spruce lignin is comprised of 94% G, 5% P, and 1% S. It is unlikely, therefore, that any edge weights leading to an S monomer will be more than say, 10%. Using this method, a narrower range can be set for each of the edge weights. As a result, this decreases the set of possible solutions and increases the efficiency of the optimization.

Utilizing both methods, simultaneously can increase the efficiency further. That is, biasing the edge weights for the reduced decision trees is also explored in this work to evaluate its effect on the efficiency of edge weight optimization.

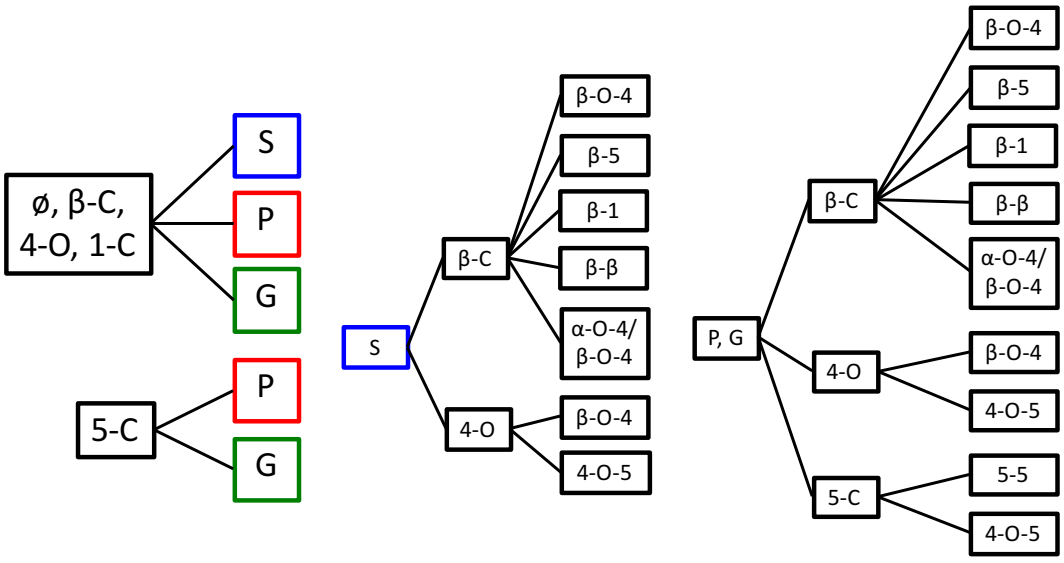


Figure 2.7. Reduced decision trees with seven degrees of freedom. The nodes β-C, 4-O, α-C, 5-C, and 1-C are the bonding positions illustrated in Figure 2.4. The nodes S (syringyl), P (*p*-hydroxyphenyl), and G (guaiacyl) are the monomers illustrated in Figure 2.1. The nodes β-O-4, β-5, 5-5, β-1, 4-O-5, α-O-4/β-O-4, and β-β are the bonding patterns illustrated in Figure 2.2. The two decision trees on the left are monomer selection trees. Note that the decision trees for the null condition and the β-C, 4-O, and 1-C positions are the same as a result of cominbing similar decision trees from Figure 2.6. The decision trees in the middle and on the right are position and bond selection trees. Again, note that the P and G trees are the same.

2.2.3. Formation of a Lignin Library

As described by Yanez and coworkers, a Markov chain Monte Carlo (MCMC) method, described in Appendix A, is utilized to produce a series of lignin molecules distributed proportionately to a target distribution of interest, which in this case is comprised of the monomers, bond types, molecular weight, and branching coefficient.[49] To form a library of a desired number of lignin molecules, a subset of the growing population of molecules is selected, and statistical tests are performed to verify agreement with experimental data. Specifically, χ^2 goodness-of-fit tests are used to confirm the monomer and bond distributions, while 1-sample t-tests are used to confirm the number-average molecular weight and the branching coefficient. For these tests, it is possible to select a region of acceptance, better known in statistics as a p-value, which is often used as a cutoff between significant results and non-significant results. For instance, in this work, a cutoff p-value of 0.90 is used for all simulations, indicating that the χ^2 test statistic must be within the bound corresponding to a p-value greater than 0.90 in order to statistically conclude that there are no significant differences between the library being tested and the experimental data. In other words, all the libraries presented in this work are produced with a restriction to be in the central 10^{th} fraction of the distribution, corresponding to a p-value greater than 0.90.

2.2.4. Experimental Values

As previously mentioned, Yanez and coworkers[49] generated lignin libraries for simplified wheatstraw, an herbaceous biomass. Here, we have extended the work to include libraries for several other sources, including softwood and hardwood biomass. Specifically, we investigate wheatstraw and miscanthus, spruce, and beech and birch, as these are the biomass sources for which all four experimental characteristics were available in the literature. The experimental

values are summarized in Table 2.1, and relevant discussion for each will be presented in turn. A more detailed summary, including the literature sources for each value in Table 2.1, can be found in Appendix A.

Table 2.1. Tabulation of four experimental properties – monomer percentage, bond percentage, molecular weight, and branching coefficient – for a variety of herbaceous, softwood, and hardwood biomass sources. Note that in the literature, the bonds may be called by the following names: β-O-4 as arylglycerol-β-aryl ether, α-O-4 as α-aryl ether or non-cyclic benzyl aryl ether, β-1 as 1,2-diarylpropane, β-5 as phenylcoumaran, β-β as resinol, 4-O-5 as diaryl ether, and 5-5 as biphenyl. In addition, the monomer percentages refer to frequencies of the monomer compared to the total number of monomers, and the bond percentages refer to frequencies of the linkage compared to the total number of linkages, as opposed to the number of phenylpropane units involved in the linkage. See Appendix A for tabulation and reference details.

		herbaceous		softwood	hardwood	
target type	details	wheatstraw	miscanthus	spruce	beech	birch
monomer percentage	syringyl	30	50	1	36	50
	p-hydroxyphenyl	6	4	5	0	0
	guaiacyl	64	46	94	64	50
bond percentage	β-O-4	77	68	51	60	62
	β-5	11	15	12	6	6
	5-5	3	0	13	2	7
	4-O-5	0	0	4	2	7
	β-1	3	0	10	15	7
	α-O-4	2	0	8	5	8
	β-β	4	17	2	10	3
molecular weight (Da)	number-average	1850	1240	6400	3690	1878
	weight-average	4210	2310	23500	5510	4600
branching coefficient		0.225	0.000	0.301	0.088	0.055

2.2.4.1. Wheatstraw

Yanez and co-workers[49] limited the bond distribution of wheatstraw to three bond types for the purpose of method development. We now extend the distribution to include all seven bonds presented in Figure 2.2. The 2D-NMR performed by Río and co-workers[69] revealed the presence of dibenzodioxocin (Figure 2.3). Although our program does not produce this structure

explicitly, we assumed an equivalent percentage of regular 5-5 bonds, 3%. Additionally, they detected spirodienone (Figure 2.3), which is suggested to degrade to the traditional β-1 dimer in an acidic environment. Given that we did not include spirodienone explicitly, we assumed an equivalent percentage of traditional β-1 bonds, 3%. The remaining five bond types are directly obtained from the 2D-NMR results. The monomer distribution, molecular weight distribution, and branching coefficient are unchanged from Yanez and coworkers' work.[49]

2.2.4.2. *Miscanthus giganteus*

Miscanthus giganteus, a C_4 herbaceous biomass, is a particularly appealing feedstock due to its high productivity in cool temperate conditions.[83] A 95% dioxane extraction procedure was performed to isolate *Miscanthus* and obtain the monomer, bond, and molecular weight distributions.[84] A 2D-NMR characterization of the isolated lignin revealed that *Miscanthus* lignin is an S/P/G type lignin (50%, 4%, 46%, respectively).[84] Additionally, the bond distribution consists of 68% β-O-4, 15% β-5, and 17% β-β.[84] A size exclusion chromatography experiment was used to conclude a number-average and weight-average molecular weight of 1240 Dalton and 2310 Dalton, respectively.[84] Finally, given the experimental bond distribution consisting of only three bond types, β-O-4, β-5, and β-β, the branching coefficient must be zero. In other words, Figure 2.4 demonstrates that three sites cannot possibly be occupied given the bond distribution, keeping in mind that a 5-β bond requires two sites, the 5-C position and the 4-O position. Thus, based on the experimental data collected, *Miscanthus* lignin is linear in nature.

2.2.4.3. Spruce

Spruce, a softwood biomass, has been studied extensively as a possible feedstock.[56,63,77] In a series of permanganate oxidative degradation experiments, Erickson and

coworkers estimated the frequencies of the monomer and bond types in spruce Björkman lignin.[61] Specifically, they reported a S:P:G ratio for spruce lignin to be 1:5:94. The bond types will be discussed in turn, as there are a few discrepancies in the literature. Originally, Erickson reported a value of 2% for the β-1 bond for spruce lignin, based on the assumption that the yields of this linkage and glyceraldehyde-2-aryl ether, revealed by ^{1}H NMR studies,[64] would be equal (see Scheme 2.5c). However, later evaluation of acidolysis products gave an estimated value of 7% for the β-1 bond.[85] Additionally, more recent HMQC (heteronuclear multiple quantum coherence) spectra have identified arylisochroman (Figure 2.3), which includes an α-6 bond and a β-1 bond.[73,76] Indeed, Erickson's results showed 2.5-3% of 2- or 6- condensed structures. It is possible that this percentage included arylisochroman structures. Furthermore, NMR has recently identified spirodienone, which is suggested to degrade to the traditional β-1 dimer in an acidic environment.[29,77] This structure may account for up to 3% of the total linkages. Thus, given the arylisochroman and spirodienone possibilities, it is reasonable to include an additional 3% with the amount for the β-1 bond, giving a total of 10% for the β-1 bond.

For the β-5 bond, the reported range is 9-12%, where the lower limit was obtained by acidolysis,[86] and the upper limit was obtained by Erickson's permanganate oxidative degradation experiments.[61] A value of 12% was chosen to be consistent with the other values chosen. For the 4-O-5 bond, Erickson reported a value of 4% from a range of 3.5-4%. For the β-β bond, Erickson reported the value of 2% obtained by acidolysis[62] for *β-β-linked* bonds. A more recent NMR study[63] confirms a value of 2% for the β-β bonds. There has been significant debate about the presence of the α-O-4 bond, also known as alpha-aryl ether, in lignin. Many authors conclude that this bond is only present in closed forms in the phenylcoumaran and dibenzodioxocin

rings.[51,80] However, Erickson reports a range of 6-8% from Adler's acidolysis experiments for the α-O-4 bond in the open form,[85] so a final value of 8% is chosen for the α-O-4 bond.

The frequency of the 5-5 (biphenyl) bonds can be quite difficult to discern from the literature. It is important to distinguish between the abundance of linkages and the abundance of C_9 moieties. The frequency of the 5-5 *linkage* is 9.5-11%, as originally reported by Erickson.[61] As a result, the frequency of the number of *phenylpropane units* involved in biphenyl linkages is 19-22%. However, more recent ^{13}C NMR examinations have shown a higher frequency of 5-5 linkages of 12-13%.[63] A final value of 13% is chosen.

The β-O-4 (beta-aryl ether) linkage is the most abundant in all types of lignin. There is widespread agreement that this linkage accounts for about 45-50% of the total linkages in softwood lignin. Erickson reported a range of 49-51%, which includes the glyceraldehyde-2-aryl ether structure discussed previously.[61] It is reasonable to include this structure with the β-O-4 linkages because it is very similar in structure. Thus, a final value of 51% is chosen for the β-O-4 linkage. The range reported by Erickson was achieved by performing oxidative degradation before and after treatment with NaOH-CuO, which cleaves β-O-4, α-O-4, and β-5 linkages. Cleavage of these bonds liberates phenolic hydroxyl groups. Therefore, by calculating the increase in phenolic hydroxyls and subtracting the amount of α-O-4 and β-5 bonds, the frequency of β-O-4 bonds can be obtained.

The number-average and weight-average molecular weights for spruce are 6400 and 23500, respectively. These values were obtained from MWL, isolated from extractive-free unbleached Norway spruce.[87] Finally, the branching coefficient for spruce lignin was calculated

to be 0.301 based on a correlation between the branching probability and the weight average molecular weight for western hemlock wood, another softwood.[78]

2.2.4.4. Beech

The monomer distribution for beech was found to be 64% G and 36% S by oxidative degradation experiments.[88] The relative proportions of the bond types for beech, a hardwood, lignin were estimated by performing lignin degradation with thioacetic acid by Nimz and coworkers.[64,72] The percentage of α-β bonds was grouped into that for the β-β bonds based on similarities in bond type. The number-average and weight-average molecular weights for native milled wood beech lignin are 3690 and 5510, respectively.[89] The branching coefficient for beech lignin was calculated to be 0.088 based on a correlation between the branching probability and the weight-average molecular weight for black cottonwood, another hardwood.[79]

2.2.4.4. Birch

For the Björkman lignin of birch hardwood, the G:S ratio is about 1:1, while there is no presence of P.[90] The frequencies of the bond types were determined from oxidative degradation,[91] with the exception of the α-O-4, β-1, and β-β bonds, which were determined from acidolysis.[24,85,92] The number-average and weight-average molecular weights for birch lignin are estimated to be 1878 and 4600, respectively.[28] The branching coefficient for birch lignin was calculated to be 0.055 based on a correlation between the branching probability and the weight-average molecular weight for black cottonwood, another hardwood.[79]

2.3. Results and Discussion

2.3.1. Reducing a Large Parameter Space

Simulations were performed for the original decision trees (27 degrees of freedom (DF)) and the reduced decision trees (7 DF) for spruce, each with both biased and unbiased edge weights, totaling four different scenarios. Figures 2.8-2.10 present the MCMC results for each scenario after sampling $1x10^9$ molecules, in comparison to the experimental values from literature for spruce lignin seen in Table 2.1. Figure 2.8 demonstrates that both scenarios with biased edge weights do an excellent job of matching the experimental *Mn*. The scenario with 7 DF and unbiased edge weights also shows significant improvement in *Mn* compared to the scenario with 27 DF and unbiased edge weights. In the latter case, the simulation is clearly hampered by the excess DF to the point that it is incapable of finding a solution that simultaneously improves all four characteristics. Figure 2.9 shows similar results, in that both scenarios with biased edge weights do the best job of matching the experimental branching coefficient, with the scenario utilizing both techniques, biasing the edge weights and reduced DF (7 DF), doing slightly better. The scenario with 7 DF and unbiased edge weights does a poor job of matching the experimental branching coefficient, but shows a trend toward the experimental value, while the scenario with 27 DF and unbiased edge weights shows a trend away from the experimental value, as the simulation attempts to match the other three properties. Finally, Figure 2.10 shows that the scenario with 27 DF and unbiased edge weights does an extremely poor job of matching the bond distribution, while the other three scenarios match the bond distributions equally well.

This analysis, although performed only on spruce lignin, suggests that using biased edge weights is a powerful technique in improving simulated values to better match experimental values. Additionally, the highly unsuccessful results of the scenario with 27 DF and unbiased edge weights demonstrates that reducing the parameter space by reducing the decision trees was a

necessary step to obtain libraries of representative lignin molecules, while also dramatically accelerating convergence with experimental values.

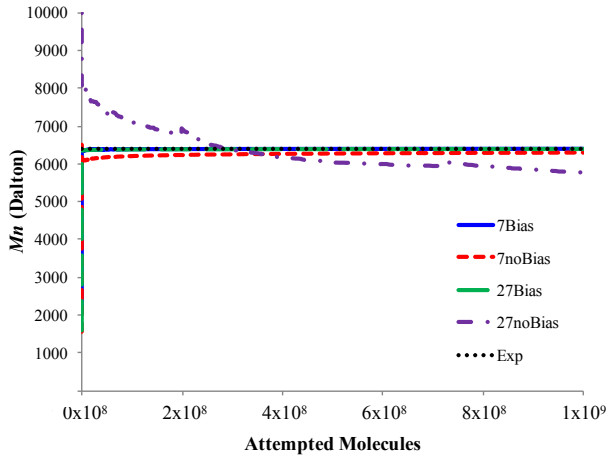


Figure 2.8. Number-average molecular weight (*Mn*) for 7 DF or 27 DF and biased or unbiased edge weights for spruce lignin as compared to the experimental *Mn* value of 6400. Note that the lines for '7Bias', '27Bias', and 'Exp' coincide. Also note that '27noBias' moves away from 'Exp' as it tries to find a solution that simultaneously matches all four properties.

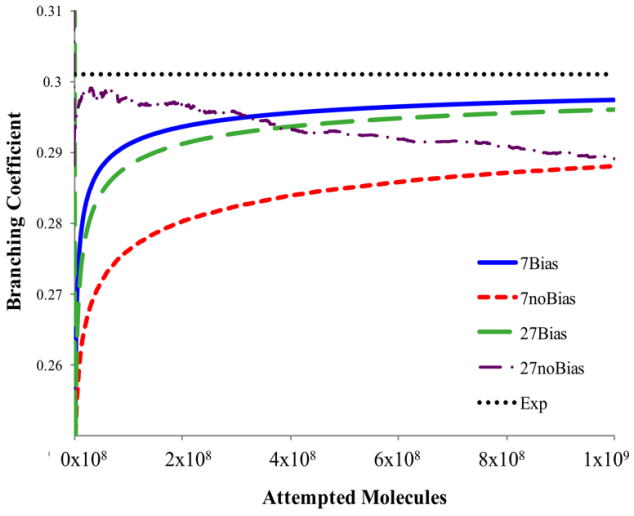


Figure 2.9. Branching coefficient for 7 DF or 27 DF and biased or unbiased edge weights for spruce lignin as compared to the experimental branching coefficient value of 0.301. Note that '27noBias' moves away from 'Exp' as it tries to find a solution that simultaneously matches all four properties.

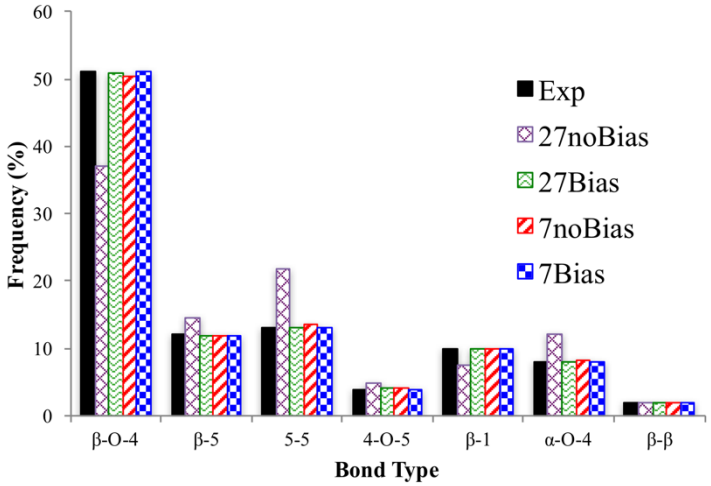


Figure 2.10. Bond distribution for 7 DF or 27 DF and biased or unbiased edge weights for spruce lignin as compared to the experimental bond distribution.

2.3.2. Successful Lignin Library Generation: *Miscanthus*, Spruce, and Birch

To validate the model with 7 DF and biased edge weights, libraries of 100 lignin structures each were generated for three different biomass sources: *Miscanthus giganteus*, spruce, and birch. The properties of one representative library of 100 molecules for each source demonstrating statistical agreement with experimental targets are listed in Table 2.2. Note that the library represented for spruce in Table 2.2 was produced from the simulation for 7 DF and biased edge weights, as described in Section 2.3.1.

Table 2.2. Simulated average properties of libraries of lignin structures compared to corresponding experimental values (target column) for herbaceous-type, softwood, and hardwood lignin. The monomer percentages refer to frequencies of the monomer compared to the total number of monomers, and the bond percentages refer to frequencies of the linkage compared to the total number of linkages.

		herbaceous		softwood		hardwood	
		miscanthus		spruce		birch	
property	details	target	simulated	target	simulated	target	simulated
monomer percentage	syringyl	50	50.00	1	1.01	50	49.84
	p-hydroxyphenyl	4	3.92	5	4.92	0	0.00
	guaiacyl	46	46.08	94	94.06	50	50.16
bond percentage	β-O-4	68	68.16	51	51.01	62	61.86
	β-5	15	14.84	12	11.61	6	5.96
	5-5	0	0.00	13	13.26	7	7.03
	4-O-5	0	0.00	4	4.28	7	6.91
	β-1	0	0.00	10	10.05	7	7.27
	α-O-4	0	0.00	8	7.73	8	7.99
	β-β	17	16.99	2	2.06	3	2.98
molecular weight (Da)	number-average	1240	1240	6400	6436	1878	1875
	weight-average	2310	2325	23500	23629	4600	4447
branching coefficient		0.000	0.0000	0.301	0.300	0.055	0.055

2.3.3. Exploration of Lignin Space via Wheatstraw and Beech Generation

2.3.3.1. Wheatstraw Lignin Space

Given the experimental data obtained from various literature sources, a library of wheatstraw lignin molecules could not be produced using any of the four approaches for determining the edge weights of the decision trees that were discussed in Section 2.3.1. The simulated characteristics of the Markov chain after $1x10^9$ attempted molecules are shown in Table 2.3. While the simulated monomer, bond, and molecular weight distributions agree well with the experimental values, the simulated branching coefficient is only 0.118 compared to an experimental value of 0.225. While the simulated branching coefficient is a measure of the branching in a Markov chain or a library/reservoir of lignin molecules, the branching propensity is an additional adjustable parameter defining the likelihood of a branch occurring and is used to

determine branching during the formation of individual lignin molecules. For this simulation, the branching propensity was a low value of 0.094, creating molecules with a small amount of branching. We discovered a direct correlation between the branching coefficient and the percentage of 5-5 bonds when we calculated these characteristics for a simple reservoir of lignin molecules yet to undergo MCMC. As expected, manually increasing the branching propensity resulted in a larger branching coefficient. However, it also led to a drastic increase in the percentage of 5-5 bonds, as seen in Figure 2.11. Based on Figure 2.4, there are 28 possible combinations of bonds that would constitute a branching situation, which can be seen in Appendix A. These 28 situations are the outcome of defining all the allowable situations with three or more positions occupied on a monomer. Note that several unlisted combinations of bonds are not allowable due to the structure of the linkage types. For instance, a single combination cannot consist of both a bond at the 1-C position and a bond at the α-C and/or the β-C position because the propanoid region containing these latter two positions detaches when a 1-β bond forms. Also, the 5-β linkage is not listed and does not occur in the decision trees because it requires two bonding positions – the 5-C and 4-O positions. The experimental data shows that wheatstraw lignin does not contain any 4-O-5 bonds. Therefore, there are only eight possible branching situations for this set of experimental data, seven of which include a 5-5 bond. The other combination is an α-O-4 bond, a β-O-4 bond, and a 4-O-β (β-O-4 bond originating from the 4-O position). However, the α-O-4 bond only accounts for 2% of the bond distribution, making this branching situation unlikely. Thus, it is highly likely that every time a branch occurs, a 5-5 bond will be involved, making it difficult to simultaneously keep the 5-5 bond percentage at a low value and the branching coefficient at a relatively high value.

Table 2.3. Tabulation of Markov chain properties (simulated column) compared to the corresponding experimental values (target column) for wheatstraw lignin. The monomer percentages refer to frequencies of the monomer compared to the total number of monomers, and the bond percentages refer to frequencies of the linkage compared to the total number of linkages.

		wheatstraw	
target type	details	target	simulated
monomer percentage	syringyl	30	29.60
	p-hydroxyphenyl	6	6.02
	guaiacyl	64	64.38
bond percentage	β-O-4	77	76.37
	β-5	11	11.00
	5-5	3	3.54
	4-O-5	0	0.00
	β-1	3	2.98
	α-O-4	2	2.15
	β-β	4	3.95
molecular weight (Da)	number-average	1850	2204
	weight-average	4210	4379
branching coefficient		0.225	0.118

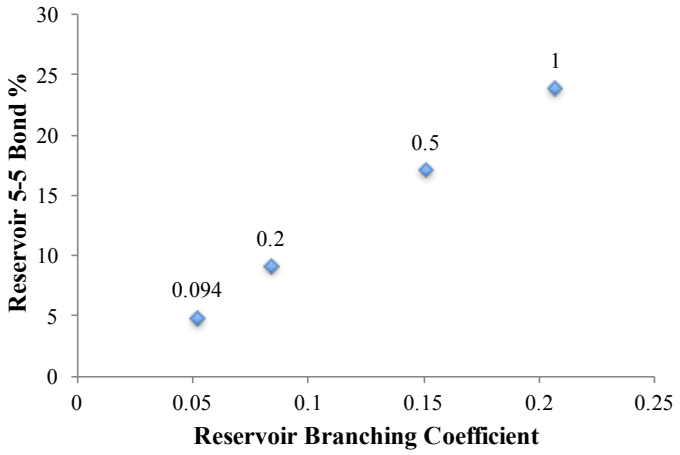


Figure 2.11. 5-5 bond percentages as a function of branching coefficient for a reservoir of lignin molecules. The data labels represent the branching propensity used in the generation of the individual lignin molecules.

Given this result, a critical analysis of the space of wheatstraw lignin structures, given the monomer and molecular weight distributions in Table 2.3, was conducted. First, the experimental branching coefficient, the target to be matched, was manually changed to determine the range of branching coefficients compatible with the monomer, bond, and molecular weight distributions in Table 2.3. For target values of the branching coefficient up to 0.13, the simulated branching coefficient matched well, and libraries were made, indicating that the specified targets were met within the statistical accuracy specified. However, for target branching coefficients higher than 0.13, the highest achieved simulated branching coefficient was 0.1325 at a target branching coefficient of 0.17, but no libraries were made, as seen in Figure 2.12. This shows that, in order to make libraries, given the monomer, bond, and molecular weight data in Table 2.3, the target branching coefficient must be a maximum of 0.13. For target branching coefficients higher than 0.17, the simulated branching coefficients were well below the target, and no libraries were made. The fact that these simulations were unable to reach the maximum achieved branching coefficient of 0.1325 can be attributed to the tradeoffs these simulations encountered as they also attempted to match the other three characteristics well.

To explore these limitations further, next the experimental 5-5 bond percentage was manually changed to determine the range of 5-5 bond percentages compatible with the other three characteristics in Table 2.3. Figure 2.13 reveals that for cases with a target 5-5 bond percentage of 6.0% or less, the simulated branching coefficient is unable to statistically match the literature value, and no libraries are made. Thus, given the monomer, molecular weight, and branching coefficient data in Table 2.3, libraries can only be produced if the 5-5 bond percentage is 7% or higher.

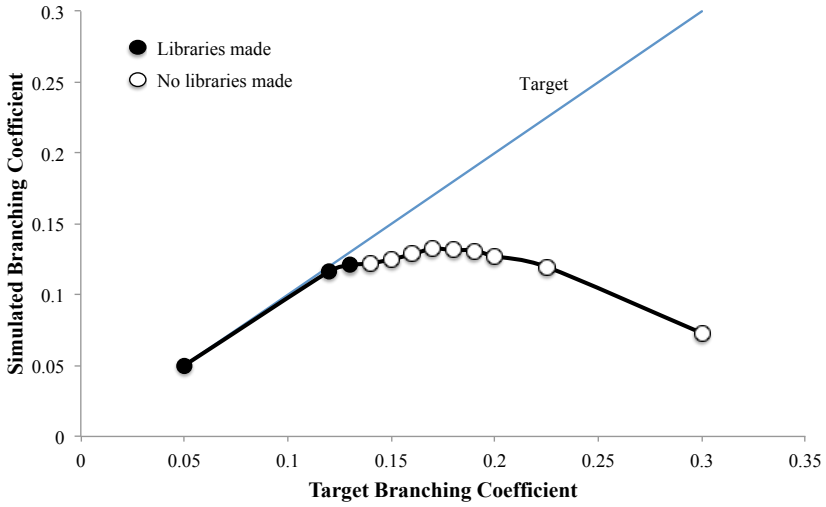


Figure 2.12. Simulated branching coefficient achieved as a function of the target experimental branching coefficient set. The line of parity is provided to underscore that simulated branching coefficients equal to the corresponding target experimental branching coefficients were not achievable in all cases.

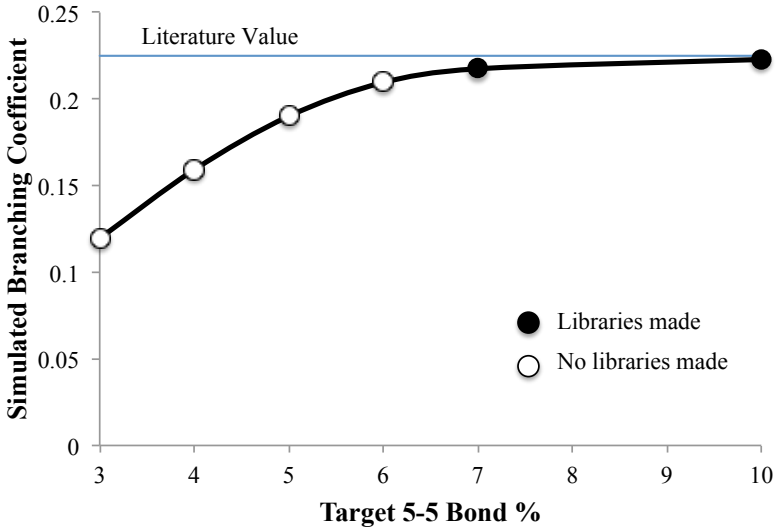


Figure 2.13. Simulated branching coefficient as a function of the target experimental bond percentage. Only when the target 5-5 bond percentage was 7% or higher were libraries able to be made and simulated branching coefficients close to the literature value of 0.225 shown by the horizontal line able to be closely achieved.

These are particularly noteworthy results, as it demonstrates the power of the simulations to reveal that the experimental data given in Table 2.3 is not physically and chemically consistent. In particular, we note that the branching coefficient was obtained from a different source than the other three characteristics. Further, the correlation used to estimate the branching coefficient was developed using western hemlock, a softwood, while wheat straw lignin is of the grass type. Thus, the simulation results underscore the need for all properties to be obtained experimentally in a consistent way, and further suggest that a correlation developed from an herbaceous lignin to quantify the general lignin branching coefficient is a gap in the current knowledge base. Extending this concept more generally, we propose that using our structure generation approach, we are able to deduce a chemical restriction on the lignin space. This not only tells researchers what types of lignin exist, but it also guides experimentalists to critically evaluate the self-consistency of reported properties of different types of lignin.

2.3.3.2. Beech Lignin Space

Similar to wheatstraw, a library of beech lignin molecules could not be produced given the experimental data. The simulated characteristics of the Markov chain after $1x10^9$ attempted molecules are shown in Table 2.4. In this case, the simulated monomer and molecular weight distributions as well as the branching coefficient match the experimental values well, but the simulated bond distribution is skewed from the experimental one. More specifically, the simulated β-O-4 bond percentage is higher than the experimental value, while the simulated β-β and β-1 bond percentages are lower. Figure 2.2 reveals that the β-β and β-1 bonds are unique in that they occupy the β-C position on both bonding monomers, while simultaneously involving the

α-C and 1-C positions. This leaves the 4-O and 5-C positions open and available for bonding. If the

Table 2.4. Tabulation of Markov chain properties (simulated column) compared to corresponding experimental values for beech lignin. Note that the β-O-4 bond percentage is too high, at the expense of the β-β and β-1 bond percentages, which are too low. The monomer percentages refer to frequencies of the monomer compared to the total number of monomers, and the bond percentages refer to frequencies of the linkage compared to the total number of linkages.

		beech	
target type	details	target	simulated
monomer percentage	syringyl	36	36.43
	p-hydroxyphenyl	0	0.00
	guaiacyl	64	63.57
bond percentage	β-O-4	60	71.77
	β-5	6	6.04
	5-5	2	2.22
	4-O-5	2	2.20
	β-1	15	5.68
	α-O-4	5	6.20
	β-β	10	5.89
molecular weight (Da)	number-average	3690	3625
	weight-average	5510	5485
branching coefficient		0.088	0.084

next bond chosen is 4-O-β, the next bonding monomer's β-C position is filled as well. It is much more likely to select the 4-O-β bond, given that its percentage of the total bonds is 60%, than it is to select a 4-O-5 or 5-5 bond, which account for only 4% of the total bonds. Therefore, there is rarely a β-C available for another β-1 or β-β bond to form. Hence, the simulated β-β and β-1 bond percentages are significantly lower than the experimental values.

This unique attribute of the β-β and β-1 bonds, in addition to the high percentage of β-O-4 bonds, leads to a low number of β-β and β-1 bonds in the molecule, regardless of the size of the molecule. Thus, we hypothesized that a higher percentage of β-β and β-1 bonds can be achieved if a lower target *Mn* is used. Figure 2.14 presents the dependence of *Mn* on the simulated β-β and β-1 bond percentage. For target *Mn* values less than or equal to 1500 Da, the simulated β-β and β-1 bond percentage was able to statistically match the literature value, and libraries were formed. However, for target *Mn* values greater than 1500, the simulated β-β and β-1 bond percentage was unable to match the literature value, and libraries could not be formed. As expected, as the target *Mn* decreases, the simulated β-β and β-1 bond percentage increases.

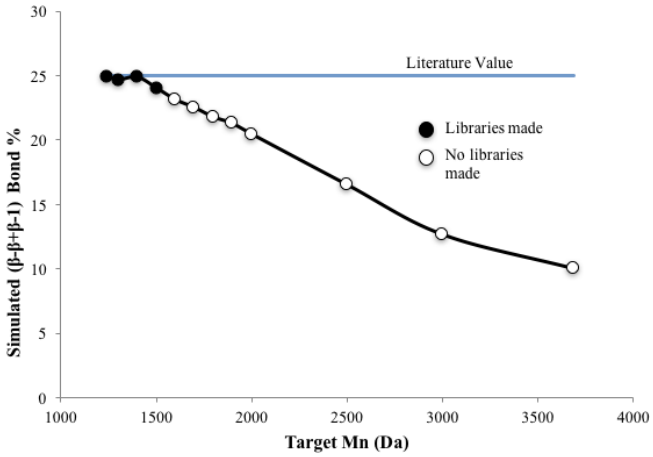


Figure 2.14. Simulated β-β and β-1 bond percentage of the target *Mn* value. Only when the *Mn* value is decreased to 1500 Da can the target percentage of β-β and β-1 bonds, shown by the horizontal line, be achieved.

As noted, it is the high percentage of β-O-4 bonds that prohibits the appearance of a β-C. Decreasing the percentage of β-O-4 bonds and increasing the percentage of 4-O-5 bonds accordingly makes the presence of the bonding location β-C more likely. Figure 2.15 presents the results of manually changing the target β-O-4 bond percentage. The percentage of β-β and β-1

bonds reaches a level close enough to the literature value to create libraries at a β-O-4 bond percentage of 50%. Thus, with a β-O-4 bond percentage of 50%, the molecular weight and percentage of β-β and β-1 bonds are compatible with one another.

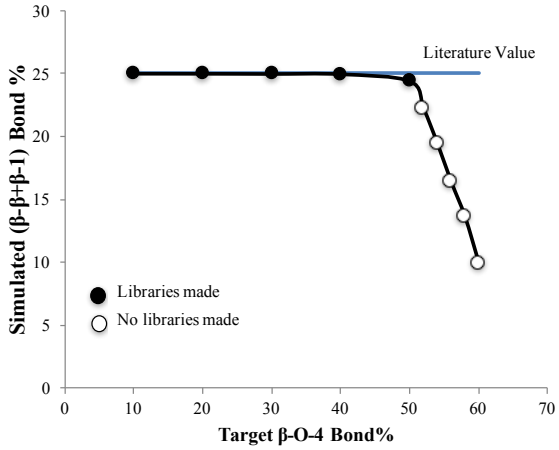


Figure 2.15. Simulated β-β and β-1 bond percentage as a function of the target β-O-4 bond percentage. When the β-O-4 bond percentage drops below 50%, the target value of β-β and β-1 bonds, shown by the horizontal line, can be achieved.

Finally, the target percentage of β-β and β-1 bonds was manually changed to determine the maximum value achievable given the other three characteristics in Table 2.4. As seen in Figure 2.16, libraries were made for target β-β and β-1 bond percentages of 19% or lower. Therefore, given the experimental data in Table 2.4, in order to produce libraries, a target β-β and β-1 bond percentage of 19% or lower is required.

Given these analyses, the simulations provide new insight into what we have termed "lignin space" by revealing that the specified values of β-β and β-1 bond percentage, β-O-4 bond percentage, and number average molecular weight cannot be matched simultaneously; a maximum of two out of these three properties can be met. As before with wheatstraw, this exposes the inconsistencies in the experimental data given in Table 2.4. Specifically, we note that the molecular

weight was obtained from a different source than the bond distribution. Further, the molecular weight was obtained from native milled wood lignin, while the bond distribution was obtained from lignin degraded with thioacetic acid. As well known, the treatment method can have a significant effect on the properties of the lignin due to potentionally degrading the structure and changing bond types, thereby emphasizing that all properties need to be obtained experimentally in a consistent way.

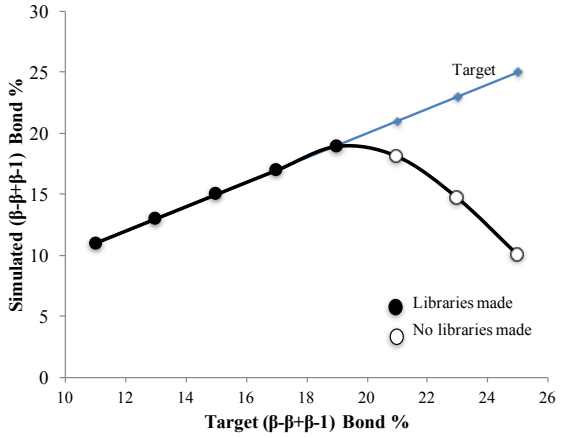


Figure 2.16. Simulated β-β and β-1 bond percentage as a function of the target β-β and β-1 bond percentage. Only when the target β-β and β-1 bond percentage drops below 19% can the simulated value of the β-β and β-1 bond percentage achieve its target value, shown by the parity line.

2.4. Conclusions

The method developed in this work is capable of developing libraries for any biomass source that has reliable and consistent experimental data, whether it be native, treated, or technical lignin. Further, this method allows for exploration of lignin space to determine incompatible characteristics, identify impossible lignin structures, and help guide researchers to identify practical lignin molecules. In this method, all possible bonding patterns, informed by lignin biosynthesis rules, are encoded in a decision tree with edge weights defining the probability of

each path. The two techniques utilized to reduce a large parameter space of edge weights, reducing the decision trees and biasing the edge weights, greatly improved the efficiency of the algorithm. Libraries of lignin structures were successfully generated for *miscanthus*, spruce, and birch. In the cases of wheatstraw and beech, relationships between experimental characteristics and limitations of certain characteristics were discovered. The libraries generated by this method can be subsequently utilized in kinetic models and molecular simulations to determine a broad range of properties, including reactivity, as a function of lignin source.

Chapter 3: Group Additivity Determination for Oxygenates, Oxonium Ions, and Oxygen-Containing Carbenium Ions

Material in this chapter is reproduced from the publication "Group Additivity Determination for Oxygenates, Oxonium Ions, and Oxygen-Containing Carbenium Ions" by Lauren D. Dellon, Chun-Yi Sung, David J. Robichaud, and Linda J. Broadbelt; *Industrial & Engineering Chemical Research* **2017**, *56*(37), 10259-10270.

3.1. Introduction

The global depletion of fossil fuel reserves has prompted the investigation into renewable sources of energy, particularly biofuels, which can be produced from many different sources, including manure, crop waste, and biomass.[1-4] A thermochemical technique called fast pyrolysis is often used to convert biomass into liquid bio-oil, a complex mixture of water and hundreds of organic compounds belonging to classes such as acids, carbonyls, and phenolics.[2,13-21] While bio-oil can be produced in yields up to 75 wt%,[14,22,23] it is characterized by high viscosity, instability, and corrosiveness, as it can contain up to 40 wt% oxygen and retain a pH as low as 2.8.[2] Thus, carefully designed catalysts are used to upgrade the pyrolysis vapors by removing oxygenated products before condensation into the liquid product. A major class of catalysts used for vapor phase upgrading is zeolites, which offer Brønsted acidic sites promoting acid-catalyzed chemistry.[2,14,15,33] In the case of catalytic fast pyrolysis (CFP) of biomass, this chemistry involves oxygen-containing carbenium ions (C^+) and oxonium ions (O^+) as intermediates.[2,34,35] The complex chemistry, in addition to other concerns, such as the short residence time of the vapors (1-2 s), mass transport issues, and rapid catalyst deactivation, have largely hindered efforts to develop empirical kinetic models for CFP. There currently exists a single empirical kinetic model

developed by Adjaye and Bakhshi[93] wherein model compounds are used to suggest simple reaction pathways of lumped components such as alcohols, aliphatic hydrocarbons, and aromatic hydrocarbons. On the other hand, computational models, such as microkinetic models, can simulate this rapid, complex chemistry at the individual species' level, as well as give insight into catalyst deactivation. Not only can these models be used to predict product distributions and investigate specific reaction pathways, but, most importantly, they can enable the design and synthesis of more effective catalysts as well as facilitate optimization of experimental process conditions.

For an accurate microkinetic model, rate constants are needed for each elementary step, which can be obtained in some cases from experiments, but more commonly must be calculated using quantum chemical calculations and transition-state theory.[94] Unfortunately, these calculations are often avoided due to their computational expense, leading to many gaps in the experimental databases for rate constants. If the common method of calculating rate constants based on the Arrhenius equation (Equation 3.1) is used

$$k = Ae^{-E_a/RT} \tag{3.1}$$

where k is the reaction rate constant, A is the pre-exponential factor, E_a is the activation energy, R is the ideal gas constant, and T is the temperature, then two parameters, A and E_a, are required. These two individual parameters can be related to thermodynamic properties using structure-reactivity relationships, such as the Evans-Polanyi relationship (Equation 3.2)[95]

$$E_a = E_o + \alpha \Delta H_{rxn} \tag{3.2}$$

where E_a is the activation energy for the reaction of interest, E_o and α are two parameters numerically fitted to match known E_a values or product yields over a wide range of process

conditions, and ΔH_{rxn} is the enthalpy of reaction for the reaction of interest. While this translates the challenge of estimating rate coefficients into calculating thermodynamic properties, which are typically easier to obtain, an enthalpy of reaction is still required for each reaction, which can be calculated from the individual enthalpies of formation of the reacting species.

Highly accurate values of enthalpies of formation may be calculated from high-level quantum chemical calculations. However, the size of many reaction networks, including the catalytic upgrading of oxygenated compounds, restrains the use of these quantum chemical calculations for all possible species. Thus, group additivity (GA) has been commonly used to estimate the enthalpies of formation for individual species.[96-104] In this approach, originally developed by Benson,[105-107] the compound is broken down into its structural constituent groups, each of which has a contributing value to the total enthalpy of formation. A group is defined as a polyvalent atom and all of its associated ligands. An example of the nomenclature of a group is C-$(C)_2(H)(O^+)$. This represents a central carbon atom bound to two carbon atoms, a hydrogen atom, and a positively charged oxygen, as seen in Figure 3.1a. Figures 3.1b-3.1d show several other examples of groups relevant to the current study. Note that a carbon atom that is connected to another carbon atom via a double bond is given the unique symbol C_d, and a carbon atom that is connected to an oxygen atom via a double bond is defined as a single entity (CO). The method of GA has been shown to be very accurate for oxygen-containing neutral organic compounds, or oxygenates.[108-111] However, additional groups, such as those with a central carbon atom bound to four different substituents (e.g. C – (C_d)(CO)(O)(H)), are required to calculate enthalpies of formation for all potential oxygenates. Additionally, GA values for the enthalpies of formation for oxygen-containing carbenium ions (C^+) and oxonium ions (O^+) are lacking in the literature.

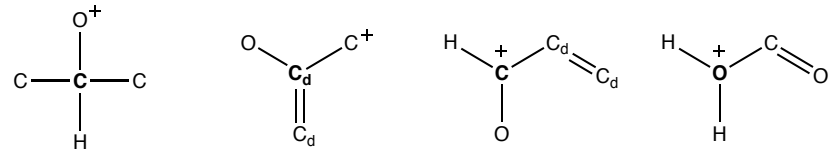


Figure 3.1. Examples of groups. The graphical representation is shown with the standard group notation directly below it. C_d is for a carbon atom comprising a double bond, and numerical subscripts represent the total number of that atom in the group.

It is important that a consistent and clear language defining these types of species be put forth here, as there can be ambiguity in the literature regarding ions containing carbon and oxygen. For the purposes of this study, molecules are defined according to the definitions put forth by Olah,[112] who has studied oxonium ions rigorously. An oxygenate is simply a neutral species containing hydrogen, carbon, and oxygen, as in Figure 3.2a. A carbenium ion is a trivalent carbocation, so an oxygen-containing carbenium ion is a trivalent carbocation with oxygen also present in the molecule but not directly bonded to the ionic carbon, as in Figure 3.2b. In the case where the charged carbon is bonded to an oxygen (Figure 3.2c), the molecule is termed a carboxonium ion (Figure 3.2c_1) or oxocarbenium ion (Figure 3.2c_2) to reflect the carbenium-oxonium resonance.[113] Finally, an oxonium ion is a compound containing a positively charged trivalent oxygen atom, seen in Figure 3.2d.[112]

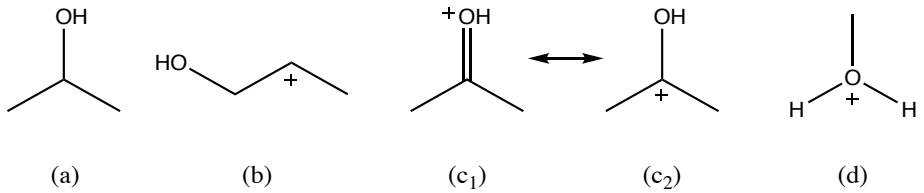


Figure 3.2. Example molecules: (a) Example of an oxygenate molecule, (b) example of an oxygen-containing carbenium ion, (c_1) example of a carboxonium ion, (c_2) example of an oxocarbenium ion, (d) example of an oxonium ion. Note that (c_2) is a subset of (b).

In this work, high-level quantum chemical calculations using Gaussian-4 on a set of 195 compounds comprising oxygenates, oxonium ions, and oxygen-containing carbenium ions were performed. We calculated enthalpies of formation for each using isodesmic reactions and derived a set of 71 GA values. Our results are compared to experimental values where available.

3.2. Methodology

Quantum mechanical (QM) calculations were performed using Gaussian-4 theory.[114] Optimized geometries were obtained using density functional theory with the B3LYP functional and the 6-31G(2*df*, *p*) basis set. Harmonic frequencies were also obtained at the B3LYP/6-31G(2*df*, *p*) level of theory and scaled by a factor of 0.9854 to account for known deficiencies.[115] The G4 energy was obtained by a series of single point correlation energy calculations with the addition of several corrections, such as those for diffuse functions, higher polarization functions, correlation effects beyond a fourth-order perturbation theory using a coupled cluster theory, larger basis set effects, the Hartree-Fock limit, and any other deficiencies.[114] Absolute enthalpies at 298 K for compounds and elements were calculated using Equation 3.3:

$$H^{298} = E_{elec} + ZPE + E_{vib}^{298} + E_{trans}^{298} + E_{rot}^{298} + \Delta PV \qquad (3.3)$$

where H^{298} is the absolute enthalpy at 298 K, E_{elec} is the electronic energy, *ZPE* is the zero-point energy, *P* is the pressure, *V* is the volume, and E_{vib}^{298}, E_{trans}^{298}, and E_{rot}^{298} are the vibrational, translational, and rotational thermal corrections, respectively, from 0 to 298 K.

As standard state values cannot be obtained directly from QM calculations, the standard enthalpy of formation of a given species was obtained using either atomization enthalpies or isodesmic reactions. Atomization enthalpies were only used to check the accuracy of the Gaussian-4 method by comparing to experimental values. A set of 27 reference molecules was defined, and

the atomization enthalpies were calculated for an oxygenate using Equation 3.4, an oxygen-containing carbenium ion using Equation 3.5, and an oxonium ion using Equation 3.6.

$$\Delta H_a^o(C_xH_yO_z) = [xH^{298}(C) + yH^{298}(H) + zH^{298}(O)]\text{-}H^{298}(C_xH_yO_z) \quad (3.4)$$

$$\Delta H_a^o(C_x^+H_yO_z) = [(x-1)H^{298}(C) + H^{298}(C^+) + yH^{298}(H) + zH^{298}(O)]\text{-}H^{298}(C_x^+H_yO_z) \quad (3.5)$$

$$\Delta H_a^o(C_xH_yO_z{}^+) = [xH^{298}(C) + yH^{298}(H) + (z-1)H^{298}(O) + H^{298}(O^+)]\text{-}H^{298}(C_xH_yO_z{}^+) \quad (3.6)$$

ΔH_a^o is the atomization enthalpy of an oxygenate ($C_xH_yO_z$), an oxygen-containing carbenium ion ($C_x^+H_yO_z$), and an oxonium ion ($C_xH_yO_z{}^+$). The standard enthalpy of formation of an oxonium ion, $\Delta H_f^o\left(C_xH_yO_z{}^+\right)$, was calculated using Equation 3.7, and the standard enthalpies of formation for oxygenates and carbenium ions were calculated in a similar manner.

$$\Delta H_f^o(C_xH_yO_z{}^+) = \left[x\Delta H_f^o(C) + y\Delta H_f^o(H) + (z-1)\Delta H_f^o(O) + \Delta H_f^o(O^+)\right]\text{-}\Delta H_a^o(C_xH_yO_z{}^+) \quad (3.7)$$

The enthalpies of formation for *C*, C^+, *H*, *O*, and O^+ at standard state (298 K), as well as the absolute enthalpies at 298 K, are given in Appendix B.

The standard enthalpies of formation for the 27 reference molecules were compared to experimental values found in the literature. For several ionic species, the enthalpy of formation was not explicitly available experimentally. Instead, the proton affinity and enthalpy of formation of the species' deprotonated counterpart were used to derive the experimental enthalpy of formation of the ionic species. The enthalpy of formation for the ionic species, MH^+, from the hypothetical protonation of *M*, as seen in Equation 3.8, is defined in Equation 3.9.[116]

$$M + H^+ \rightarrow MH^+ \quad (3.8)$$

$$\Delta H_f^o(MH^+) = \Delta H_f^o(M) + \Delta H_f^o(H^+)\text{-}PA(M) \quad (3.9)$$

where $\Delta H_f^o(MH^+)$ is the derived experimental enthalpy of formation for the ionic species, and $\Delta H_f^o(M)$ is the experimental enthalpy of formation for the ionic species' neutral counterpart. At

298 K, $\Delta H_f^o(H^+)$, the enthalpy of formation of the proton, is 365.690 kcal/mol.[116] The proton affinities of the neutral species, $PA(M)$, were obtained from the extensive list of proton affinities compiled by Hunter and Lias.[117]

Next, a second method based on isodesmic reactions was used to validate the G4 method and to calculate the values of the enthalpies of formation for the species used to regress the GA values. Isodesmic reactions are those that conserve the number of the same types of bonds in the reactants and the products.[118] As a result, they are known to cancel out any errors that may arise in the QM calculations. Isodesmic reactions have proven to be an effective method for estimating enthalpies of formation.[96,102,109,119-123] Isodesmic reactions were constructed using the 27 reference molecules to calculate enthalpies of formation for a set of 195 species, seen in Figure 3.3. As an example, the following reaction, Equation 3.10, is an isodesmic reaction used to calculate the enthalpy of formation for species 1 (*sp1*) from Figure 3.3.

$$H_2C^+OH \quad + \quad C_3H_7^+ \quad \rightarrow \quad sp1 \quad + \quad CH_3^+ \tag{3.10}$$

where H_2C^+OH, $C_3H_7^+$, and CH_3^+ are reference molecules with known experimental enthalpies of formation. The enthalpy of formation of species 1 was calculated using Equation 3.11:

$$\Delta H_{f,x}^o = H_x^{298} + \sum_i \upsilon_i (H^{298}\text{-}\Delta H_f^o)_{A_i} \tag{3.11}$$

where *x* is the species number, ΔH_f^o is the enthalpy of formation, H^{298} is the absolute enthalpy from the G4 QM calculations, *i* is the number of reference molecules in the isodesmic reaction, and A_i is a reference molecule in the reaction with stoichiometric coefficient υ_i, following the usual convention of negative for reactants and positive for products. Two isodesmic reactions were written for each species in Figure 3.3, and the average of the two enthalpies of formation was taken

to be that species' enthalpy of formation. These average enthalpies of formation were compared to experimental data when available, as shown in Table 3.3.

Finally, the enthalpies of formation that were calculated based on isodesmic reactions were used to obtain GA values for ΔH_f^o. First, 55 groups were fixed to either known values from literature or analogous values, with a rank analysis applied to identify the number of degrees of freedom that needed to be specified. Next, multiple linear regression was used to obtain the GA values for 71 groups containing 9 oxygenates, 25 oxonium ions, 36 oxygen-containing carbenium ions, and 1 next-nearest-neighbor correction. Further details on multiple linear regression, as applied for GA value determination, can be found elsewhere.[98,109] Six of the oxygenate GA values were already reported in the literature by Paraskevas et al.,[110] but were not explicitly used in this study, as their database of fixed groups is significantly different than the one used here. Thus, while each database of groups may individually produce accurate estimates of enthalpies of formation for molecules, a comparison across databases of the group values themselves is not meaningful.

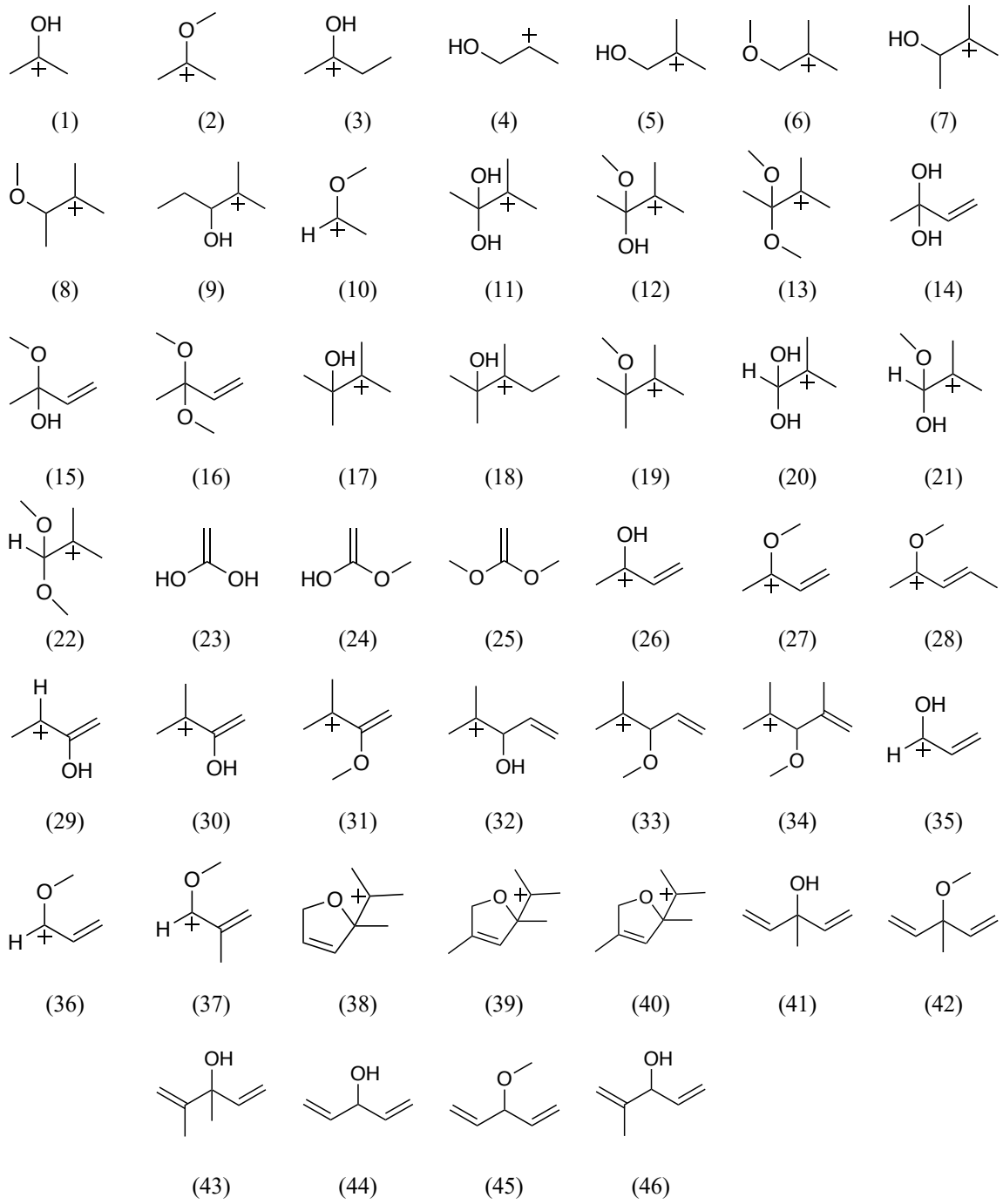


Figure 3.3. Molecules used to derive GA values for enthalpies of formation. Their enthalpies of formation were calculated from Gaussian-4 quantum chemical calculations and isodesmic reactions, which were constructed from a set of 27 reference molecules. Part 1 of 4.

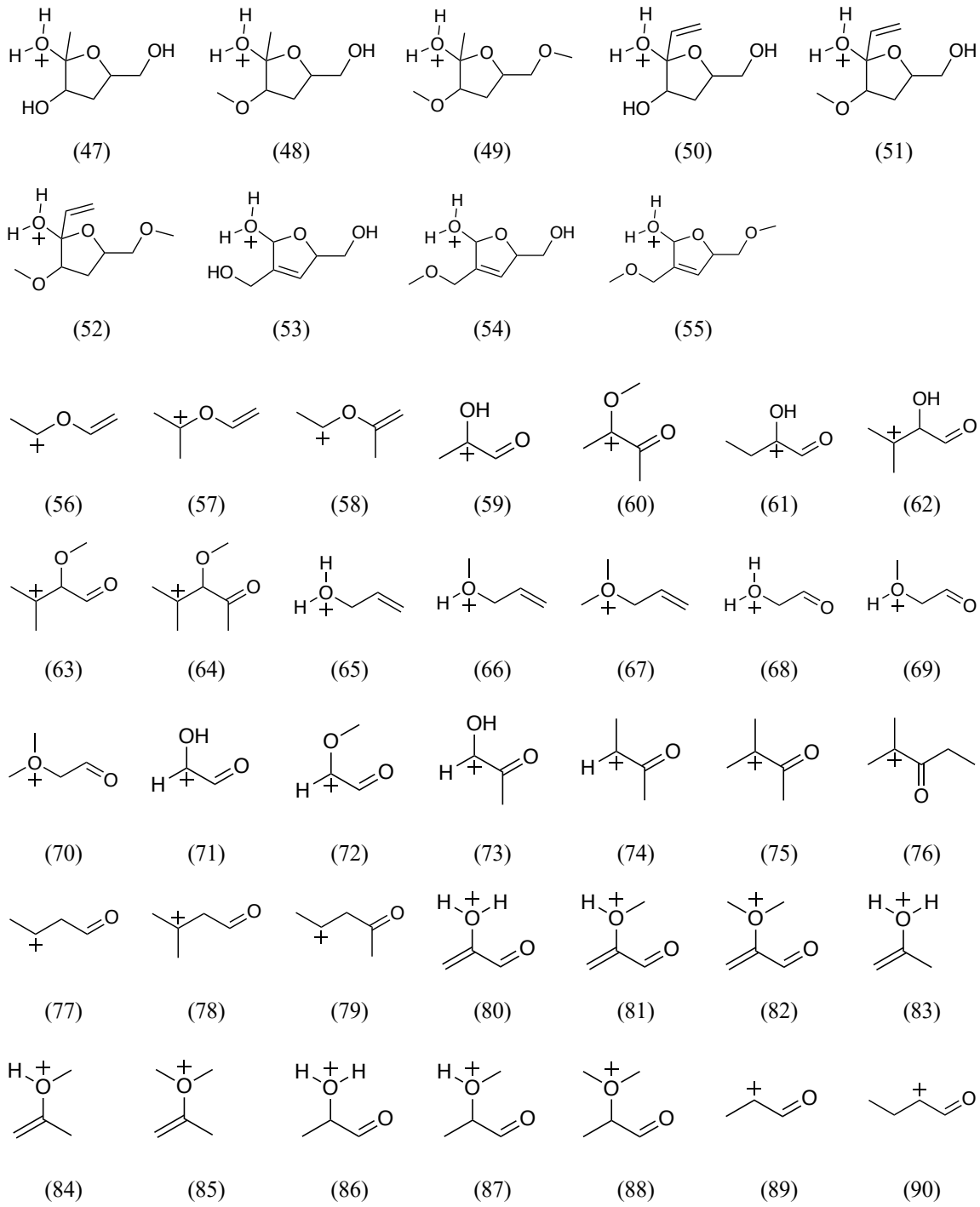


Figure 3.3. Part 2 of 4.

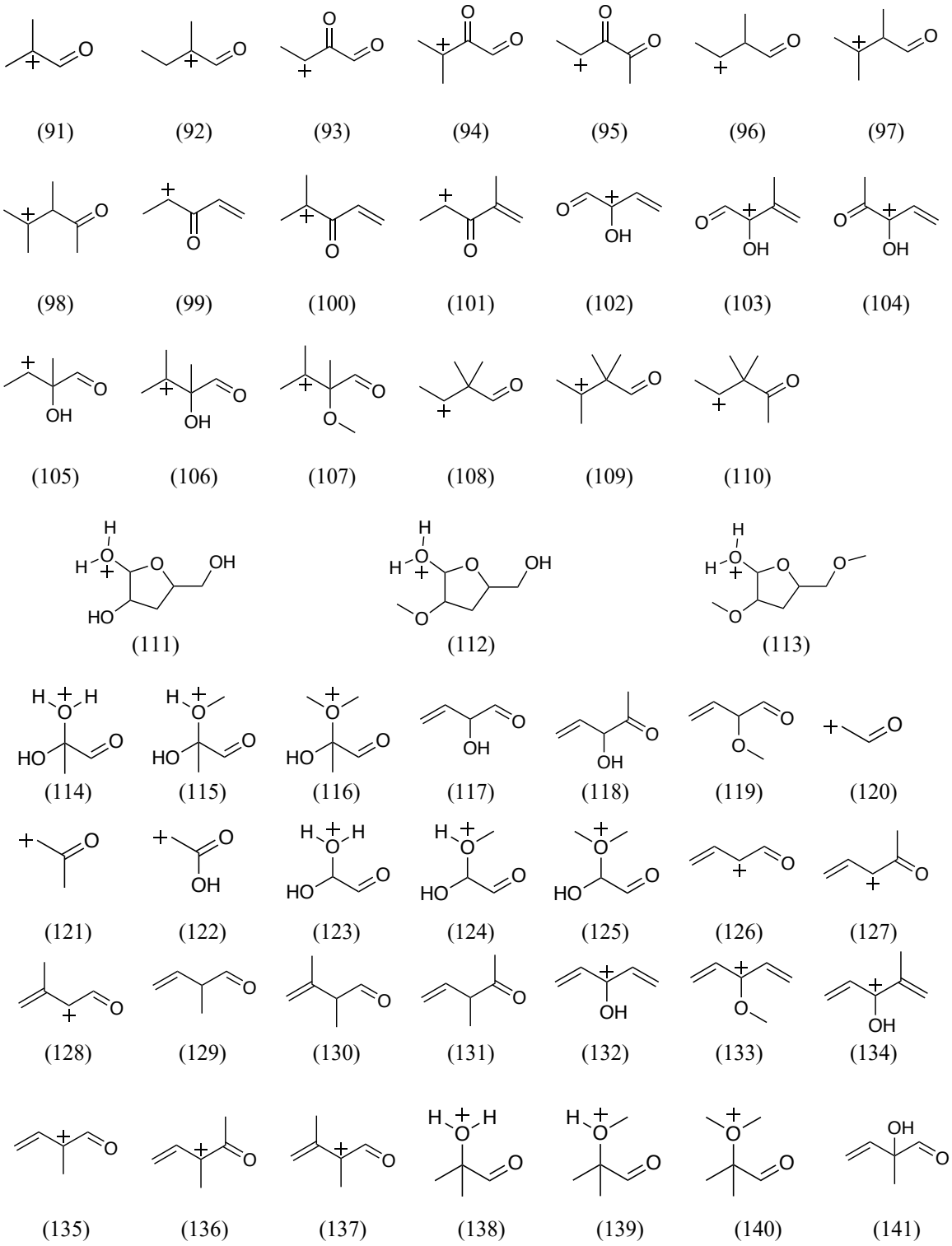


Figure 3.3. Part 3 of 4.

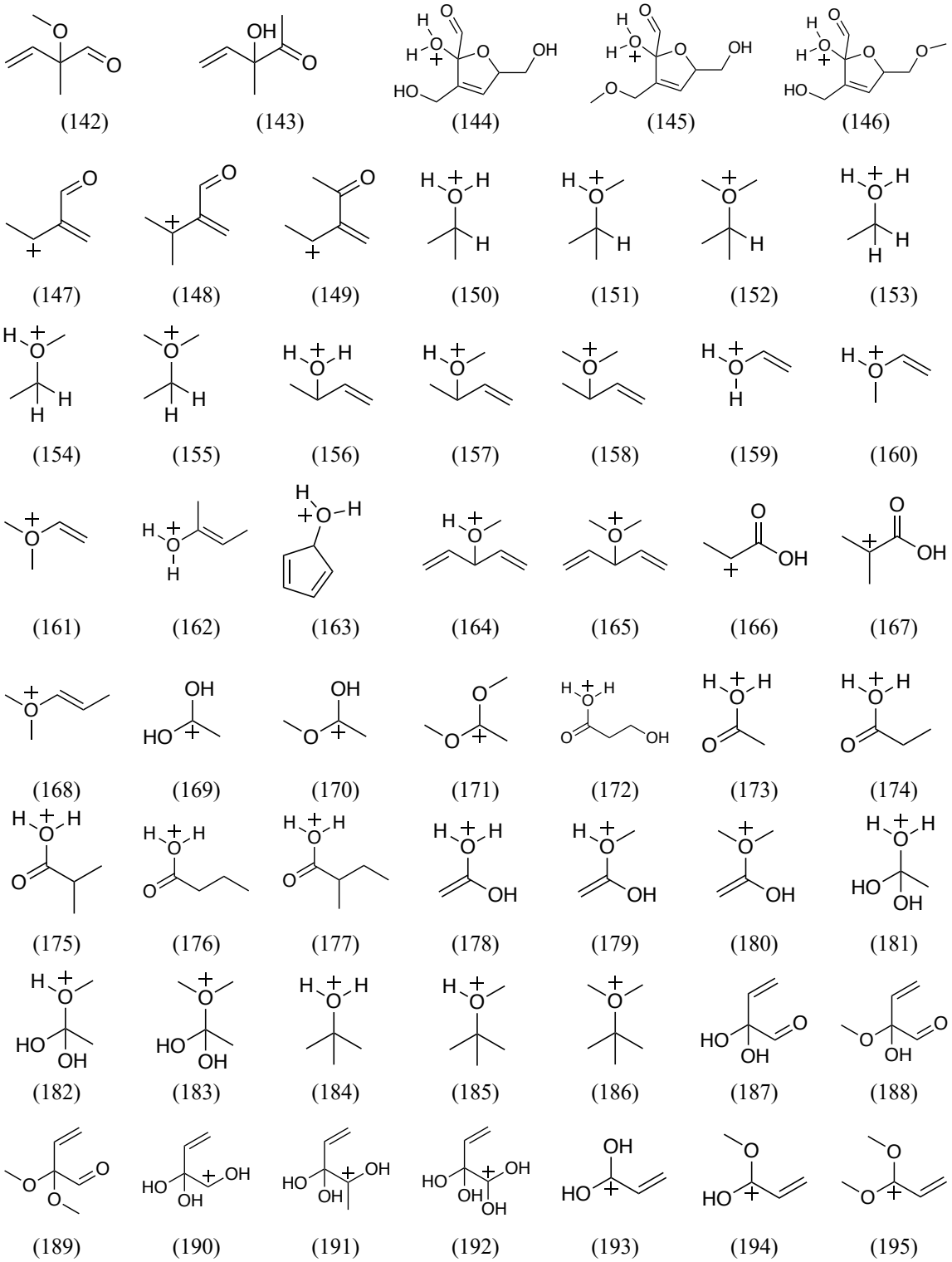


Figure 3.3. Part 4 of 4.

3.3. Results and Discussion

G4 calculations were performed for 27 reference molecules and the 195 species in Figure 3.3 to obtain an optimized geometry and the absolute enthalpy, H^{298}, for each. The optimized geometries can be found in an attached Excel file titled 'Optimized Geometries'. The absolute enthalpies were subsequently used to calculate atomization enthalpies and construct isodesmic reactions for method validation. The absolute enthalpies at 298 K and the atomization enthalpies for the reference molecules and the species with available experimental data can be found in Appendix B. The enthalpies of formation from the isodesmic reactions were then used to regress 71 GA values for oxygenates, oxonium ions, and oxygen-containing carbenium ions.

3.3.1. Method Validation from Atomization Enthalpies

The atomization enthalpies were calculated for a set of 27 reference molecules using Equations 3.4-3.6. The enthalpies of formation were subsequently calculated using Equation 3.7. The 27 reference molecules and their corresponding enthalpies of formation based on the method of atomization enthalpies are provided in Table 3.1. The experimental enthalpies of formation in the literature are also provided. In some instances, the enthalpy of formation was not directly available. Thus, proton affinity calculations were used to calculate the experimental enthalpy of formation, as seen in Table 3.2. The mean absolute deviation between the experimental and calculated enthalpies of formation was 1.07 kcal/mol, with a maximum absolute deviation of 3.01 kcal/mol and a minimum absolute deviation of 0.00 kcal/mol. Generally, the absolute deviations show that the method of atomization enthalpies provides reasonable estimates of the enthalpies of formation, offering validation of G4 as a QM method for obtaining absolute enthalpies.

The enthalpies of formation based on the method of atomization enthalpies were also calculated for the species in Figure 3.3 for which experimental data was available. The experimental and calculated enthalpies of formation for these species can be seen in Appendix B. The results are similar to those for the reference molecules, further validating the G4 method in estimating enthalpies of formation.

3.3.2. Method Validation from Isodesmic Reactions

Next, the reference molecules were used to form isodesmic reactions for the species in Figure 3.3 which had available experimental data. The calculated enthalpy of formation of the unknown species was calculated using Equation 3.11, and the average enthalpy of formation was computed from two isodesmic reactions for each species, which can be seen in Appendix B. The calculated values are summarized in Table 3.3, with the structures corresponding to the species' number shown in Figure 3.3, and compared to the experimental enthalpies of formation. Note that some of the experimental enthalpies of formation were obtained from proton affinity calculations, which can be seen in Appendix B. The mean absolute deviation between experimental and calculated enthalpies of formation using the method of isodesmic reactions was 0.58 kcal/mol, a remarkably low deviation given that the experimental uncertainties for the enthalpies of formation in this study are in the general range of 0.01-2 kcal/mol. Additionally, this value is significantly lower than that obtained from using atomization enthalpies on these species, which was 2.57 kcal/mol (Appendix B), as anticipated. These observations confirm that not only is the method of isodesmic reactions more accurate than the method of atomization enthalpies, but it also further distinguishes G4 as an appropriate and accurate method for estimating enthalpies of formation for the types of species in this study.

Table 3.1. Comparison of experimental and calculated enthalpies of formation in kcal/mol at 298 K for reference molecules using the method of atomization enthalpies. The top section includes carbenium ion species, the middle section includes oxonium ion species, and the bottom section includes neutral species. Experimental uncertainties are listed where available.

Molecule	Reference	Exp ΔH_f^o	Calc ΔH_f^o	Abs Dev
CH_3^+	Traeger and McLoughlin[124]	261.3 ± 0.4	262.73	1.43
$C_3H_7^+$	Traeger and McLoughlin[124]	191.8 ± 0.4	194.81	3.01
$C_3H_5^+$	Traeger[14,125]	227.0 ± 0.3	229.73	2.73
$C_4H_9^+$	Traeger[126]	170.0 ± 0.3	172.35	2.35
H_2C^+-OH	PA[a]	169.24 ± 0.72	170.80	1.56
H_3C-HC^+-OH	PA[a]	142.29 ± 0.75	144.19	1.90
H_3O^+	PA[a]	142.69 ± 0.71	145.56	2.87
$CH_3O^+H_2$	PA[a]	137.39 ± 1.90	139.64	2.25
CH_3-O^+H-CH_3	PA[a]	132.40 ± 2.01	134.77	2.37
CH_4	Pittam and Pilcher[127]	-17.80 ± 0.10	-17.79	0.01
C_2H_6	Pittam and Pilcher[127]	-20.04 ± 0.07	-19.81	0.23
C_2H_4	Manion[128]	12.5 ± 0.1	12.57	0.07
C_3H_8	Pittam and Pilcher[127]	-25.02 ± 0.12	-24.60	0.42
C_3H_6	Furuyama et al.[129]	4.878[b]	5.06	0.18
C_4H_6	Prosen et al.[130]	26.00 ± 0.19	26.86	0.86
H_2O	Cox et al.[131]	-57.798 ± 0.010	-57.25	0.54
CH_3OH	Hine and Arata[132]	-48.00[b]	-47.89	0.11
C_2H_5OH	Chao and Rossini[133]	-56.1 ± 0.1	-55.73	0.37
H_3C-O-CH_3	Pilcher et al.[134]	-43.99 ± 0.12	-43.99	0.00
H_2C=O	da Silva et al.[135]	-26.05 ± 0.42	-26.60	0.55
H_3C-CH=O	da Silva et al.[135]	-39.70 ± 0.12	-39.52	0.18
CH_3-CO-CH_3	Wiberg et al.[136]	-52.23 ± 0.14	-51.36	0.87
H_2C=CH-OH	Turecek and Havlas[137]	-30.6[b]	-29.28	1.32
H_2C=CH-O-CH=CH_2	Pilcher et al.[138]	-3.03 ± 0.20	-2.35	0.68
H_2C=CH-C(CH_3)=O	Guthrie[139]	-27.39 ± 2.63	-26.22	1.17
O=CH-HC=O	Fletcher and Pilcher[140]	-50.66 ± 0.19	-51.39	0.73
O=CH-OH	Guthrie[141]	-90.49[b]	-90.26	0.23
mean absolute error				1.07
minimum absolute error				0.00
maximum absolute error				3.01

Abbreviations: Exp = experimental value from literature; Calc = value from this study; Abs Dev = absolute deviation, |Calc-Exp|; PA = proton affinity calculation

[a] Experimental values were calculated from proton affinity calculations, as shown in Table 3.2

[b] Experimental uncertainty not reported in the literature.

Table 3.2. Proton affinity calculations to calculate experimental enthalpies of formation for reference molecules (Species MH^+). PA = proton affinity from Hunter.[117] Enthalpy of formation of a proton from Chase.[142] All values are in kcal/mol. Experimental uncertainties are listed where available.

M			H^+	→	MH^+	
Species M	**ΔH_f^o**	**PA**	**ΔH_f^o**		**Species MH^+**	**ΔH_f^o**
$H_2C{=}O$	-26.05 ± 0.42 [135]	170.4 ± 0.3	365.690		H_2C^+-OH	169.24 ± 0.72
H_3C-CH=O	-39.70 ± 0.35 [135]	183.7 ± 0.4	365.690		H_3C-HC^+-OH	142.29 ± 0.75
H_2O	-57.798 ± 0.010 [14,131]	165.2 ± 0.7	365.690		H_3O^+	142.69 ± 0.71
CH_3OH	-48.00 [132], *a*	180.3 ± 1.9	365.690		$CH_3O^+H_2$	137.39 ± 1.90
H_3C-O-CH_3	-43.99 ± 0.11 [134]	189.3 ± 1.9	365.690		CH_3-O^+H-CH_3	132.40 ± 2.01

a Experimental uncertainty not reported in the literature.

Table 3.3. Comparison of experimental and calculated enthalpies of formation in kcal/mol at 298 K for species using the method of isodesmic reactions. Experimental uncertainties are listed where available.

Charge	Species	Exp ΔH_f^o	Calc ΔH_f^o	Abs Dev
C^+	(1)	119.36 ± 2.04 *a*	119.43	0.07
	(2)	116.29 ± 1.90 *a*	117.19	0.90
	(3)	110.97 ± 2.10 *a*	112.13	1.16
	(26)	138.69 ± 4.53 *a*	140.40	1.71
O^+	(150)	110.99 ± 2.00 *a*	110.29	0.70
	(151)	107.99 ± 2.10 *a*	107.85	0.14
	(153)	123.99 ± 2.00 *a*	124.07	0.08
	(154)	120.66 ± 2.06 *a*	120.31	0.35
0	(25)	-67.09 ± 0.90 [143]	-67.02	0.07
mean absolute dev				0.58

a Calculated from proton affinity calculations.
Abbreviations: Exp = experimental value from literature; Calc = value from this study; Abs Dev = absolute deviation, |Calc-Exp|; C^+ = carbenium ion species; O^+ = oxonium ion species; 0 = neutral species.

3.3.3. Enthalpies of Formation from Isodesmic Reactions

With the G4 method validated, enthalpies of formation for the complete set of 195 species in Figure 3.3 were calculated using isodesmic reactions. As was the case for the species examined for method validation, two isodesmic reactions were constructed for each species, and the average enthalpy of formation was computed. A complete list of isodesmic reactions with their corresponding average enthalpy of formation can be seen in Appendix B.

3.3.4. GA Value Regression and Discussion

Before regressing new groups, 55 GA values were fixed based on known values in the literature. In some instances, literature sources fixed certain GA values to analogous ones; those same values were fixed in this study for consistency. For instance, the C-(C^+)($H)_3$ value was set to that for C-(C)($H)_3$. However, seven of these 55 known GA values, seen in Table 3.4, were independently set to analogous values in this study to solve a rank deficiency in the matrix used for regression. Specifying seven degrees of freedom ensured that all rows of the matrix were linearly independent and a unique solution was obtained. All 55 fixed GA values can be seen in Appendix B. The additional C^+–C–C_d correction was originally introduced by Bjorkman et al.[123] to account for the bending of the molecule to bring the charged carbon closer to the double bonded carbon, thereby stabilizing the molecule. Species 32 is an example of a species containing this group. The allylic *cis* correction was also originally introduced by Bjorkman et al.[123] to represent the nonbonded, next-nearest neighbor interactions of groups on either side of a resonant C^+–C_d bond. Species 26 is an example of a species containing this group. Ring strain corrections for tetrahydrofuran (e.g. species 51) and 2,5-dihydrofuran (e.g. species 53) and a correction for the

alkane gauche non-next-neighbor interaction (e.g. species 7), as defined by Sabbe et al.,[98] were included.

Table 3.4. List of seven GA values in kcal/mol that were independently set to analogous values in this study. The references that regressed the value for the analogous groups are also provided.

Group	Analogous Group	ΔH_f^o Lit. Value	Value Regressed By
CO – $(C)(O^+)$	CO – (C)(O)	-34.86	Khan et al.[109]
CO – $(C^+)(C_d)$	CO – $(C)(C_d)$	-32.71	Khan et al.[109]
C_d – $(C)(O^+)$	C_d – (C)(O)	8.94	Khan et al.[109]
O – $(C^+)(C_d)$	O – $(C)(C_d)$	-30.36	da Silva/Bozzelli[144]
C – $(C)_2(O)(O^+)$	C – $(C)_2(O)_2$	-16.2	Cohen[108]
C – $(C_d)(H)(O)(O^+)$	C – $(C_d)(H)(O)_2$	-0.3	Cohen[108]
C – $(H)_3(O^+)$	C – $(H)_3(O)$	-10	Cohen[108]

A matrix was formulated from the 195 molecules and the 71 unknown groups, which can be found in an attached Excel file titled 'GA Matrix'. Multiple linear regression was used to determine 71 unknown GA values, reduced from 78 due to the rank deficiency in the matrix, using the enthalpies of formation from the isodesmic reactions in Appendix B. These groups are listed in Table 3.5. While a majority of the groups are traditional atom-centered groups, one interaction, H–CO–X–C^+, where X can be either C or CO, was introduced to allow for next-nearest neighbor interactions within molecules involving this group that were not captured by the traditional atom-centered groups. This group accounts for the next-nearest neighbor interaction between the electronegative oxygen in the CO group and the carbocation. Species 62 is an example of a species containing this group. Non-bonded interactions such as this are commonly utilized in GA value regression.[99,107,110] Figure 3.4 shows a comparison of 195 enthalpies of formation calculated from the G4 calculations and isodesmic reactions against those calculated from the newly regressed GA

values. Specific values for all 195 species are listed in Appendix B. For the enthalpies of formation in Figure 3.4, the average deviation between the G4 and GA values was 1.91 kcal/mol, and 85% of the GA values within 4% of the G4 values. For the molecules where experimental data was available, it was compared to our GA values in Table 3.6. The mean absolute deviation was 1.23 kcal/mol. The goodness of fit of our GA values is excellent, and our results are particularly noteworthy as a significant extension to the GA values in the literature, which previously contained no GA values for oxonium ions. We report 71 groups, 65 of which were not reported previously in the literature, for a specific family of molecules, namely, oxonium ions and oxygen-containing carbenium ions. Note, however, that it is important to use the set of 71 new group values along with the fixed values from the literature used to obtain them in order to ensure that the entire set of values is self-consistent. The results also demonstrate that G4 calculations and isodesmic reactions are reasonable methods for obtaining estimates of gas-phase enthalpies of formation of a variety of species, including ions.

Table 3.5. List of 71 regressed groups with enthalpy of formation GA values in kcal/mol.[a]

Group	ΔH_f^o Value	Group	ΔH_f^o Value
C – (C)(C_d)(CO)(H)	-1.01	C_d – $(O)_2$	6.04
C – (C)(C_d)(CO)(O)	-5.83	C_d – (C^+)(CO)	27.34
C – (C)(C_d)(H)(O)	-21.91	C_d – (C^+)(O)	14.51
C – (C)(C_d)$(O)_2$	-17.25	C_d – (H)(O^+)	14.25
C – (C)$(C_d)_2$(O)	-4.27	C_d – (CO)(O^+)	23.59
C – (C_d)(CO)(H)(O)	-4.49	C_d – (O)(O^+)	13.45
C – (C_d)(CO)$(O)_2$	-12.93	CO – (C)(C^+)	-23.64
C – $(C_d)_2$(H)(O)	-5.66	CO – (C^+)(H)	-13.41
C – (C)(C^+)(CO)(O)	-15.31	CO – (C^+)(CO)	-30.48
C – (C)(C^+)(CO)(H)	-14.35	CO – (C^+)(O)	-16.68
C – (C)(C^+)(C_d)(O)	-3.30	C^+ – (C)(C_d)(CO)	194.43
C – (C)(C^+)$(O)_2$	-7.31	C^+ – (C)(CO)(H)	212.79
C – (C)(C^+)(H)(O)	-5.64	C^+ – (C)(CO)(O)	178.63
C – $(C)_2$(C^+)(O)	-7.06	C^+ – (C)(C_d)(O)	169.81
C – $(C)_2$(C^+)(CO)	-16.60	C^+ – (C)(H)(O)	185.86
C – (C^+)(CO)$(H)_2$	-17.57	C^+ – (C)$(O)_2$	155.57
C – (C^+)(CO)(H)(O)	-10.11	C^+ – $(C)_2$(CO)	204.69
C – (C^+)(C_d)(H)(O)	-0.20	C^+ – $(C)_2$(O)	174.22
C – (C^+)$(H)_2$(O)	-4.37	C^+ – (C_d)(H)(O)	176.04
C – (C^+)(H)$(O)_2$	-4.89	C^+ – (C_d)(CO)(H)	204.13
C – (C^+)(C_d)$(O)_2$	-3.91	C^+ – (C_d)(CO)(O)	168.31
C – (C)$(H)_2$(O^+)	27.31	C^+ – (C_d)$(O)_2$	154.46
C – (C)(C_d)(H)(O^+)	24.37	C^+ – $(C_d)_2$(O)	166.93
C – (C)(CO)(H)(O^+)	29.99	C^+ – (CO)$(H)_2$	221.29
C – (C)(H)(O)(O^+)	-10.60	C^+ – (CO)(H)(O)	192.42
C – (C)(CO)(O)(O^+)	23.03	O – (C)(C^+)	-26.85
C – (C)(C_d)(O)(O^+)	-12.31	O – (C^+)(H)	-35.53
C – (C)$(O)_2$(O^+)	-4.29	O^+ – (C)$(H)_2$	105.74
C – $(C)_2$(H)(O^+)	24.30	O^+ – (C)(C_d)(H)	148.27
C – $(C)_2$(CO)(O^+)	27.96	O^+ – $(C)_2$(H)	112.21
C – $(C)_3$(O^+)	22.54	O^+ – $(C)_2$(C_d)	152.77
C – (C_d)(CO)(O)(O^+)	7.67	O^+ – $(C)_3$	122.86
C – (C_d)$(H)_2$(O^+)	27.19	O^+ – (C_d)$(H)_2$	143.50
C – $(C_d)_2$(H)(O^+)	27.96	O^+ – (CO)$(H)_2$	124.75
C – (CO)(H)(O)(O^+)	27.83	H – CO – X – C^+	13.69
C – (CO)$(H)_2$(O^+)	37.81		

[a] The nomenclature follows standard GA notation as discussed in text. C_d is for a carbon atom comprising a double bond, and numerical subscripts represent the number of that atom in the group. X = C or CO.

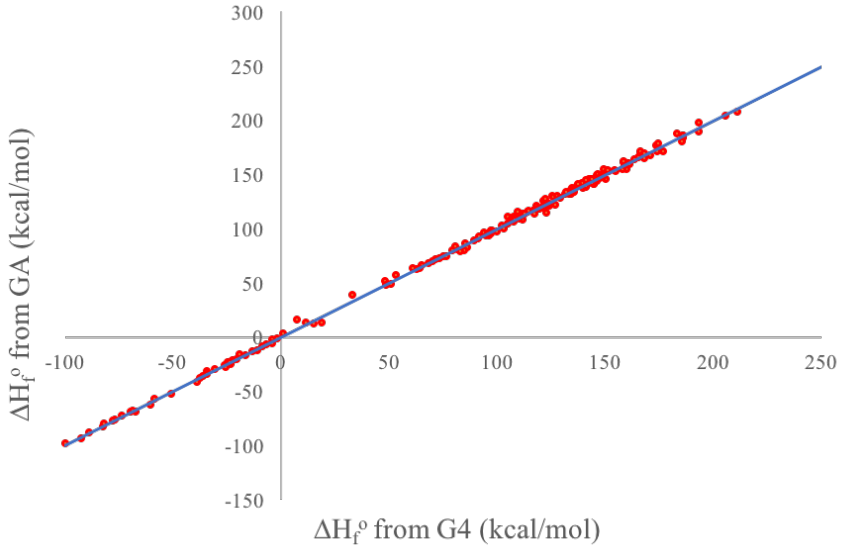


Figure 3.4. Enthalpies of formation obtained from GA values compared to G4 values for 195 species. A line of parity is plotted for comparison.

Table 3.6. Experimental enthalpies of formation compared to those calculated from GA values regressed in this study. Note that the species and their associated experimental values are the same as those reported in Table 3.3. All values are in kcal/mol.

Charge	Species	Exp ΔH_f^o	GA ΔH_f^o	Abs Dev
C^+	(1)	119.36 ± 2.04 [a]	118.20	1.16
	(2)	116.29 ± 1.90 [a]	116.87	0.58
	(3)	110.97 ± 2.10 [a]	113.30	2.33
	(26)	138.69 ± 4.53 [a]	138.55	0.14
O^+	(150)	110.99 ± 2.00 [a]	109.55	1.44
	(151)	107.99 ± 2.10 [a]	106.71	1.28
	(153)	123.99 ± 2.00 [a]	122.80	1.19
	(154)	120.66 ± 2.06 [a]	119.27	1.39
0	(25)	-67.09 ± 0.90 [143]	-68.68	1.59
	mean absolute deviation			1.23

[a] Calculated from proton affinity calculations.
Abbreviations: Exp = experimental value from literature; GA = group additivity; Abs Dev = absolute deviation, |Calc-Exp|; C^+ = carbenium ion species; O^+ = oxonium ion species; 0 = neutral species.

Finally, a bootstrapping analysis was carried out to determine the sensitivity of the regression to the specific molecules used as employed in the work by Khan et al.[109] In this method, 10% of the molecules, or 20 molecules, were removed from the matrix of 195 molecules. A regression with the remaining 175 molecules was performed, and new GA values were obtained. The enthalpies of formation for the 20 molecules removed were calculated using these new GA values. They were compared to the enthalpies of formation obtained when using the GA values from the full regression with 195 molecules. This process was repeated in four different runs, which was sufficient for drawing conclusions about the sensitivity of the regression. The results were compiled in histograms showing the deviations between the enthalpies of formation of the 20 molecules when they were included in the matrix versus when they were not included in the matrix versus when they were not included in the matrix. The histograms can be found in Appendix B. A majority of the molecules have deviations within ±2 kcal/mol, while the deviation between the experimental enthalpies of formation and those obtained from the GA values was 1.23 kcal/mol. This indicates that the GA method is still valid for the molecules not included in the matrix for regression and that no one molecule or subset of molecules is significantly perturbing the regression. Additionally, the histogram values are mostly centered around zero, showing no skewing of the data. This shows that the regression was consistent, regardless of the database of molecules used.

While Benson's group additivity scheme is relatively easy to use and supplies accurate thermodynamic properties, the number of GA values needed is high, as it requires a GA value for every distinct atom site. Lay et al.[145] developed an alternative method called the hydrogen bond increment (HBI) method with a goal of decreasing the number of parameters needed. In this

method, originally developed for radicals, a thermodynamic value for a radical (ΔH_f^o, S^{298}, or C_p) is calculated using the corresponding value for its parent molecule, in addition to an increment corresponding to the change due to the loss of a *H* atom. Several research groups have successfully developed HBI parameters for complex radical chemistry.[99,110,145,146] The same ideology could be applied for complex ionic chemistry, whereby the thermodynamic value of a carbenium or oxonium ion is calculated using the corresponding value for its radical, in addition to an increment based on the ionization energy. Therefore, as noted by Bjorkman et al.[123], not only does this method require the thermodynamic value for the corresponding radical, it also requires the selection and subsequent justification of either vertical or adiabatic ionization, as well as an accurate estimation of these ionization energies. Additionally, Sabbe et al.[99] has pointed out that different HBIs can often correspond to the same radical, introducing considerable ambiguity into the system. Nevertheless, while the investigation of this method is out of the scope of this work, it offers a well-regarded alternative to Benson's method that should be studied in the future for ionic chemistry.

3.4. Conclusions

In this work, 71 new GA values were regressed for 195 species including oxygenates, oxonium ions, and oxygen-containing carbenium ions. This significantly expands the thermodynamic data available for acid-catalyzed chemistry. In particular, these newly regressed GA values can be used to calculate estimates of enthalpies of formation for species involved in catalytic fast pyrolysis of biomass using a zeolite. Subsequently, the enthalpies of formation can be used to obtain an enthalpy of reaction followed by an activation energy, which is one of the main parameters needed to develop an accurate kinetic model. For a process as complex as CFP

of biomass, kinetic models are extremely powerful tools in predicting product concentrations and optimizing process conditions. Therefore, with this work, we have made significant progress in developing a microkinetic model of CFP of biomass. Current work is being done to combine a network of acid-catalyzed elementary reactions with the required parameters to develop an accurate microkinetic model capable of predicting product distributions and optimizing process conditions for the CFP of biomass.

Chapter 4: Microkinetic Modeling of the Vapor Phase Upgrading of Biomass-Derived Oxygenates

Material in this chapter is reproduced from the manuscript "Microkinetic Modeling of the Vapor Phase Upgrading of Biomass-Derived Oxygenates" by Lauren D. Dellon, Chun-Yi Sung, David J. Robichaud, and Linda J. Broadbelt.

4.1. Introduction

Renewable energy has garnered attention as a practical solution to the negative environmental effects imparted by the continually decreasing supply of fossil fuels. Biomass is a unique renewable energy source, in that it can be burned directly or converted to liquid biofuels or biogas. One such method for this conversion is fast pyrolysis, a thermochemical technique characterized by rapid heating, low residence times, and the absence of oxygen. Fast pyrolysis produces primarily a liquid product, known as bio-oil, a complex mixture of water and hundreds of organic compounds belonging to classes such as acids, carbonyls, and phenolics.[13-21] Unfortunately, bio-oil is incompatible with current fuel technologies by reason of its high viscosity, instability, corrosiveness, and acidity. Thus, the quality and stability of bio-oil must be improved by means of catalytic fast pyrolysis (CFP) if it is to be competitive with conventional hydrocarbon fuels. This can be achieved by employing carefully designed catalysts to upgrade the pyrolysis vapors, primarily by deoxygenation, before condensation into a liquid product.[2,15,147,148]

While there exist a host of different catalysts capable of bio-oil upgrading, zeolites have received much attention as vapor phase upgrading catalysts due to their relatively low cost, wide availability, and absence of a required H_2 input stream. Zeolites are best described as microporous crystalline aluminosilicates composed of SiO_4 and $[AlO_4]^-$ tetrahedra. These tetrahedra are interconnected in such a way that forms pores, or cavities, which can host exchangeable cations

that compensate for the negative charge imposed by the $[AlO_4]^-$. One of the most common and most effective zeolites used for bio-oil upgrading is HZSM-5, due to its high acidity, shape selectivity, and hydrothermal stability.[149,150] The protons provided by HZSM-5 enact acid-catalyzed chemistry, characterized by chemisorbed intermediates, which include alkoxides or carbenium ions (C^+) and oxonium ions (O^+).

The experimental exploration of the CFP of biomass and its quantitative description have largely been impeded by the complex chemistry imparted by the catalyst, short residence time of the vapors (1-2 s), mass transport issues, and rapid catalyst deactivation. To our knowledge, there currently exists a single empirical kinetic model in the open literature developed by Adjaye and Bakhshi wherein model compounds are used to suggest simple reaction pathways of lumped components such as alcohols, aliphatic hydrocarbons, and aromatic hydrocarbons.[93] On the other hand, computational tools, particularly microkinetic models, can be developed to simulate CFP. Microkinetic models consider all elementary steps in the reaction mechanism and make no assumptions about the rate-determining step. These microkinetic models allow for the prediction of product yields, investigation of specific reaction pathways, optimization of experimental operating conditions (e.g. temperature, pressure), and design and synthesis of more effective catalysts.

The development of a microkinetic model for the catalytic upgrading of bio-oil is quite intimidating, as the product contains more than 200 compounds.[151] Therefore, it is reasonable to begin by developing a microkinetic model for various model compounds, which are species resembling the product vapors from fast pyrolysis. This chapter discusses the development of a microkinetic model for the catalytic upgrading of two simple oxygenates present in bio-oil, acetone

and acetic acid. We explain the generation of a reaction network through the use of reaction families and the estimation of key kinetic parameters. The validity of the model is quantitatively assessed by its ability to capture the experimental data reported by Gayubo and coworkers.[152,153]

4.2. Experimental Details

Gayubo and coworkers utilized an isothermal fixed-bed reactor to observe and analyze the catalytic upgrading of various oxygenates.[152,153] The reaction equipment and procedure for product analysis are described elsewhere.[154] The physical properties of the catalyst, HZSM-5 zeolite, are as follows: the SiO_2/Al_2O_3 (SAR) is 48, and the surface area is 131 $m^2\ g^{-1}$.[152] The acid site density used in this work, 1.7 μmol/m^2, is approximated from a HZSM-5 zeolite with SAR 50.[155] The reaction conditions for the catalytic upgrading of acetone are as follows: catalyst amount, 0.420 g; space time, 0.840 (g of catalyst) h (g of acetone)$^{-1}$; pressure, atmospheric; partial pressure of acetone, 9 kPa; temperature, 400°C; water content, 50% by volume; diluted with nitrogen.[153] The reaction conditions for the catalytic upgrading of acetic acid are as follows: catalyst amount, 0.325 g; space time, 0.325 (g of catalyst) h (g of acetic acid)$^{-1}$; pressure, atmospheric; partial pressure of acetic acid, 9 kPa; temperature, 450°C; water content, 50% by volume; diluted with nitrogen.

4.3. Methodology

4.3.1. Physisorption and Chemisorption

It is well-known that molecules can adsorb on zeolites, but the term “adsorption” can be ambiguous. As such, it is necessary to clearly define the terminology used in this work, namely, physisorption and chemisorption. Physisorption is the process by which a fluid (gas-phase) species enters the pores of the zeolite, forming a π-complex. The physisorbed π-complex, which is a non-covalent interaction, is primarily characterized by hydrogen bonding, van der Waals forces, and

dispersive interactions with the zeolite framework.[156] Chemisorption occurs when the physisorbed π-complex directly interacts with the surface of the zeolite, forming a σ-complex. The precise nature of the chemisorbed intermediate has been widely debated.[157-162] More specifically, the chemisorbed species has been proposed as either a covalent alkoxide, in which the species is explicitly connected to an oxygen atom in the zeolite framework, or an ion pair with the negatively charged zeolite framework (see Figure 4.1). This distinction is necessary when assigning kinetic parameters, as alkoxides and ions differ in entropy and enthalpy.

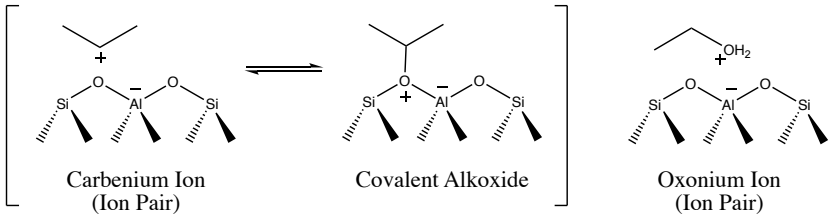


Figure 4.1. Examples of chemisorption. A carbenium ion can exist as an ion pair or a covalent alkoxide, while an oxonium ion can only exist as an ion pair. Note that the carbenium ion above is for demonstration purposes, but secondary species are more likely present as covalent alkoxides.

Kazansky and Senchenya were the first to suggest that alkoxides are the stable intermediates in zeolites, while carbenium ions are present only as transition states.[163] Since their original quantum chemical studies in 1989, various other authors have reported on the presence or absence of free carbenium ions and alkoxides.[157-161,164-166] Cnudde and coworkers used molecular dynamics simulations to show that the dominant form, either alkoxide or carbenium ion, is dependent upon the temperature, geometry of the acid site, and branching of the species.[161] The key factor, according to Boronat and Corma, is the species' accessibility to oxygen atoms in the zeolite framework.[157] As primary species have the most access to the zeolite framework, while tertiary species are limited by steric hindrance and bulkiness, the stability for alkoxides is as

follows: primary > secondary > tertiary. This is opposite to the stability trend for free carbenium ions, tertiary > secondary > primary, in which the positive charge is better stabilized by increased branching (i.e. tertiary).

Primary carbenium ions are rarely considered as intermediates due to their instability,[167,168] but primary alkoxides can exist at both low and high temperatures, as witnessed by the presence of isobutoxide in Cnudde and coworkers' molecular dynamics simulations.[161] While the form of the secondary species is more uncertain due to effects of temperature and branching, the consensus in the literature agrees with the conclusion made by Kazansky and Senchenya, namely that both primary and secondary species form stable alkoxides. Tertiary species, on the other hand, have been analyzed by studying isobutene protonation. For instance, Tuma and Sauer concluded that the tert-butyl carbenium ion is more stable than tert-butoxide due to the large entropy loss in forming tert-butoxide.[160] Cnudde and coworkers also found that tert-butoxide is limited by steric constraints, leading to the more likely presence of the tert-butyl carbenium ion.[161] Based on the reasoning above, in this work, all primary and secondary chemisorbed species are considered to be alkoxides, and all tertiary chemisorbed species are considered to be carbenium ions. For the remainder of this chapter, the term "chemisorbed intermediate" refers to either an alkoxide or carbenium/oxonium ion, while the term "physisorbed intermediate" refers to a physisorbed π-complex, described earlier.

4.3.2. Reaction Families

Formulating all the elementary steps for a large system presents a significant challenge. Further, rate constants for elementary steps cannot always be obtained directly or regressed from experimental data. To combat these issues, a reaction family approach is utilized, which operates

under the assumption that reactions with similar chemistry can be grouped together into a single type of reaction. The reaction families in this work are divided into two categories: traditional hydrocarbon chemistry and chemical reactions involving oxygen and oxonium ions.

4.3.2.1. Traditional Hydrocarbon Chemistry

Acid-catalyzed hydrocarbon chemistry, distinguished by the formation of a chemisorbed intermediate, is characterized by the addition, removal, and shifting of protons, in addition to carbon-carbon bond scission and formation.[169] The reaction families for hydrocarbon chemistry are pictured in Table 4.1. Protonation is the reaction family describing the addition of a proton to a double bond, while deprotonation is the reaction family for the removal of a proton from a chemisorbed intermediate. Hydride transfer describes the transfer of a proton from a physisorbed intermediate to a chemisorbed intermediate. Isomerization-type reaction families include hydride shift, methyl shift, and branching. Oligomerization describes the bonding between a physisorbed alkene and a chemisorbed species, while β-scission describes the breaking of a carbon-carbon bond. Finally, cyclization is the transformation of a chemisorbed intermediate into a cyclic chemisorbed intermediate.

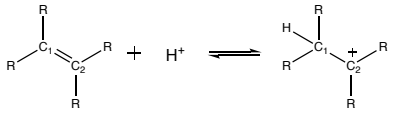
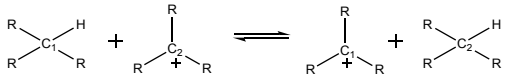
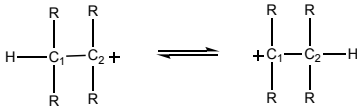
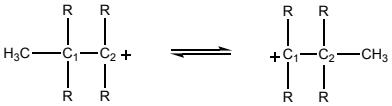
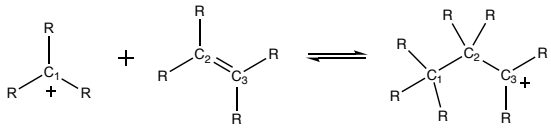
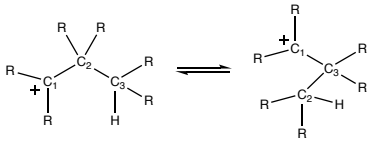
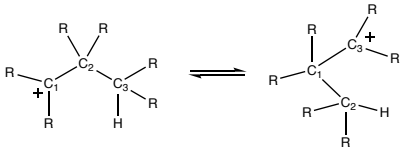
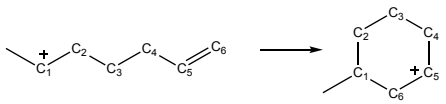
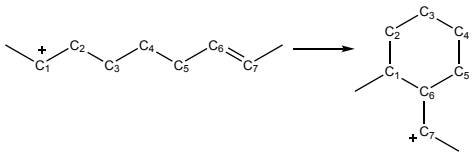
4.3.2.2. Chemistry Involving Oxygenates and Oxonium Ions

The reaction families involving oxygenates and oxonium ions are pictured in Table 4.2. Alcohol protonation is the addition of a proton to a hydroxyl group, forming an oxonium ion, and oxodeprotonation is the reverse reaction.[112] Keto protonation is the addition of a proton to a carbonyl to form a chemisorbed intermediate, and alcohol deprotonation is the reverse reaction.[112] Hydration is the reaction family describing the addition of a water molecule to a chemisorbed intermediate to form an oxonium ion, while dehydration is the reaction family for the removal of

a water molecule from a chemisorbed intermediate to form a new chemisorbed intermediate.[112] Acylium ion addition is the addition of an acylium ion to acetic acid, a very specific part of acetic acid ketonization proposed by Gumidyala and coworkers.[170] Decarboxylation, also a distinct step in ketonization, is a concerted reaction describing the removal of carbon dioxide from a β-ketoacid.[171] There are two versions of aldol condensation presented in the literature. The first is the addition of a protonated ketone/aldehyde to a ketone/aldehyde.[172] The second is the addition of a protonated ketone/aldehyde to an enol tautomer of a ketone/aldehyde.[173] Aldol condensation is further discussed in the Results and Discussion section. Finally, the Prins reaction is the addition of a protonated ketone/aldehyde to an alkene, a subset of oligomerization reactions.[174,175]

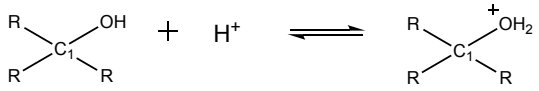
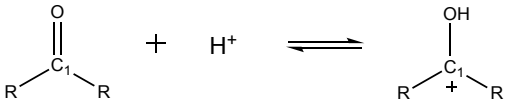
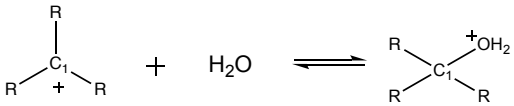
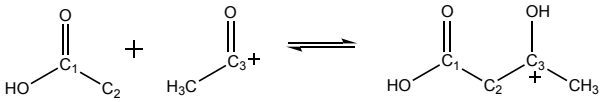
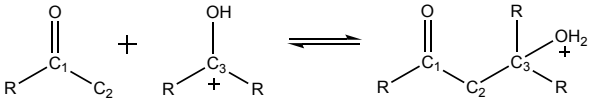
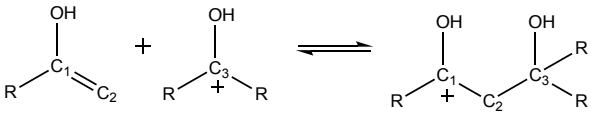
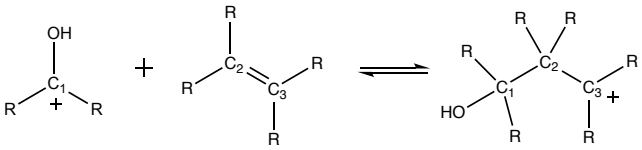
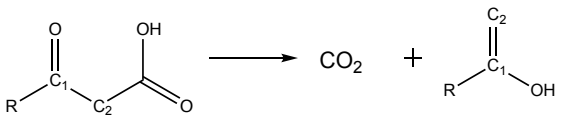
Two additional reaction families, physisorption and dephysisorption, are necessary for determining the amounts of a particular species in the fluid-phase and in the pores of the zeolite. The ratio of the frequency factor for physisorption to that for dephysisorption is estimated by calculating the entropic contribution of the physisorption of isobutene. The activation energy for physisorption is set to zero, assuming no barrier, and the activation energy for dephysisorption is set to the heat of physisorption of the species.

Table 4.1. Traditional Hydrocarbon Chemistry. [a]

Reaction Family	Example
Protonation / Deprotonation [b]	
Hydride Transfer	
Hydride Shift	
Methyl Shift	
Oligomerization / β-Scission [b]	
α PCP (Protonated Cyclopropane) Branching	
β PCP (Protonated Cyclopropane) Branching	
Cyclization to form endo-cyclohexane alkoxides or carbenium ions [c]	
Cyclization to form exo-cyclohexane alkoxides or carbenium ions [c]	

[a] Note that all chemisorbed species are shown as carbenium ions for ease of demonstration. However, primary and secondary chemisorbed species exist as alkoxides.
[b] The first reaction family listed is the forward reaction, and the second is the reverse reaction.
[c] The reverse reactions of cyclization are a form of β-scission.

Table 4.2. Reaction Families Involving Oxygenates and Oxonium Ions. [a,b]

Reaction Family	Example
Alcohol Protonation / Oxodeprotonation	
Keto Protonation / Alcohol Deprotonation	
Hydration / Dehydration	
Acylium Ion Addition / Acylium Ion Removal	
Aldol Condensation / Retro Aldol (Keto)	
Aldol Condensation / Retro Aldol (Enol)	
Prins / Reverse Prins	
Decarboxylation	

[a] Note that all chemisorbed species are shown as carbenium ions for ease of demonstration. However, primary and secondary chemisorbed species exist as alkoxides.
[b] The first reaction family listed is the forward reaction, and the second is the reverse reaction. Decarboxylation only occurs in the forward direction, as this is a quick removal of CO_2.

4.3.3. Automated Network Generation and Seeding

This work employed the automated network generator, NetGen, developed by Broadbelt and coworkers.[50,176-178] NetGen creates a network of elementary reactions by applying reaction families to the reactants and their progeny. The underlying technique relies on the representation of molecules as matrices and operations based on graph theory. More specifically, each species in the model is described by a bond and electron (BE) matrix defining the connectivity of the atoms in the species.[179] Each reaction family is assigned a reaction matrix that quantifies the changes in the electronic configuration and the connectivity among the atoms affected by the reaction. A chemical reaction can then be carried out via addition of the reaction matrix to the BE matrix of the reactant molecule.

When applying the reaction families to the reactants and their progeny, the network may become too large, in that commercial solvers are unable to solve such a large system of differential-algebraic equations describing the reactor of interest. For this reason, a technique known as seeding is used as a guide to reduce the size of the network and direct the growth of the network toward the empirically-observed species.[180] The method of seeding begins by selecting target molecules, usually those that are experimentally observed. Next, a network with a high enough rank, or extent of reaction, to obtain target species is generated. In this network, pathways from the reactant to the target molecules are traced, and intermediate species are chosen as seed molecules. These seed molecules act as reactants, or rank 0 species, and allow for a lower rank network to be generated, reducing the number of species in the network.

In this study, termination criteria[177] were applied. More specifically, products with a rank greater than one or a carbon count greater than nine were not allowed, although it is important to

note that many products with a true rank relative to the reactant, acetic acid or acetone, were formed due to the effect of seeding on rank. The rank-based termination criterion limited the extent of reaction, while the carbon-based termination criterion limited the oligomerization-type reactions, including acylium ion addition, aldol condensation and Prins reaction, so as to prevent the formation of heavy ionic intermediates not associated with a rank increase.

4.3.4. Estimation of Kinetic Parameters

The rate constant for each elementary step is calculated from the Arrhenius equation (Equation 4.1):

$$k = Ae^{-E_a/RT} \tag{4.1}$$

where k is the reaction rate constant, A is the frequency factor (or pre-exponential factor), E_a is the activation energy, R is the ideal gas constant, and T is the temperature. The activation energy is related to the enthalpy of reaction using a structure-reactivity relationship known as the Evans-Polanyi relationship (Equation 4.2):

$$E_a = E_o + \alpha \Delta H_{rxn} \tag{4.2}$$

where E_a is the activation energy for the reaction of interest, ΔH_{rxn} is the enthalpy of reaction for the reaction of interest, E_o is the intrinsic activation barrier, and α is a measure of how far along the transition state is on the reaction coordinate. Reactions in the same family share the same A, E_o, and α, thus differing only by ΔH_{rxn}. The following subsections describe the calculation of ΔH_{rxn} and the estimation of the kinetic parameters A, E_o, and α.

4.3.4.1. Enthalpies of Reaction

The enthalpies of reaction on the zeolite surface are calculated from the gas phase enthalpy of reaction and the adjustments for surface effects, which include physisorption and

chemisorption. For example, for a hydride transfer reaction between a molecule R_1H and a carbenium ion R_2^+ (Equation 4.3):

$$R_1H + R_2^+ \rightarrow R_1^+ + R_2H \tag{4.3}$$

the gas phase enthalpy of reaction can be calculated as the algebraic sum of the enthalpies of formation for the products and the reactants (Equation 4.4):

$$\Delta H_{rxn,g} = \Delta H_{f,g}(R_2H) + \Delta H_{f,g}(R_1^+) - \Delta H_{f,g}(R_2^+) - \Delta H_{f,g}(R_1H) \tag{4.4}$$

Similarly, the enthalpy of reaction on the surface can be calculated from the enthalpies of formation of the species in the adsorbed state (Equation 4.5):

$$\Delta H_{rxn,ads} = \Delta H_{f,ads}(R_2H) + \Delta H_{f,ads}(R_1^+) - \Delta H_{f,ads}(R_2^+) - \Delta H_{f,ads}(R_1H) \tag{4.5}$$

The enthalpies of formation of a molecule and a carbenium ion in the adsorbed states can be related to those in the gas phase, respectively, by means of Equations 4.6 and 4.7:

$$\Delta H_{f,ads}(R_iH) = \Delta H_{f,g}(R_iH) + \Delta H_{ads}(R_iH) \tag{4.6}$$

$$\Delta H_{f,ads}(R_i^+) = \Delta H_{f,g}(R_i^+) - \Delta q(R_i^+) \tag{4.7}$$

where $\Delta H_{ads}(R_iH)$ and $\Delta q(R_i^+)$ denote the enthalpy of adsorption of the molecule R_iH and the stabilization energy for the carbenium ion R_i^+, respectively, on the zeolite surface. Note that the enthalpy of adsorption is a negative value. Equation 4.5 can then be expressed in terms of $\Delta H_{rxn,g}$ (Equation 4.8):

$$\Delta H_{rxn,ads} = \Delta H_{rxn,g} - \Delta H_{ads}(R_1H) + \Delta q(R_2^+) - \Delta q(R_1^+) + \Delta H_{ads}(R_2H) \tag{4.8}$$

4.3.4.2. Gas Phase Enthalpy of Reaction and Enthalpies of Formation

To calculate the enthalpy of reaction in the gas phase via Equation 4.4, $\Delta H_{rxn,g}$, a hierarchy is employed for obtaining the enthalpy of formation of a gaseous species, $\Delta H_{f,g}$, in the following order with decreasing priority: first, experimental values from the NIST database[181];

second, theoretical values from quantum chemical calculations reported in previous studies; last, group additivity values developed in previous studies based on Benson's group additivity scheme via the decomposition of a molecule into structural groups.[98,99,101,105,108,110,123,135,144,182,183]

4.3.4.3. Physisorption Enthalpies

Physisorption enthalpies of neutral molecules, ΔH_{ads}, are estimated using a group contribution method, whereby each atom or group of atoms is assigned a physisorption enthalpy that contributes to the total physisorption enthalpy for the neutral species. The enthalpy of physisorption for each molecule is formulated based on a contribution for each alkane carbon, alkene carbon, aromatic carbon, -OH group, =O group, and -OOH group that does not depend on its neighboring environment. These group contributions to the total physisorption enthalpy are seen in Table 4.3.

Table 4.3. Group Contributions for Physisorption Enthalpies

Group	Contribution (kcal/mol)
Alkane Carbon	3.0
Alkene Carbon	6
Aromatic Carbon	2.5
-OH Group	23
=O Group	22.1
-OOH Group	23

Appendix C shows an example of the physisorption calculation for a molecule. These group contribution values were obtained by correlating the enthalpy of physisorption, obtained either experimentally or theoretically, with the carbon number. To obtain the enthalpy of physisorption per alkane carbon, the experimental enthalpies of physisorption for propane, n-butane, n-pentane, and n-hexane were regressed as a function of carbon number.[184] The experimental enthalpies are found in Appendix C. To determine the contribution from an alkene

carbon, the enthalpies of physisorption for 1/2/3/4-alkenes, obtained by Marin's group using DFT calculations, were analyzed.[185] More specifically, the contribution from alkane carbons was subtracted from the total enthalpy of physisorption to obtain only the contribution per alkene carbon. The contribution for each alkene was plotted, and the average contribution per alkene carbon was calculated. A similar procedure was carried out for the -OH[186] and -OOH groups[187]. The contributions for the alkene carbons, -OH groups, and -OOH groups can be found in Appendix C. For the =O group, the contribution from alkane carbons was subtracted from the experimental enthalpy of physisorption for acetone,[188] 31.07 kcal/mol, to obtain only the contribution per =O group. For the aromatic carbon group, the enthalpy of physisorption for benzene,[189] 15.2 kcal/mol, was divided by six to obtain the contribution per aromatic carbon. The resulting plots and the aforementioned physisorption enthalpy calculations can be seen in Appendix C.

The physisorption values from the previous calculations can be considered as valid for adsorption to a single empty site. However, it is widely known that the observed physisorption enthalpy is dependent upon surface coverage of the zeolite. That is, when a site is empty, a molecule adsorbs with its full physisorption enthalpy. When a site is occupied, a subsequent molecule will adsorb with a decreased physisorption enthalpy. To determine this decreased physisorption enthalpy, we refer to coadsorption of two molecules. For example, methanol to dimethyl ether conversion requires the coadsorption of two methanol molecules. The first methanol physisorbs with an energy of -0.63 eV, while the second physisorbs with an energy of -0.40 eV.[190] Similarly, ethanol condensation requires the coadsorption of two ethanol molecules. The first ethanol physisorbs with an energy of -49.8 kcal/mol, while the second physisorbs with an energy of -29.7 kcal/mol.[191] Based on these two coadsorption examples, it is reasonable to

assume that for a fully covered surface, the relevant regime in this work, the physisorption of a molecule is 60% of its total physisorption energy. That is, the physisorption value used in Equation 4.8 is 60% of the physisorption value calculated by the group contribution method discussed above.

4.3.4.4. Stabilization Energies

The stabilization energy, Δq, of an ionic species is defined as the difference in energy between the gas phase ionic species and the chemisorbed complex. It can be calculated for a carbenium ion or alkoxide using Equation 4.9:

$$\Delta q(RH^+) = \Delta q(H^+) - \Delta E_{chem} - PA(R) \tag{4.9}$$

and for an oxonium ion using Equation 4.10:

$$\Delta q(ROH_2^+) = \Delta q(H^+) - \Delta E_{chem} - PA(ROH) \tag{4.10}$$

where ΔE_{chem}, $\Delta q(H^+)$, and PA (proton affinity) are defined in Figure 4.2. Note the difference in notation in Equations 4.9 and 4.10 from Equation 4.3. A stabilization energy is needed for each ionic species of which there are 8 types, depicted in Figure 4.3. The stabilization energy for a proton, $\Delta q(H^+)$, is zeolite dependent; for HZSM-5, the value is 289 kcal/mol.[192] Proton affinities, PA, were estimated by either De Moor and coworkers[193] or obtained from the NIST WebBook[194]. The stabilization energy, Δq, is assumed to have a carbon number-dependent portion, which is constant for all chemisorbed intermediates of types 1-5, in addition to an ion-dependent portion. The following discussion is for the estimation of the carbon number-dependent portion, followed by the ion-dependent portion for all chemisorbed intermediates of types 1-5.

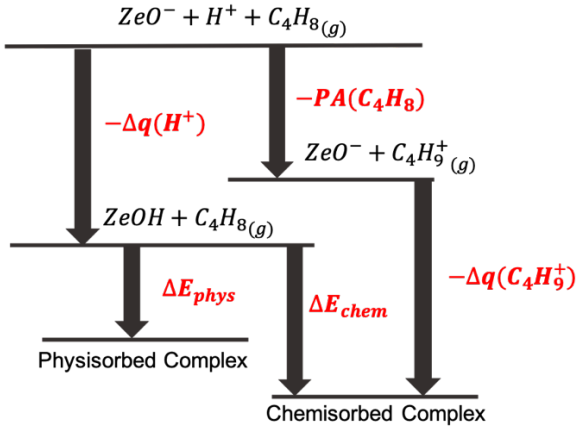


Figure 4.2. Energy diagram for the example of isobutene. Adapted from De Moor and co-workers.[186]

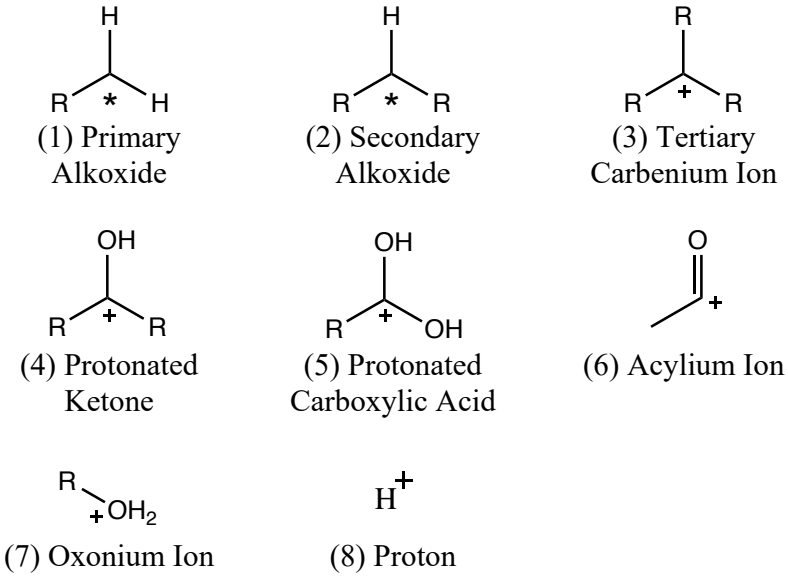


Figure 4.3. Chemisorbed species' types. Note that an asterisk represents a connection to an oxygen atom on the zeolite framework, while a positive charge represents an ion. 'R' represents any additional group that is not a hydrogen. Note that the acylium ion and proton are single species, while the remainder of the types represent groups of species.

Nguyen and coworkers used periodic DFT-D and statistical thermodynamics calculations to estimate the chemisorption energy, ΔE_{chem}, for the transformation of various 1-alkenes to their corresponding 2-alkoxides.[185] These values are found in Appendix C. The chemisorption energy for 1-nonene, which was not estimated by De Moor and coworkers, was linearly extrapolated from the chemisorption energies for carbon numbers 3 through 8 (excluding 7, as the proton affinity for 1-heptene was not available). The proton affinity for 1-nonene was assumed to be equal to the proton affinity for 1-octene under the assumption that the proton affinity levels off at carbon number 8.

The stabilization energies, Δq, are calculated from Equation 4.9 and plotted in Figure 4.4. The stabilization energy follows a parabolic trend with respect to the carbon number. This can be rationalized by considering that the stabilization energy includes effects from both local and non-local interactions. Local interactions are those of and near the charge center with the zeolite oxygen atom and the ability to distribute this charge among the other atoms in the ion. Thus, as an ion grows in carbon number, it has better distribution of charge, increasing the likelihood that it will remain in the fluid phase (i.e. decreasing its stabilization). However, this effect plateaus at a carbon number of about 6, as the charge can only be distributed to a certain extent. As the carbon number continues to increase, non-local interactions come into play, which are those of the other atoms in the ion with the zeolite framework. These non-local interactions will include hydrogen bond and dispersive interactions, much like the physisorption interactions. Under the assumption that the carbon number-dependent quantity remains constant for all chemisorbed intermediates of types 1-5, the stabilization energy for any chemisorbed intermediate in the model, excluding acylium and oxonium ions, will have the form of Equation 4.11:

$$\Delta q = 0.6(CN)^2 - 8.2(CN) + \Delta q^o \quad (4.11)$$

where CN is the carbon number and Δq^o is the ion-dependent portion of the stabilization energy. The Δq^o value for each type of chemisorbed intermediate 1-5 is seen in Table 4.4. Note that these Δq^o and Δq values were derived from theory and correlations and were not adjustable parameters in the model.

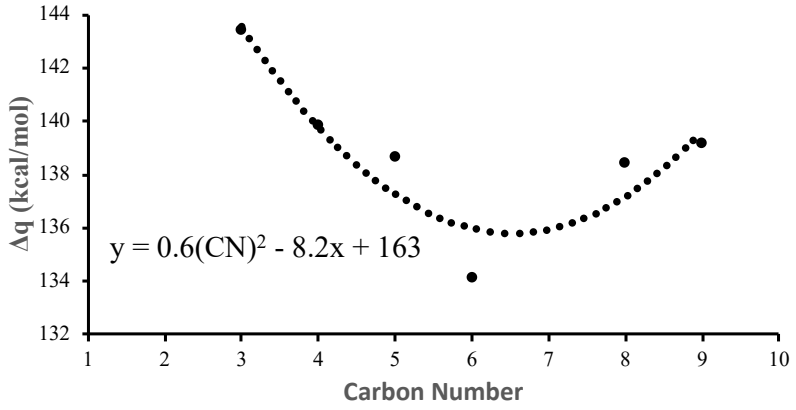


Figure 4.4. Stabilization energies for the transformation of 1-alkenes to 2-alkoxides. Each data point was calculated using Equation 4.9. There is no data point for carbon number 7 as the proton affinity of 1-heptene was not available in the literature. The parabolic trend highlights the effects from local and non-local interactions. Local interactions are responsible for the left portion (decreasing stabilization with increasing carbon number) of the graph, while non-local interactions come into play as the carbon number increases beyond 6.

Table 4.4. Stabilization Energies.

$\Delta q = 0.6(CN)^2 - 8.2(CN) + \Delta q^o$	
Chemisorbed Intermediate	Δq^o (kcal/mol)
(1) Primary Alkoxide	166
(2) Secondary Alkoxide	163
(3) Tertiary Carbenium Ion	138
(4) Protonated Ketone	150
(5) Protonated Carboxylic Acid	151
(6) Acylium Ion	$\Delta q = 139$
(7) Oxonium Ion	$\Delta q = 0.3(CN)^2 + 136$
(8) Proton	$\Delta q = 289$

(CN) = carbon number

The Δq^o value for secondary alkoxides is obtained directly from Figure 4.4. The Δq^o value for primary alkoxides is obtained from the stabilization energy for the transformation of ethene to ethoxy.[185] The Δq^o value for tertiary carbenium ions is obtained from the stabilization energy for the transformation of isobutene to the tert-butyl carbenium ion.[162] The Δq^o value for the protonated ketone intermediate, considered to be a tertiary species, is calculated by requiring the heat of reaction for the protonation of acetone to be consistent with the results obtained by Herrmann and Iglesia[172] and then using Equation 4.8 to back out the stabilization energy. We used the work of Kouskoulli and coworkers[195], in which they investigated stabilization energies for oxy-substituted carbocations, to estimate the stabilization energies of the protonated carboxylic acid chemisorbed intermediate. More specifically, they determined that the stabilization energy per hydroxyl group is 61.6 kcal/mol for one hydroxyl and 51.4 kcal/mol per hydroxyl when there are two hydroxyls. We estimated the stabilization energy per methyl group using the stabilization energy of the protonated ketone and the stabilization energy per hydroxyl group for one hydroxyl. Using the stabilization energy per methyl group and per hydroxyl group, we estimated the stabilization energy for the protonated carboxylic acid chemisorbed intermediate. The stabilization energy for the acylium ion was estimated by requiring the protonation of ketene to the acylium ion to be consistent with the results from Yan and coworkers.[196]

Nguyen and coworkers also used DFT calculations to estimate the chemisorption energy of the transformation of various n-alcohols to their corresponding oxonium ion complexes.[186] The chemisorption and stabilization energies of the oxonium complexes and the proton affinities for the alcohols are found in Appendix C. Note that the chemisorption energies for n-alcohols with carbon numbers 5-9 were linearly extrapolated from the chemisorption

energies for the n-alcohols with carbon numbers 1-4. The plot of stabilization energies, calculated from Equation 4.10, is seen in Figure 4.5. This plot follows the same trend as that seen in Figure 4.4. The stabilization energy for any chemisorbed intermediate of the oxonium ion type will have the form of Equation 4.12:

$$\Delta q = 0.3(CN)^2 + 136 \tag{4.12}$$

where CN is the carbon number.

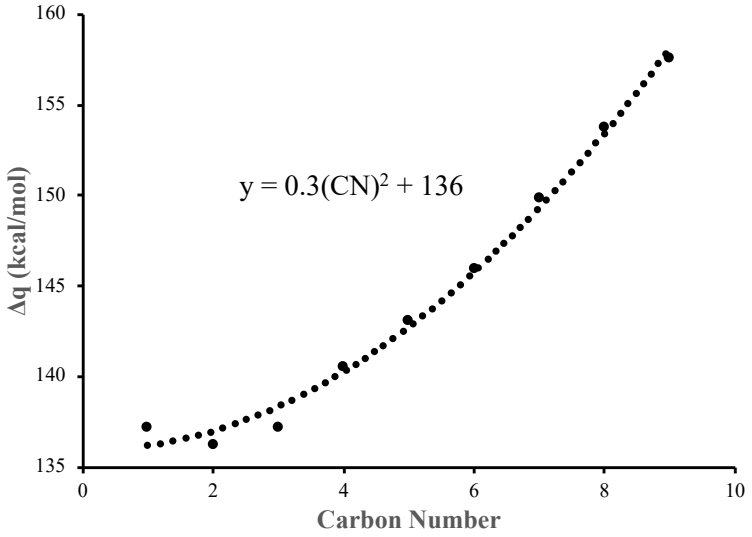


Figure 4.5. Stabilization energies for the transformation of n-alcohols to oxonium complexes. Each data point was calculated using Equation 4.10. The slight parabolic trend highlights the effects from non-local interactions.

As mentioned, the stabilization energy is a combination of both local and non-local interactions. The Δq^o value accounts for the local interactions of the charge center with the oxygen atom in the zeolite framework. However, local interactions also include the contributions from oxygen atoms *near* the charge center. For a first or second local neighbor (ketone, alcohol, or carboxylic acid) to a positively charged carbon atom, a 16.8 kcal/mol contribution is added to the stabilization energy for that chemisorbed species. Additionally, for a first, second, or third local

neighbor to a positively charged oxygen atom, the stabilization energy is decreased by an additional 3.8 kcal/mol. These values were obtained by backing out the stabilization energies of protonated diacetone alcohol and protonated mesityl oxide from Equation 4.8, where the heat of reaction in the adsorbed phase is provided by the DFT results of Herrmann and Iglesia.[172]

Non-local interactions also contribute to the stabilization energy. The carbon number-dependent portion of the stabilization energy accounts for the non-local interactions from the carbon atoms in the species. However, non-local interactions also include the contributions from oxygen atoms *far* from the charge center. These contributions are equal to the physisorption energy experienced by a ketone, alcohol, or carboxylic acid. More specifically, for a hydroxyl moiety far from the charge center, 23 kcal/mol (from Table 4.3) is added to the stabilization energy. Similarly, for a carbonyl moiety far from the charge center, 22.1 kcal/mol is added to the stabilization energy. Finally, for a carboxylic acid group far from the charge center, 23 kcal/mol is added to the stabilization energy. Appenix C provides examples of the calculation of the stabilization energy for a carbenium ion and an oxonium ion, respectively, including the local and non-local interactions.

4.3.4.5. Evans-Polanyi Constants

The Evans-Polanyi equation, Equation 4.2, includes two parameters, E_o and α. E_o is the activation energy of the reference reaction of the reaction family, and α characterizes the position of a transition state along the reaction coordinate such that $0 \leq \alpha \leq 1$. The value of α is estimated from the exothermicity of the reaction family. More exothermic reaction families are assigned a value of α closer to 0, while endothermic reaction families are assigned a value of α closer to 1. The microreversibility of a reaction is maintained by setting $\alpha_{reverse} = 1 - \alpha_{forward}$,

so that the difference in the activation energies of the forward and reverse directions equals the enthalpy of reaction. The Evans-Polanyi relationship assumes that reactions in the same family share the same E_o and α.

4.3.4.6. Frequency Factors

On the basis of transition state theory that assumes that the transformation of reactants to products proceeds through an activated complex (or transition state structure) in quasi-equilibrium with the reactants, the frequency factor in the rate constant in the Arrhenius form (Equation 4.1) can be estimated as:

$$A = \frac{k_B T}{h} e^{\Delta S^{\ddagger}/R} e^{(1-\Delta n^{\ddagger})} (c^0)^{\Delta n^{\ddagger}} \tag{4.13}$$

where k_B denotes Boltzmann's constant, T is the temperature, h is Planck's constant, c^0 represents the standard state concentration (1 M), $\Delta n^{\ddagger}$ is the change in the number of moles in going from the reactants to the transition state, $\Delta S^{\ddagger}$ signifies the entropy change between the reactants and the activated complex, and R is the gas constant. A discussion of the estimated and optimized frequency factors is found in the Results and Discussion section.

4.3.5. Model Parameters and Solution

The rate expressions for the elementary steps based on mass action kinetics are incorporated into the appropriate reactor design equation. The fixed bed reactor used in the experiments of Gayubo and coworkers[153] can be modeled as a plug-flow reactor. The resulting system of ordinary differential equations, describing the change in concentration of each species in the model, and one algebraic equation, describing the mass balance for surface coverage, was solved with the DDASL solver. Parameter estimation was carried out using a gradient-based local

optimizer (GREG). The error between the model results and the experimental data was quantified according to the sum of squared error (SSE):

$$SSE = \sum_{i} (x_{i,exp} - x_{i,model})^2 \tag{4.14}$$

where i represents a species or group of species with experimental data and x is the weight percent of the species or group of species. Equation 4.14 is also the objective function that was minimized by GREG.

4.4. Results and Discussion

4.4.1. Reaction Network

By applying the reaction families in Tables 4.1 and 4.2 to the reactants, acetic acid and acetone, a reaction network with over 10,000 unique species was generated. The resulting network of ordinary differential equations was too large for solution by a conventional solver due to the significant stiffness in the system. Therefore, a reduction in the size of the reaction network was necessary. As mentioned, seeding is a common method to reduce the size of a network by "seeding" intermediate neutral species as reactant species in order to direct the growth toward the empirically observed species. Pathways were traced from the original reactant, acetic acid, to the target molecules, namely, aromatics (i.e. benzene, xylene, toluene, trimethylbenzene). The intermediate neutral species were selected as seeds. A complete list of the 17 seed molecules, including the original reactants, acetic acid and acetone, can be found in Appendix C.

The final reaction network includes a total of 314 unique neutral species, in addition to the inert gas N_2, which is assumed to have no interaction with any of the species in the model, and 266 surface (chemisorbed) species. There is a total of 2,160 reactions, 1,532 of which are the reaction

families listed in Tables 4.1 and 4.2, and 628 are for physisorption and dephysisorption (314 each) of each neutral species. An inventory of the reactions can be found in Appendix C.

4.4.2. Model Parameters

The process of specifying and optimizing the model parameters involved several steps. First, an estimate for each parameter was made. For each reaction family in Tables 4.1 and 4.2, three parameters were estimated: $A, E_o,$ and α. A tabulation of the estimated frequency factors is seen in Table 4.5. The original estimates of the frequency factor were estimated from transition state theory. As mentioned earlier, chemisorbed species can exist as either alkoxides for primary and secondary species, or carbenium ions for tertiary species. The distinction is necessary when estimating the frequency factors because the entropy changes upon reaction are dependent upon the type of chemisorbed species involved in the reaction. Nguyen and coworkers calculated the entropy changes, depicted in Appendix C, for the protonation of physisorbed isobutene via the transition state toward isobutoxide and tert-butyl carbenium ion on H-ZSM5 with periodic density functional theory calculations and statistical thermodynamics.[162] An entropy loss of 70 J $mol^{-1}K^{-1}$ was calculated from physisorbed isobutene to the transition state toward isobutoxide. Similarly, an entropy loss of 54 J $mol^{-1}K^{-1}$ was calculated from physisorbed isobutene to the transition state toward t-butyl carbenium ion. Therefore, we estimate the frequency factor for protonation to be 4.08×10^3 $Pa^{-1}s^{-1}$ for primary and secondary species (alkoxides) and 2.80×10^4 $Pa^{-1}s^{-1}$ for tertiary species (carbenium ions). While these estimates were derived from isobutene protonation, the frequency factor for alkoxides can be generalized to reactions of the same molecularity (i.e. bimolecular), including alcohol protonation, hydration, aldol condensation, acylium ion addition,

and oligomerization. The order of magnitude of the frequency factor for hydride transfer was estimated from a kinetic model for isobutane cracking over zeolites.[197]

Similarly, an entropy loss of 5 J $mol^{-1}K^{-1}$ was calculated from isobutoxide to the transition state toward physisorbed isobutene, and an entropy loss of 71 J $mol^{-1}K^{-1}$ was calculated from tert-butyl carbenium ion to the transition state toward physisorbed isobutene. Therefore, we estimate the frequency factor for deprotonation to be $2.09x10^{13}$ s^{-1} for primary and secondary species (alkoxides) and $7.45x10^{9}$ s^{-1} for tertiary species (carbenium ions). Again, while these estimates were derived from chemisorbed isobutene deprotonation, the frequency factor for alkoxides can be generalized to reactions of the same molecularity (i.e. unimolecular), including oxodeprotonation, dehydration, retro aldol, acylium ion removal, and β-scission. We assume an entropy change, $\Delta S^{\neq}$, of zero for all isomerization reactions, corresponding to a frequency factor of $3.81x10^{13}$ s^{-1}. Finally, a frequency factor for dephysisorption was assumed, ensuring that it was sufficiently high for quasi-equilibrium between the fluid phase and physisorbed species to be achieved. The frequency factor of physisorption was calculated from the ratio of the frequency factors for physisorption to dephysisorption, which was calculated to be $8.72x10^{-7}$, corresponding to an entropy loss of 116 J $mol^{-1}K^{-1}$ upon physisorption.[162]

It is also important to note that several reaction families of a similar type were grouped together to have a single frequency factor, so as to lower the number of independent parameters. For example, the frequency factor for alcohol deprotonation was equated to the frequency factor for deprotonation of tertiary carbenium ions, as the chemisorbed intermediates for these two reaction families are of the same type. Similarly, all isomerization reaction families were assigned the same estimated frequency factor. Thus, there are 17 total independent frequency factors.

A list of the intrinsic activation barriers, E_o, can be seen in Table 4.6. As with the frequency factors, several reaction families of a similar type were grouped together to have a single intrinsic activation barrier. To further reduce the number of independent parameters, the ratio of the intrinsic activation barrier for protonation for tertiary species to the intrinsic activation barrier for protonation for primary/secondary species, 0.37, was used, which was estimated from the protonation of isobutene.[162] Additionally, the intrinsic activation barrier for the isomerization reactions, including ring cyclization, 1,2-hydride shift and 1,2-methyl shift was set equal to zero. As a result, there are 10 total independent intrinsic activation barriers.

The estimated α values can be seen in Table 4.7. As mentioned earlier, the exothermic reactions were assigned a value of α less than 0.50, while the endothermic reactions were assigned a value of α greater than 0.50. Isomerization reactions were assigned a value of α of 0.50. All of the α values were fixed.

The experiments provided a total of 20 data points. Both the transformation of acetone and the transformation of acetic acid had 10 data points, one for each of the following categories: acetone, acetic acid, olefins, propene, aromatics, ethene, isobutene, n-butenes, paraffins, and oxygenates. As a result, a maximum of 20 independent parameters could be optimized by GREG. After the first step of estimation, there are a total of 29 independent parameters. Thus, not all of the parameters could be optimized. Next, a sensitivity analysis was performed to determine the effect of various parameters. The results were not very sensitive to the frequency factors for hydration, hydride transfer for primary and secondary species, isomerization reactions, decarboxylation, and physisorption and dephysisorption. As a result, the frequency factors for these reaction families were not adjusted from their estimated values. The sensitivity analysis

revealed that a higher frequency factor for the deprotonation reactions was required, so the frequency factor for the deprotonation reactions was fixed to a reasonable value. Additionally, for cases in which the results were not very sensitive to the parameter, it was assigned a reasonable value and removed as an optimizable parameter. After this process, there were a total of 16 independent parameters and 20 data points to perform the optimization.

Table 4.5. Estimated and optimized frequency factors for each elementary reaction family.

Reaction Family	Estimated Frequency Factor, A (s^{-1} or $Pa^{-1} s^{-1}$)	Fixed A [a]	Optimized A
Protonation (p, s) [b]; Alcohol Protonation	4.08×10^{3}	-	2.36×10^{3}
Protonation (t) [b]; Keto Protonation	2.80×10^{4}	-	2.55×10^{4}
Deprotonation (p, s) [b]; Oxodeprotonation; Dehydration	2.09×10^{13}	2.09×10^{15}	-
Deprotonation (t) [b]; Alcohol Deprotonation	7.45×10^{9}	5.98×10^{12}	-
Hydration	4.08×10^{3}	[c]	-
Aldol Condensation	4.08×10^{3}	-	1.89×10^{4}
Retro Aldol	2.09×10^{13}	2.55×10^{13}	-
Acylium Ion Addition	4.08×10^{3}	8.14×10^{2}	-
Acylium Ion Removal	2.09×10^{13}	2.77×10^{13}	-
Oligomerization	4.08×10^{3}	-	4.89×10^{4}
β-Scission	2.09×10^{13}	-	1.35×10^{13}
Hydride Transfer (p, s) [b]	1.00×10^{4}	[c]	-
Hydride Transfer (t) [b]; Cyclic Hydride Transfer	1.00×10^{4}	-	2.33×10^{3}
Ring Cyclization (endo/exo); 1,2-Hydride Shift; 1,2-Methyl Shift; α/β PCP Branching	3.81×10^{13}	[c]	-
Decarboxylation	1.00×10^{13}	[c]	-
Physisorption [d]	8.72×10^{2}	[c]	-
Dephysisorption [d]	1.00×10^{9}	[c]	-

[a] After testing the sensitivity of parameters, several frequency factors were fixed.

[b] (p, s) is primary and secondary alkoxides, while (t) is tertiary carbenium ions.

[c] Frequency factors for hydride transfer (p,s), hydration, ring cyclization, 1,2-hydride shift, 1,2-methyl shift, α/β PCP branching, decarboxylation, and physisorption and dephysisorption were not adjusted from their estimated values.

[d] A frequency factor for dephysisorption was assumed, and the frequency factor for physisorption was determined from a ratio calculated from entropy estimates for the physisorption of isobutene.

Table 4.6. Intrinsic activation barriers, E_o, for each elementary reaction family. Values that were fixed are denoted in the footnoes; all other values were optimized.

Reaction Family	E_o (kcal/mol)
Protonation/Deprotonation (p, s) [a]	9.8
Protonation/Deprotonation (t) [a]; Keto Protonation/Alcohol Deprotonation	3.6 [b]
Alcohol Protonation/Oxodeprotonation	7.8
Hydration/Dehydration	6.5
Aldol Condensation/Retro Aldol	17.7 [c]
Acylium Ion Addition/Removal	10.1
Oligomerization/β-Scission	11.4
Hydride Transfer (p, s) [a]	13.7
Hydride Transfer (t) [a]	6.9
Cyclic Hydride Transfer	4.0
Ring Cyclization (endo/exo); 1,2-Hydride Shift; 1,2-Methyl Shift	0.0 [c]
α PCP Branching	0.4
β PCP Branching	16.6

[a] (p, s) is primary and secondary alkoxides, while (t) is tertiary carbenium ions.

[b] This intrinsic activation barrier was determined by a ratio.

c This intrinsic activation barrier was fixed after performing a sensitivity analysis.

Table 4.7. Estimated α values for each elementary reaction family.

Reaction Family	Estimated α
Protonation/Deprotonation (p, s, t) [a]	0.30/0.70
Alcohol Protonation/Oxodeprotonation	0.30/0.70
Keto Protonation/Alcohol Deprotonation	0.30/0.70
Hydration/Dehydration	0.35/0.65
Aldol Condensation/Retro Aldol	0.40/0.60
Acylium Ion Addition/Removal	0.30/0.70
Oligomerization/β-Scission	0.35/0.65
Hydride Transfer (p, s, t) [a]; Cyclic Hydride Transfer; Ring Cyclization (Endo); Ring Cyclization (Exo); 1,2-Hydride Shift; 1,2-Methyl Shift; α PCP Branching; β PCP Branching	0.50

[a] (p, s, t) is primary/secondary alkoxides and tertiary carbenium ions.

Note: Reverse reactions have a value of 1-α to ensure microscopic reversibility.

4.4.3. Model Validation

As described in the Experimental Details section, the model was validated against the experimental data provided by Gayubo and coworkers.[153] The authors provide the evolution with time on stream of the distribution of the hydrocarbon products in the transformation of acetone or acetic acid as reactants. However, deactivation of the catalyst is a concern, especially for the transformation of acetic acid. Thus, we validated our model against the experimental data at a time on stream of zero, so as to simulate a fresh catalyst. Further experiments quantifying the rate of catalyst deactivation would allow for the validation at multiple times on stream.

Figures 4.6 and 4.7 provide the model and experimental product yields in weight percent for the transformation of acetone at 400°C and the transformation of acetic acid at 450°C, respectively. The products are separated into the same seven categories defined by Gayubo and coworkers: C5+ olefins, C4+ paraffins, aromatics, propene, ethene, isobutene and n-butenes. While Gayubo and coworkers do not report which species make up a particular group, our model provides complete speciation. While the C5+ olefins group consists of 193 species, 11 of these species comprise the majority (~99 wt%) of the total mass yield for the group. As shown Appendix C, which provides a depiction of these 11 species, seven of them are diolefins, while four of them are monoolefins. The C4+ paraffins group is comprised of 17 species, which is much less than the number of olefins due to the fact that only 49 of the 193 olefins are monoolefins, limiting the number of $C_nH_{2n+1}^+$ ions in the system. There is not an equal number of paraffins and monoolefins because the termination criterion prevents rank 1 species from protonating, and some of the monoolefins are formed via β-scission, rather than via deprotonation of the corresponding $C_nH_{2n+1}^+$ ion, which would also lead to the corresponding paraffin. Additionally, isobutane is the

only paraffin with a noticeable mass yield in the product stream. This is due to the build-up of isobutene (rank 0), and hence, the tert-butyl cation, which can undergo hydride transfer to form isobutane. Finally, while the aromatics group consists of mesitylene, xylene isomers, ethylbenzene and benzene, mesitylene (trimethylbenzene) is the only aromatic with a noticeable yield, which will be elaborated on later when the net rate analysis is discussed. However, it is important to note that bimolecular disproportionation and transalkylation of aromatics[198] were not included as reaction families in the model. Thus, while mesitylene is the only aromatic product in the model with a high yield, it could react further to form other aromatic species if higher rank transformation's typical of conversion of aromatics on acid zeolites had been allowed.

In the transformation of acetone at 400°C, the model shows very good agreement with experiment for most products. The yields of isobutene and acetone are noticeably underestimated, which corresponds to an overestimation of their conversion. The sum of squared error (SSE) for the transformation of acetone is 1.5×10^{-2}. In the transformation of acetic acid at 450°C, the model again shows very good agreement with experiment, although the yield of acetone is again underestimated in this case. The SSE for the transformation of acetic acid is 8.2×10^{-3}.

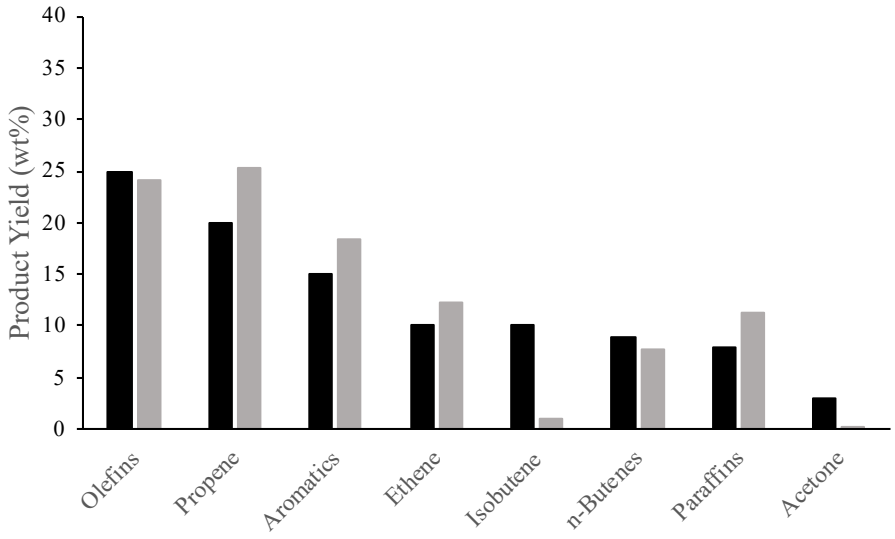


Figure 4.6. Experimental (black) vs. model (gray) product yields (wt%) for the transformation of acetone at 400°C. Note that the yields of CO_2 and H_2O were not included in this calculation.

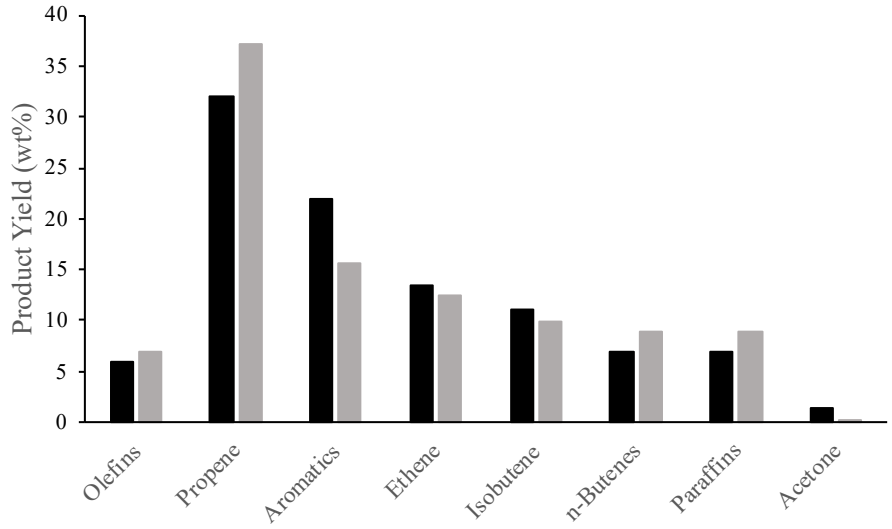


Figure 4.7. Experimental (black) vs. model (gray) product yields (wt%) for the transformation of acetic acid at 450°. Note that the yields of CO_2 and H_2O were not included in this calculation.

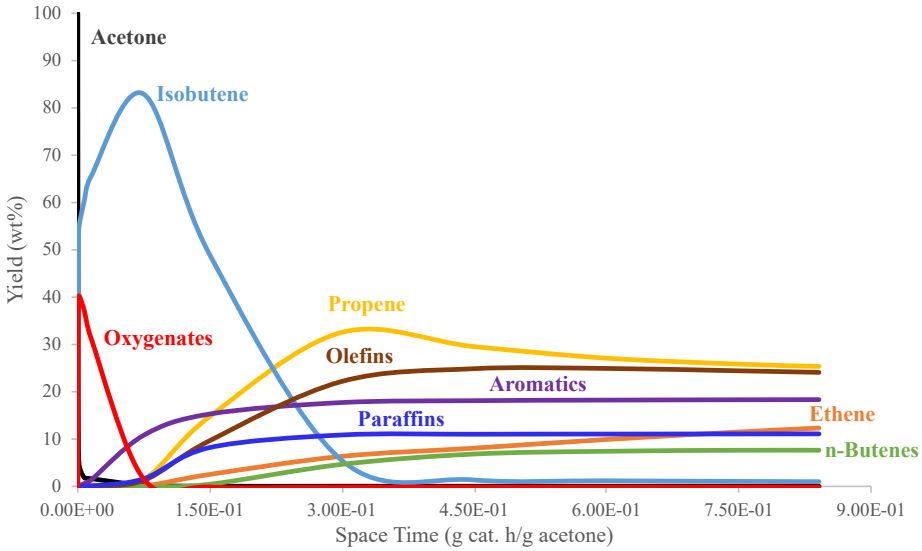


Figure 4.8. Progression with space time of acetone conversion at 400°C to various product classes. Note that the conversion of acetone to oxygenates and isobutene is so rapid that the decrease in acetone coincides with the y-axis.

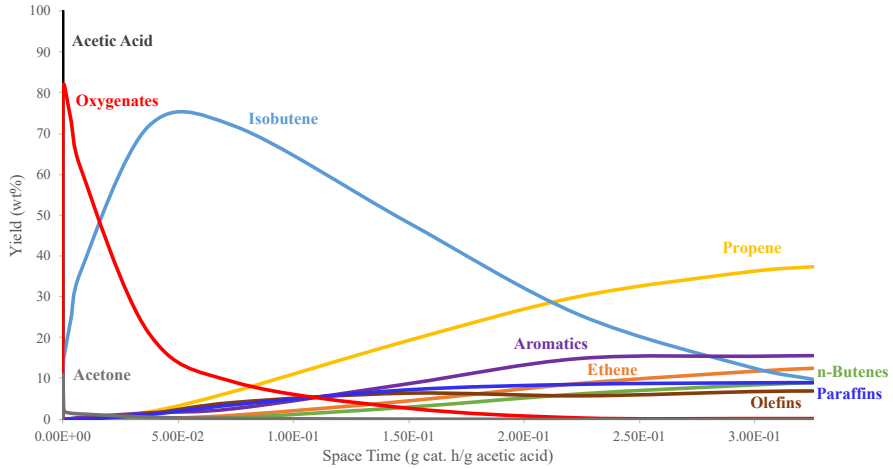


Figure 4.9. Progression with space time of acetic acid conversion at 450°C to various product classes. Note that the conversion of acetic acid to oxygenates is so rapid that the decrease in acetic acid and the increase in oxygenates and acetone coincide with the y-axis.

4.4.4. Reaction Mechanism and Net Rate Analysis

While the experimental data provides only the yields at the exit, the model tracks the yield profiles as a function of the length along the PFR, i.e., the number of sites. These predicted profiles are shown in Figures 4.8 and 4.9 for acetone and acetic acid conversion, respectively.

An analysis of the net rates provides vital information regarding the mechanism by which acetic acid and acetone are transformed into the different products and product classes and can be used to interpret the results as a function of space time. There are three notable "phases" in the reaction mechanism: conversion of acetic acid to acetone, aldol condensation of acetone, and formation of olefins and aromatics.

4.4.4.1. Conversion of Acetic Acid to Acetone

As proposed by Gayubo and coworkers[153], the first overall step in the transformation of acetic acid is the formation of acetone. Gayubo and coworkers proposed that the deoxygenation of acetic acid takes place by decarboxylation and dehydration given the overall stoichiometry of the reaction. Our model allows us to test this proposal, identify the key reaction intermediates, and uncover if there are other competitive routes that form acetone from acetic acid as well as divert mass to other primary products. Figure 4.10 shows the net rate analysis for the transformation of acetic acid to acetone. First, the hydroxyl group on acetic acid is protonated, followed by dehydration to form the acylium ion. The acylium ion adds to acetic acid, which is followed by alcohol deprotonation to form acetoacetic acid. Acetoacetic acid then undergoes decarboxylation to form the enol version of acetone, which tautomerizes to the keto version of acetone. In addition to adding to acetic acid, the acylium ion can deprotonate to form ketene.

However, the flux toward oligomerization of ketene is very small, as the ketene primarily desorbs, creating a build-up of ketene in the gas phase, which can be seen in Figure 4.9. Also, in addition to protonation of the hydroxyl group, acetic acid can undergo ketone protonation, starting a pathway that ultimately regenerates the acylium ion, in addition to forming acetone.

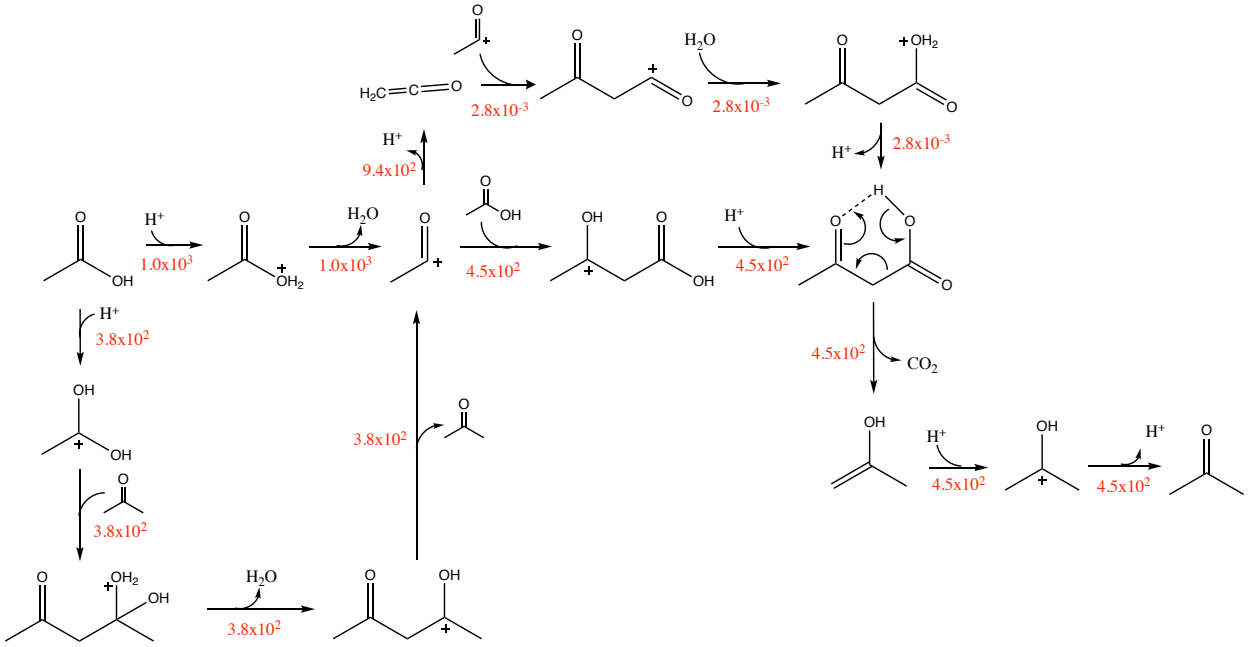


Figure 4.10. Net rate analysis for the transformation of acetic acid to acetone at 450°C and a space time of 7.46×10^{-7} g cat. h/g acetic acid. Net rates (red) are in s^{-1}. Note that all chemisorbed species are shown as carbenium/oxonium ions, but the primary/secondary species exist as alkoxides. All species are either physisorbed or chemisorbed (no gas-phase species shown). The arrows point to species with positive rate of formation.

4.4.4.2. Aldol Condensation of Acetone

As postulated by Salvapati and coworkers[199], acetone is transformed into subsequent products via catalytic self-condensation, but the product distribution is clearly complex such that there are many other subsequent and competing pathways. Even for the step of aldol condensation, there is debate about the dominant mechanistic route. Some literature reports contend that acetone aldol condensation occurs through the enol form of acetone[173,200], while our

net rate analysis in Figure 4.11 reveals that aldol condensation actually occurs primarily through the keto form of acetone, as suggested by Herrmann and Iglesia.[172] The keto-enol equilibrium of acetone clearly favors the keto form, leading to a much larger rate for acetone aldol condensation via the keto form than acetone aldol condensation via the enol form. The chemisorbed intermediate formed from acetone aldol condensation via the keto form can undergo either oxodeprotonation to form diacetone alcohol or dehydration to form another chemisorbed intermediate that serves as the starting point for the next phase of the reaction mechanism.

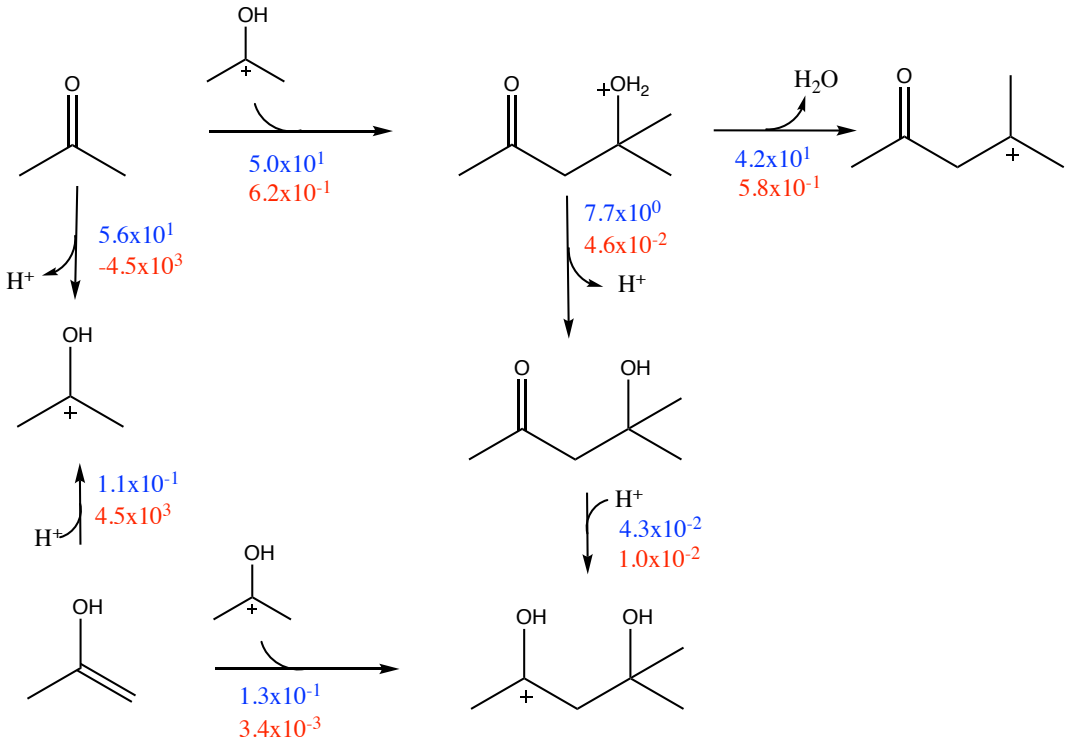


Figure 4.11. Net rate (in s^{-1}) analysis for the aldol condensation of acetone for a space time of 7.46×10^{-7} g cat. h/g reactant. The net rates in blue are from the transformation of acetone at 400°C, while the net rates in red are from the transformation of acetic acid at 450°C. Although they are presented in the same figure, no comparisons are made between the net rates for the transformation of acetone and acetic acid, as the space time corresponds to a different effective conversion of each reactant. Note that all chemisorbed species are shown as carbenium/oxonium ions, but the primary/secondary species exist as alkoxides. All species are either physisorbed or chemisorbed (no gas-phase species shown).

4.4.4.3. Formation of Olefins and Aromatics

The rightmost chemisorbed intermediate in Figure 4.11 serves as a key branching point in the mechanism. More specifically, this chemisorbed intermediate can undergo β-scission to form isobutene and regenerate an acylium ion, or it can undergo deprotonation to a physisorbed intermediate, as seen in Figure 4.12. Each of these pathways will be discussed in turn. We will demonstrate that xylene is produced from pathways stemming from isobutene, while mesitylene is produced from pathways stemming from the two physisorbed intermediates, one of which is mesityl oxide, in Figure 4.12.

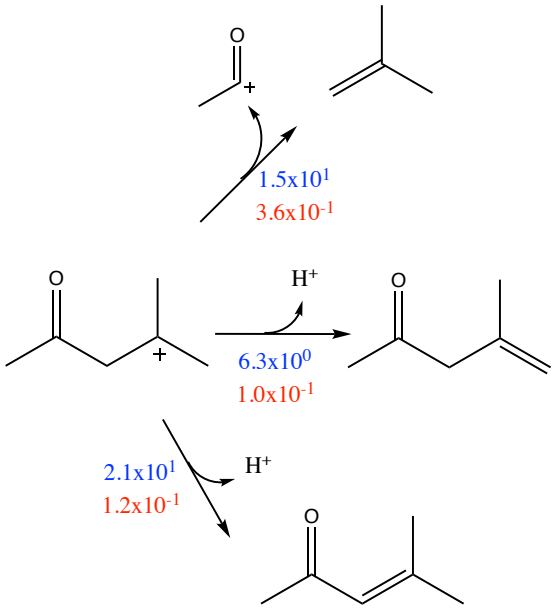


Figure 4.12. Net rate (in s^{-1}) analysis for a key branching point in the mechanism for a space time of 7.46×10^{-7} g cat. h/g reactant. The net rates in blue are from the transformation of acetone at 400°C, while the net rates in red are from the transformation of acetic acid at 450°C. Although they are presented in the same figure, no comparisons are made between the net rates for the transformation of acetone and acetic acid, as the space time corresponds to a different effective conversion of each reactant. All species are either physisorbed or chemisorbed (no gas-phase species shown).

A pathway from isobutene to various aromatics is shown in Figure 4.13. Note that this is not a comprehensive set of elementary steps leading to the xylene isomers seen on the right in the figure. Rather, Figure 4.13 is a compilation of elementary steps forming a single pathway to xylene and is shown as a representative example. The pathway begins with the oligomerization of isobutene. The net rates show that the resulting chemisorbed intermediate undergoes a hydride shift to form a different chemisorbed intermediate. These two chemisorbed intermediates follow identical pathways, though the location of a double bond varies. Rather than undergoing the deprotonation seen in Figure 4.13, the two chemisorbed intermediates favor isomerization pathways that ultimately produce the olefins seen in Appendix C. For example, one of the C_8H_{14} species in Figure S4 is produced from the pathway seen in Figure 4.14, which includes shifting and branching. Further, each route in Figure 4.13 requires two hydride transfer reactions, which according to the net rates, are clearly not favored. As a result, the fluxes toward the xylene isomers, and thus their yields, are insubstantial.

On the other hand, mesitylene is produced from pathways stemming from the two bottom physisorbed intermediates in Figure 4.12. As can be seen in Figure 4.15, aldol condensation and dehydration reactions form vital steps in the pathway from mesityl oxide to mesitylene. The net rates reflect that these aldol condensation steps proceed through the keto version of acetone, confirming our previous conclusion.

In summary, we have concluded that the yield of aromatics from the isobutene pathway is insubstantial. Rather, aromatics (i.e. mesitylene) stem from aldol condensation of mesityl oxide. However, it is possible that this conclusion is an artifact of the mechanism

construction. The oligomerization of isobutene forms a branched C8 olefinic ion. For cyclization of this ion to occur, there must be a double bond and positive charge located at the one and six positions, respectively. Given this requirement, the C8 species must be a primary ion. When constructing the mechanism, we allowed for hydride transfer to form primary ions to ensure that there were pathways to aromatics. However, in reality, the flux through these primary ionic intermediates is too small for any noticeable yield of aromatics to be formed from this pathway. Rather, we determine that the more facile pathway to aromatics is that in Figure 4.15, which does not require a primary intermediate. We note that it is possible that if we lift the carbon number limitation of 9, allowing for the formation of C12 species, there will be pathways from isobutene to aromatics that occur via secondary or tertiary chemisorbed intermediates which are much more favored than those through the primary chemisorbed intermediate.

The kinetic parameters used in this work were estimated and tuned based on the limited experimental data provided by Gayubo and coworkers.[153] However, if further experiments were performed, the robustness of the model could potentially be improved further and additional mechanistic details could be gleaned from the model output. More specifically, experiments studying the transformation of acetone or acetic acid at multiple temperatures and conversion levels is desired. This would allow for confirmation of the results as a function of space time to be validated further and confirm that the rapid conversion of acetic acid or acetone at the inlet is borne out. Additionally, this study was only able to simulate a fresh catalyst. Further experimentation quantifying a catalyst deactivation rate would allow mass yields to be estimated as a function of time on stream.

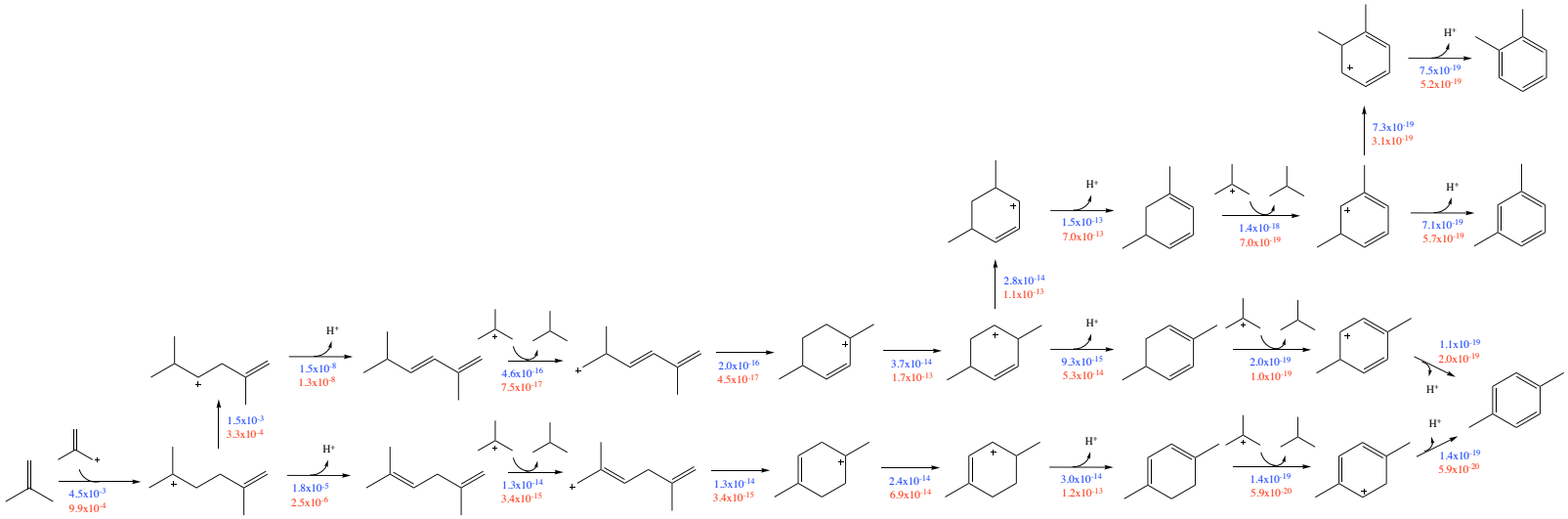


Figure 4.13. Net rate (in s^{-1}) analysis for the pathway from isobutene to aromatics for a space time of 7.46×10^{-5} g cat. h/g reactant. The net rates in blue are from the transformation of acetone at 400°C, while the net rates in red are from the transformation of acetic acid at 450°C. Although they are presented in the same figure, no comparisons are made between the net rates for the transformation of acetone and acetic acid, as the space time corresponds to a different effective conversion of each reactant. Note that all chemisorbed species are shown as carbenium/oxonium ions, but the primary/secondary species exist as alkoxides. All species are either physisorbed or chemisorbed (no gas-phase species shown).

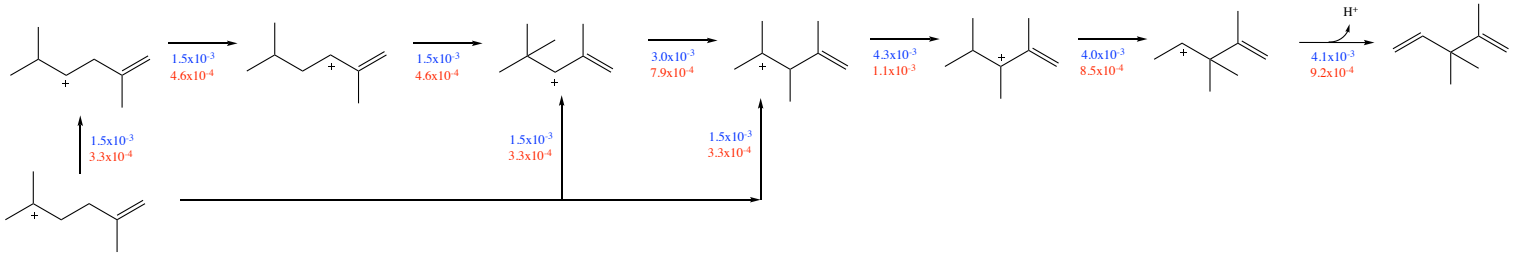


Figure 4.14. Net rate (in s^{-1}) analysis for example pathway to C8H14 for a space time of 7.46×10^{-5} g cat. h/g reactant. The net rates in blue are from the transformation of acetone at 400°C, while the net rates in red are from the transformation of acetic acid at 450°C. Although they are presented in the same figure, no comparisons are made between the net rates for the transformation of acetone and acetic acid, as the space time corresponds to a different effective conversion of each reactant. Note that all chemisorbed species are shown as carbenium/oxonium ions, but the primary/secondary species exist as alkoxides. All species are either physisorbed or chemisorbed (no gas-phase species shown).

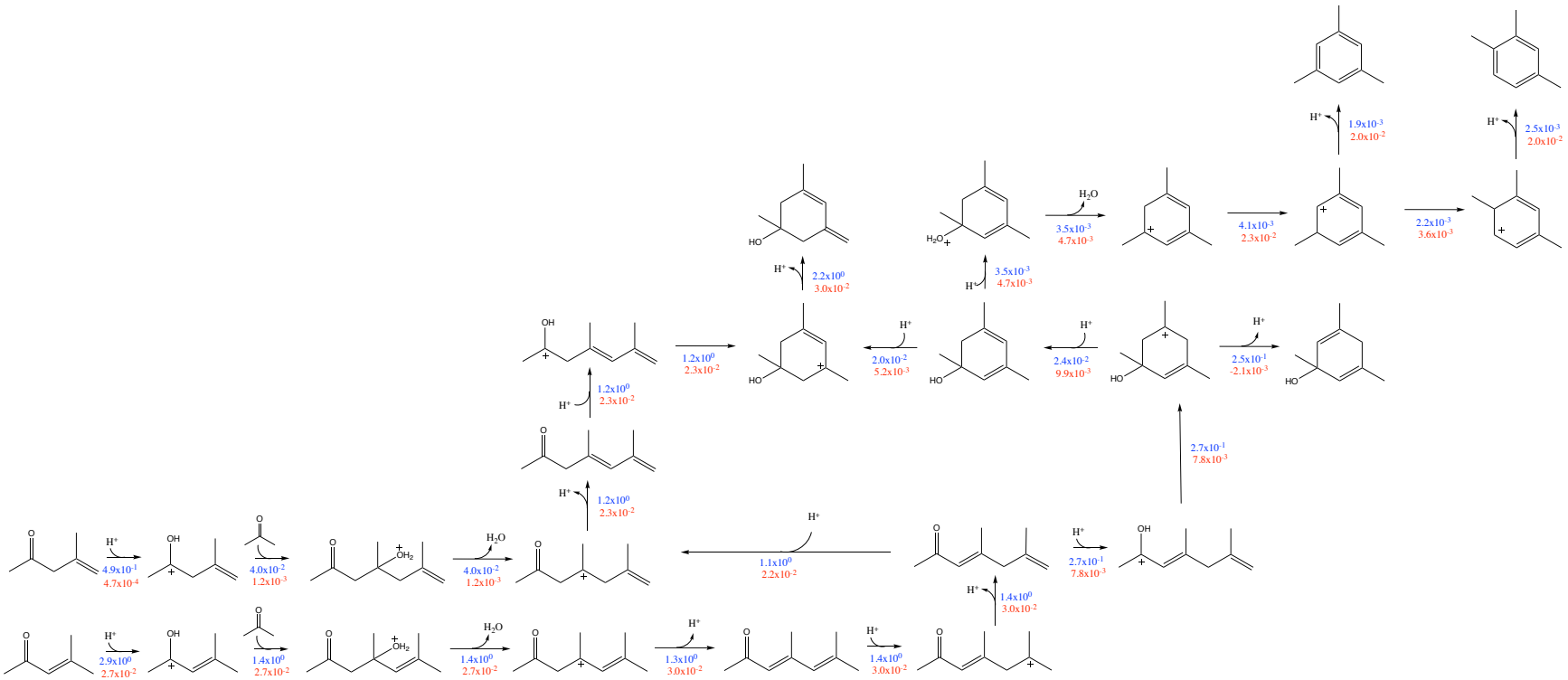


Figure 4.15. Net rate (in s^{-1}) analysis for formation of mesitylene for a space time of 7.46x10-5 g cat. h/g reactant. The net rates in blue are from the transformation of acetone at 400°C, while the net rates in red are from the transformation of acetic acid at 450°C. Although they are presented in the same figure, no comparisons are made between the net rates for the transformation of acetone and acetic acid, as the space time corresponds to a different effective conversion of each reactant. Note that all chemisorbed species are shown as carbenium/oxonium ions, but the primary/secondary species exist as alkoxides. All species are either physisorbed or chemisorbed (no gas-phase species shown).

4.5. Conclusions

In this work, a microkinetic model was developed to describe the transformations of acetic acid and acetone, two key oxygenates present in bio-oil, on HZSM-5. Reaction families involving hydrocarbon chemistry and chemistry involving oxygenates were elucidated. A mechanism reduction technique known as seeding was used to reduce the network to a manageable size. An automated network generator was used to build a kinetic network consisting of 580 unique species and 2,160 unique reactions. A variety of group contribution methods and DFT calculations was used to estimate enthalpies of reaction in the surface phase and entropy changes. There was very good agreement between model and experiment for both the transformation of acetic acid and the transformation of acetone. Net rate analysis quantified that the first overall step in the transformation of acetic acid is dehydration followed by decarboxylation to form acetone. Acetone is transformed through catalytic self-condensation of the keto form rather than the enol form as is often proposed in the literature. Finally, a key branching point was identified, separating the mechanism into a pathway characterized by isobutene oligomerization to heavier olefins and a pathway characterized by further aldol condensation to aromatics. While the current microkinetic model is a powerful tool to simulate the transformations of acetic acid and acetone, further experimentation would allow for increased robustness of the model, more mechanistic details, and the implementation of a catalyst deactivation rate.

Chapter 5: Summary and Future Perspectives

This dissertation has demonstrated the power of computational methods in understanding biomass pyrolysis. More specifically, two well-known challenges were addressed. First, we sought to tackle the uncertainty of the structure of lignin by using a kinetic Monte Carlo algorithm to generate representative structures of lignin. Second, we utilized a variety of techniques, including group additivity, reaction network generation and parameter estimation and optimization, to gain a better understanding of the catalytic upgrading of bio-oil.

Chapter 2 presented the extension of a stochastic method to produce libraries of diverse lignin molecules. This extension incorporates more complexity and allows for the structure generation of lignin from all types of biomass, including hardwood, softwood, and herbaceous biomass. Using this method, we were able to develop libraries of lignin molecules whose properties – monomer, bond, and molecular weight distributions and the branching coefficient – on average match the experimental data presented in the literature. Additionally, we determined incompatible characteristics and identified impossible lignin structures by exploring "lignin space", which includes all possible structures of lignin given the four structural characteristics.

It is possible, however, that the incompatible characteristics are a result of incomplete experimental data. For instance, the monomer and bond distributions may have been obtained from different literature sources. Additionally, the branching coefficient is estimated from a correlation with molecular weight. It would be beneficial to conduct experiments that can directly measure the branching coefficient of a lignin molecule. In addition, the inconsistencies in the data highlight the need for self-consistent, reliable experiments that provide all four structural characteristics.

Naturally, these lignin libraries can be utilized in molecular simulations. Vermaas and coworkers developed LigninBuilder, which is a framework for transforming lignin libraries into three-dimensional atomic models.[201] The subsequent molecular dynamics simulations of these models provided insight into atomic-level lignin interactions. Additionally, a kinetic model has been constructed with the aim of predicting the product distribution from lignin fast pyrolysis.[202] More specifically, important reaction families were derived from model compound studies and applied to a library of lignin molecules. However, the bond types of interest were β-O-4, β-5 and 5-5, whereas the structure generation program developed in this work includes seven bond types. Thus, in future work, additional reaction families can be created that describe transformations involving the α-O-4, β-β, 4-O-5 and β-1 linkages. With these additional reaction families, the kinetic models can be used to simulate the fast pyrolysis of lignin from any biomass source.

Chapters 3 and 4 moved into the second aim of this dissertation, namely, to gain an understanding of the catalytic upgrading of bio-oil. A detailed microkinetic model was developed to describe the transformations of acetic acid and acetone, two vapor products of biomass pyrolysis, to valuable fuels and chemicals. A great deal of Chapter 4 was devoted to explaining the methodology behind this microkinetic model. In particular, care was taken to describe how the activation energy and frequency factor, two vital parameters, were estimated. While the frequency factor was directly estimated from transition-state theory, the activation energy was calculated from the Evans-Polanyi equation. A key thermodynamic quantity in this equation is the enthalpy of reaction in the surface phase, which is calculated from the enthalpy of reaction in the gas phase, stabilization energies of the chemisorbed species, and physisorption enthalpies of the neutral species. The stabilization energies and physisorption enthalpies were estimated from density

functional theory calculations, while the enthalpy of reaction in the gas phase is simply a combination of the enthalpies of formation of each species in the reaction. Although it is likely that these enthalpies of formation are available in the literature for the neutral species, the same cannot be said of the ionic species. Therefore, Chapter 3 discusses a group additivity method for calculating enthalpies of formation and a multiple linear regression study to determine 71 group additivity values for oxygenates, oxonium ions, and oxygen-containing carbenium ions.

In addition to discussing the thermodynamic and kinetic parameters, Chapter 4 also presented the reaction families that were applied to the two reactants, acetic acid and acetone, and their progeny. An automated network generator was used to construct the reaction network and the reactor design ordinary differential equations. In general, there was good agreement between the model and experimental product yields for both the transformations of acetic acid at 450°C and acetone at 400°C. Additionally, an analysis of the net rates led to three conclusions. First, the first phase of the transformation of acetic acid is the dehydration and decarboxylation to acetone. Second, acetone undergoes aldol condensation through its keto form rather than its enol form, as suggested in the literature. Finally, heavy olefins were formed from the oligomerization of isobutene, while aromatics were formed from the continued aldol condensation of mesityl oxide.

The microkinetic model developed in this work has many other applications. For instance, the model was intended for the upgrading of acetic acid and acetone, but there are many other model compounds in bio-oil. While acetic acid and acetone are derived from hemicellulose and cellulose, compounds such as guaiacol are derived from lignin. Additional reaction families may be required to develop a model for the upgrading of guaiacol, but the methodology is largely the same. Further, mixtures of model compounds can also be explored to better represent bio-oil. In

addition to testing different model compounds at different operating conditions, different catalysts, such as H-FAU or H-BEA, can be used to determine the effect of acidity and pore size on the product yields.

Coking and deactivation are huge concerns when working with zeolites. It would be beneficial if experimentation were to provide a deactivation rate of the catalyst that can be incorporated into the model. This would allow us to compare the progress with time on stream from the model with that from the experiments, instead of simply matching the data that represents a fresh catalyst. Additionally, metals can be incorporated into the aluminosilicate framework. For example, [Ga]-H-ZSM5 has been found to increase bio-oil yield and decrease gas yield.

Ultimately, while there are several ways to refine and expand the microkinetic model developed in this work, it shows promise as a basic framework for studying zeolitic upgrading of bio-oil.

References

1. Xiu, S.N.; Shahbazi, A. *Renew. Sust. Energ. Rev.* **2012,** *16*(7), 4406-4414.
2. Mortensen, P.M.; Grunwaldt, J.D.; Jensen, P.A.; Knudsen, K.G.; Jensen, A.D. *Appl. Catal. A-Gen.* **2011,** *407*(1-2), 1-19.
3. Balat, M. *Energy Convers. Manage.* **2011,** *52*(2), 858-875.
4. Limayem, A.; Ricke, S.C. *Prog. Energ. Combust.* **2012,** *38*(4), 449-467.
5. Schneider, U.A.; Mccarl, B.A. *Environ Resour Econ* **2003,** *24*(4), 291-312.
6. Wang, H.M.; Male, J.; Wang, Y. *ACS Catal.* **2013,** *3*(5), 1047-1070.
7. Ruddy, D.A.; Schaidle, J.A.; Ferrell, J.R.; Wang, J.; Moens, L.; Hensley, J.E. *Green Chem.* **2014,** *16*(2), 454-490.
8. Sutton, D.; Kelleher, B.; Ross, J.R.H. *Fuel Process Technol.* **2001,** *73*(3), 155-173.
9. Bridgwater, A.V. *Fuel* **1995,** *74*(5), 631-653.
10. Mckendry, P. *Bioresource Technol.* **2002,** *83*(1), 55-63.
11. Sun, Y.; Cheng, J.Y. *Bioresource Technol.* **2002,** *83*(1), 1-11.
12. Chum, H.L.; Johnson, D.K.; Black, S.; Baker, J.; Grohmann, K.; Sarkanen, K.V.; Wallace, K.; Schroeder, H.A. *Biotechnol. Bioeng.* **1988,** *31*(7), 643-649.
13. Lu, Q.; Li, W.Z.; Zhu, X.F. *Energy Convers. Manage.* **2009,** *50*(5), 1376-1383.
14. Bridgwater, A.V. *Biomass Bioenerg.* **2012,** *38*, 68-94.
15. Dickerson, T.; Soria, J. *Energies* **2013,** *6*(1), 514-538.
16. French, R.; Czernik, S. *Fuel Process Technol.* **2010,** *91*(1), 25-32.
17. Patwardhan, P.R.; Satrio, J.A.; Brown, R.C.; Shanks, B.H. *J. Anal. Appl. Pyrol.* **2009,** *86*(2), 323-330.
18. Vinu, R.; Broadbelt, L.J. *Energy Environ. Sci.* **2012,** *5*(12), 9808-9826.
19. Zhou, X.W.; Nolte, M.W.; Mayes, H.B.; Shanks, B.H.; Broadbelt, L.J. *Ind. Eng. Chem. Res.* **2014,** *53*(34), 13274-13289.
20. Zhou, X.W.; Nolte, M.W.; Shanks, B.H.; Broadbelt, L.J. *Ind. Eng. Chem. Res.* **2014,** *53*(34), 13290-13301.
21. Williams, P.T.; Nugranad, N. *Energy* **2000,** *25*(6), 493-513.
22. Onay, O.; Kockar, O.M. *Fuel* **2006,** *85*(12-13), 1921-1928.
23. Mukarakate, C.; Zhang, X.D.; Stanton, A.R.; Robichaud, D.J.; Ciesielski, P.N.; Malhotra, K.; Donohoe, B.S.; Gjersing, E.; Evans, R.J.; Heroux, D.S.; Richards, R.; Iisa, K.; Nimlos, M.R. *Green Chem.* **2014,** *16*(3), 1444-1461.
24. Adler, E. *Wood Sci. Technol.* **1977,** *11*(3), 169-218.
25. Sakakibara, A. *Wood Sci. Technol.* **1980,** *14*(2), 89-100.
26. Chakar, F.S.; Ragauskas, A.J. *Ind. Crop. Prod.* **2004,** *20*(2), 131-141.
27. Dodd, A.P.; Kadla, J.F.; Straus, S.K. *ACS Sustain. Chem. Eng.* **2015,** *3*(1), 103-110.
28. Dolgonosov, B.M.; Gubernatorova, T.N. *Water Resources* **2010,** *37*(3), 320-331.
29. Sette, M.; Wechselberger, R.; Crestini, C. *Chem-Eur. J.* **2011,** *17*(34), 9529-9535.
30. Wan, S.L.; Wang, Y. *Front. Chem. Sci. Eng.* **2014,** *8*(3), 280-294.
31. Choi, S.J.; Park, S.H.; Jeon, J.K.; Lee, I.G.; Ryu, C.; Suh, D.J.; Park, Y.K. *Renew. Energ.* **2013,** *54*, 105-110.
32. Park, H.J.; Jeon, J.K.; Suh, D.J.; Suh, Y.W.; Heo, H.S.; Park, Y.K. *Catal. Surv. Asia* **2011,** *15*(3), 161-180.

33. Mihalcik, D.J.; Mullen, C.A.; Boateng, A.A. *J. Anal. Appl. Pyrol.* **2011,** *92*(1), 224-232.
34. Chiang, H.; Bhan, A. *J. Catal.* **2010,** *271*(2), 251-261.
35. Dejaifve, P.; Vedrine, J.C.; Bolis, V.; Derouane, E.G. *J. Catal.* **1980,** *63*(2), 331-345.
36. Hallac, B.B.; Ragauskas, A.J. *Biofuel. Bioprod. Bior.* **2011,** *5*(2), 215-225.
37. Mccoy, B.J.; Madras, G. *Chem. Eng. Sci.* **2001,** *56*(8), 2831-2836.
38. Zhou, X.W.; Li, W.; Mabon, R.; Broadbelt, L.J. *Energy Technol.* **2016,** *4*, 1-29.
39. Ebringerova, A. *Macromol. Symp.* **2006,** *232*, 1-12.
40. Scheller, H.V.; Ulvskov, P. *Annu. Rev. Plant Biol.* **2010,** *61*, 263-289.
41. Ranzi, E.; Cuoci, A.; Faravelli, T.; Frassoldati, A.; Migliavacca, G.; Pierucci, S.; Sommariva, S. *Energy Fuel* **2008,** *22*(6), 4292-4300.
42. Horton, S.R.; Mohr, R.J.; Zhang, Y.; Petrocelli, F.P.; Klein, M.T. *Energy Fuel* **2016,** *30*(3), 1647-1661.
43. Zhou, X.W.; Li, W.J.; Mabon, R.; Broadbelt, L.J. *Energy Technol.* **2017,** *5*(1), 52-79.
44. Zakzeski, J.; Jongerius, A.L.; Bruijnincx, P.C.A.; Weckhuysen, B.M. *ChemSusChem* **2012,** *5*(8), 1602-1609.
45. Stewart, D. *Industrial Crops and Products* **2008,** *27*(2), 202-207.
46. Freudenberg, K. *Science* **1965,** *148*(3670), 595-600.
47. Glasser, W.G.; Glasser, H.R.; Morohoshi, N. *Macromolecules* **1981,** *14*(2), 253-262.
48. Train, P.M.; Klein, M.T. *ACS Sym. Ser.* **1988,** *376*, 241-263.
49. Yanez, A.J.; Li, W.; Mabon, R.; Broadbelt, L.J. *Energy Fuels* **2016,** *30*(7), 5835-5845.
50. Broadbelt, L.J.; Stark, S.M.; Klein, M.T. *Ind. Eng. Chem. Res.* **1994,** *33*(4), 790-799.
51. Zakzeski, J.; Bruijnincx, P.C.A.; Jongerius, A.L.; Weckhuysen, B.M. *Chem. Rev.* **2010,** *110*(6), 3552-3599.
52. N.-S. Hon, D.; Shiraishi, N.; *Chemistry of Lignin*, 2nd; New York, 2001; 109-173.
53. Heitner, C.; Dimmel, D.R.; Schmidt, J.a.E. *Lignin and Lignans: Advances in Chemistry*, Boca Raton, FL, 2010.
54. Vanholme, R.; Demedts, B.; Morreel, K.; Ralph, J.; Boerjan, W. *Plant Physiol.* **2010,** *153*(3), 895-905.
55. Bjorkman, A. *Svensk papperstidning* **1956,** *59*(13), 477-485.
56. Guerra, A.; Filpponen, I.; Lucia, L.A.; Argyropoulos, D.S. *J. Agr. Food Chem.* **2006,** *54*(26), 9696-9705.
57. Chang, H.M.; Cowling, E.B.; Brown, W.; Adler, E.; Miksche, G. *Holzforschung* **1975,** *29*(5), 153-159.
58. Erickson, M.; Miksche, G.E.; Somfai, I. *Holzforschung* **1973,** *27*(4), 113-117.
59. Erickson, M.; Miksche, G.E.; Somfai, I. *Holzforschung* **1973,** *27*(5), 147-150.
60. Erickson, M.; Larsson, S.; Miksche, G.E. *Acta Chem. Scand.* **1973,** *27*(1), 127-140.
61. Erickson, M.; Larsson, S.; Miksche, G.E. *Acta Chem. Scand.* **1973,** *27*(3), 903-914.
62. Lundquist, K. *Acta Chem. Scand.* **1970,** *24*(3), 889-907.
63. Capanema, E.A.; Balakshin, M.Y.; Kadla, J.F. *J. Agr. Food Chem.* **2004,** *52*(7), 1850-1860.
64. Nimz, H.H.; Tschirner, U.; Stahle, M.; Lehmann, R.; Schlosser, M. *J. Wood Chem. Technol.* **1984,** *4*(3), 265-284.
65. Lundquist, K. *Acta Chem. Scand. B* **1980,** *34*(1), 21-26.
66. Freudenberg, K. *Science* **1965,** *148*, 595-600.
67. Adler, E. *Ind. Eng. Chem.* **1957,** *49*(9), 1377-1383.

68. Del Rio, J.C.; Prinsen, P.; Rencoret, J.; Nieto, L.; Jimenez-Barbero, J.; Ralph, J.; Martinez, A.T.; Gutierrez, A. *J. Agr. Food Chem.* **2012,** *60*(14), 3619-3634.
69. Del Rio, J.C.; Rencoret, J.; Prinsen, P.; Martinez, A.T.; Ralph, J.; Gutierrez, A. *J. Agr. Food Chem.* **2012,** *60*(23), 5922-5935.
70. Ralph, J. *Phytochem Rev.* **2010,** *9*(1), 65-83.
71. Ralph, J.; Hatfield, R.D.; Quideau, S.; Helm, R.F.; Grabber, J.H.; Jung, H.J.G. *J. Am. Chem. Soc.* **1994,** *116*(21), 9448-9456.
72. Nimz, H. *Angew. Chem. Int. Edit.* **1974,** *13*(5), 313-321.
73. Ralph, J. *US Dairy Forage Research Center, USDA-Agricultural Research Service* **1999**.
74. Argyropoulos, D.S.; Jurasek, L.; Kristofova, L.; Xia, Z.C.; Sun, Y.J.; Palus, E. *J. Agr. Food Chem.* **2002,** *50*(4), 658-666.
75. Karhunen, P.; Rummakko, P.; Sipila, J.; Brunow, G.; Kilpelainen, I. *Tetrahedron Lett.* **1995,** *36*(1), 169-170.
76. Ralph, J.; Lundquist, K.; Brunow, G.; Lu, F.; Kim, H.; Schatz, P.F.; Marita, J.M.; Hatfield, R.D.; Ralph, S.A.; Christensen, J.H.; Boerjan, W. *Phytochem Rev.* **2004,** *3*, 29-60.
77. Zhang, L.M.; Gellerstedt, G.; Ralph, J.; Lu, F.C. *J. Wood Chem. Technol.* **2006,** *26*(1), 65-79.
78. Dolk, M.; Pla, F.; Yan, J.F.; Mccarthy, J.L. *Macromolecules* **1986,** *19*(5), 1464-1470.
79. Pla, F.; Dolk, M.; Yan, J.F.; Mccarthy, J.L. *Macromolecules* **1986,** *19*(5), 1471-1477.
80. Ek, M.; Gellerstedt, G.; Henriksson, G.; *Lignin*, 2009; 1, 121-145.
81. Holmgren, A. Biochemical Control Aspects in Lignin Polymerization. Stockholm: Royal Institute of Technology; 2008.
82. Aimi, H.; Matsumoto, Y.; Meshitsuka, G. *J. Wood Sci.* **2005,** *51*(3), 252-255.
83. Somerville, C.; Youngs, H.; Taylor, C.; Davis, S.C.; Long, S.P. *Science* **2010,** *329*(5993), 790-792.
84. Bauer, S.; Sorek, H.; Mitchell, V.D.; Ibanez, A.B.; Wemmer, D.E. *J. Agr. Food Chem.* **2012,** *60*(33), 8203-8212.
85. Adler, E.; Miksche, G.E.; Johansson, B. *Holzforschung* **1968,** *22*(6), 171-174.
86. Adler, E.; Lundquist, K. *Acta Chem. Scand.* **1963,** *17*(1), 13-26.
87. Guerra, A.; Filpponen, I.; Lucia, L.A.; Saquing, C.; Baumberger, S.; Argyropoulos, D.S. *J. Agr. Food Chem.* **2006,** *54*(16), 5939-5947.
88. Miksche, G.E.; Yasuda, S. *Liebigs Ann Chem.* **1976,** (7-8), 1323-1332.
89. Brosse, N.; El Hage, R.; Chaouch, M.; Petrissans, M.; Dumarcay, S.; Gerardin, P. *Polym. Degrad. Stabil.* **2010,** *95*(9), 1721-1726.
90. Gellerstedt, G.; Pranda, J.; Lindfors, E.L. *J. Wood Chem. Technol.* **1994,** *14*(4), 467-482.
91. Larsson, S.; Miksche, G.E. *Acta Chem. Scand.* **1971,** *25*(2), 647-662.
92. Lundquist, K. *Acta Chem. Scand.* **1973,** *27*(7), 2597-2606.
93. Adjaye, J.D.; Bakhshi, N.N. *Biomass Bioenerg.* **1995,** *8*(3), 131-149.
94. Laidler, K.J.; King, M.C. *J. Phys. Chem.* **1983,** *87*(15), 2657-2664.
95. Evans, M.G.; Polanyi, M. *T. Faraday Soc.* **1938,** *34*(1), 11-24.
96. Lay, T.H.; Bozzelli, J.W. *J. Phys. Chem. A* **1997,** *101*(49), 9505-9510.
97. Marsi, I.; Viskolcz, B.; Seres, L. *J. Phys. Chem. A* **2000,** *104*(19), 4497-4504.
98. Sabbe, M.K.; De Vleeschouwer, F.; Reyniers, M.F.; Waroquier, M.; Marin, G.B. *J. Phys. Chem. A* **2008,** *112*(47), 12235-12251.

99. Sabbe, M.K.; Saeys, M.; Reyniers, M.F.; Marin, G.B.; Van Speybroeck, V.; Waroquier, M. *J. Phys. Chem. A* **2005,** *109*(33), 7466-7480.
100. Sebbar, N.; Bozzelli, J.W.; Bockhorn, H. *J. Phys. Chem. A* **2004,** *108*(40), 8353-8366.
101. Sumathi, R.; Green, W.H. *J. Phys. Chem. A* **2002,** *106*(46), 11141-11149.
102. Sumathi, R.; Green, W.H. *J. Phys. Chem. A* **2002,** *106*(34), 7937-7949.
103. Swihart, M.T.; Girshick, S.L. *J. Phys. Chem. B* **1999,** *103*(1), 64-76.
104. Wong, H.W.; Nieto, J.C.A.; Swihart, M.T.; Broadbelt, L.J. *J. Phys. Chem. A* **2004,** *108*(5), 874-897.
105. Benson, S.W.; Buss, J.H. *J. Chem. Phys.* **1958,** *29*(3), 546-572.
106. Benson, S.W.; Cruickshank, F.R.; Golden, D.M.; Haugen, G.R.; Oneal, H.E.; Rodgers, A.S.; Shaw, R.; Walsh, R. *Chem. Rev.* **1969,** *69*(3), 279-324.
107. Benson, S.W. *Thermochemical Kinetics: Methods for the Estimation of Thermochemical Data and Rate Parameters*, 2nd Ed.; New York, 1976.
108. Cohen, N. *J. Phys. Chem. Ref. Data* **1996,** *25*(6), 1411-1481.
109. Khan, S.S.; Yu, X.R.; Wade, J.R.; Malmgren, R.D.; Broadbelt, L.J. *J. Phys. Chem. A* **2009,** *113*(17), 5176-5194.
110. Paraskevas, P.D.; Sabbe, M.K.; Reyniers, M.F.; Papayannakos, N.; Marin, G.B. *Chem-Eur. J.* **2013,** *19*(48), 16431-16452.
111. Eigenmann, H.K.; Golden, D.M.; Benson, S.W. *J. Phys. Chem.* **1973,** *77*(13), 1687-1691.
112. Olah, G.A. *Onium Ions*, New York, 1998.
113. Prakash, G.K.S.; Bae, C.S.; Rasul, G.; Olah, G.A. *J. Org. Chem.* **2002,** *67*(4), 1297-1301.
114. Curtiss, L.A.; Redfern, P.C.; Raghavachari, K. *J. Chem. Phys.* **2007,** *126*(8).
115. Curtiss, L.A.; Redfern, P.C.; Raghavachari, K.; Pople, J.A. *J. Chem. Phys.* **2001,** *114*(1), 108-117.
116. Lias, S.G.; Bartmess, J.E.; Liebman, J.F.; Holmes, J.L.; Levin, R.D.; Mallard, W.G. *J. Phys. Chem. Ref. Data* **1988,** *17*, 1-861.
117. Hunter, E.P.L.; Lias, S.G. *J. Phys. Chem. Ref. Data* **1998,** *27*(3), 413-656.
118. Ponomarev, D.A.; Takhistov, V.V. *J. Chem. Educ.* **1997,** *74*(2), 201-203.
119. Colson, A.O.; Becker, D.; Eliezer, I.; Sevilla, M.D. *J. Phys. Chem. A* **1997,** *101*(47), 8935-8941.
120. Chen, C.J.; Bozzelli, J.W. *J. Phys. Chem. A* **2000,** *104*(21), 4997-5012.
121. Viskolcz, B.; Berces, T. *Phys. Chem. Chem. Phys.* **2000,** *2*(23), 5430-5436.
122. Sun, H.Y.; Bozzelli, J.W. *J. Phys. Chem. A* **2002,** *106*(15), 3947-3956.
123. Bjorkman, K.R.; Sung, C.Y.; Mondor, E.; Cheng, J.C.; Jan, D.Y.; Broadbelt, L.J. *Ind. Eng. Chem. Res.* **2014,** *53*(50), 19446-19452.
124. Traeger, J.C.; Mcloughlin, R.G. *J. Am. Chem. Soc.* **1981,** *103*(13), 3647-3652.
125. Traeger, J.C. *Int. J. Mass Spectrom.* **1984,** *58*, 259-271.
126. Traeger, J.C. *Rapid Commun. Mass Sp.* **1996,** *10*(1), 119-122.
127. Pittam, D.A.; Pilcher, G. *J. Chem. Soc. Farad. T. 1* **1972,** *68*(12), 2224-2229.
128. Manion, J.A. *J. Phys. Chem. Ref. Data* **2002,** *31*(1), 123-172.
129. Furuyama, S.; Golden, D.M.; Benson, S.W. *J. Chem. Thermodyn.* **1969,** *1*(4), 363-375.
130. Prosen, E.J.; Maron, F.W.; Rossini, F.D. *J. Res. Nat. Bur. Stand.* **1951,** *46*(2), 106-112.
131. Cox, J.D.; Pilcher, G. *Thermochemistry of Organic and Organometallic Compounds*, London, 1970.

132. Hine, J.; Arata, K. *B. Chem. Soc. Jpn.* **1976,** *49*(11), 3089-3092.
133. Chao, J.; Rossini, F.D. *J. Chem. Eng. Data* **1965,** *10*(4), 374-379.
134. Pilcher, G.; Coleman, D.J.; Pell, A.S. *T. Faraday Soc.* **1964,** *60*(4953), 499-505.
135. Da Silva, G.; Bozzelli, J.W.; Sebbar, N.; Bockhorn, H. *ChemPhysChem* **2006,** *7*(5), 1119-1126.
136. Wiberg, K.B.; Hao, S. *J. Org. Chem.* **1991,** *56*(17), 5108-5110.
137. Turecek, F.; Havlas, Z. *J. Org. Chem.* **1986,** *51*(21), 4066-4067.
138. Pilcher, G.; Pell, A.S.; Skinner, H.A.; Pope, A.E. *T. Faraday Soc.* **1963,** *59*(482), 316-330.
139. Guthrie, J.P. *Can. J. Chem.* **1978,** *56*(7), 962-973.
140. Fletcher, R.A.; Pilcher, G. *T. Faraday Soc.* **1970,** *66*(568), 794-799.
141. Guthrie, J.P. *J. Am. Chem. Soc.* **1974,** *96*(11), 3608-3615.
142. Chase, M.W. *NIST-JANAF Thermochemical Tables*, 4th Ed.; 1998.
143. Guthrie, J.P.; Liu, Z. *Can. J. Chem.* **1995,** *73*(9), 1395-1398.
144. Da Silva, G.; Bozzelli, J.W. *J. Phys. Chem. A* **2006,** *110*(48), 13058-13067.
145. Lay, T.H.; Bozzelli, J.W.; Dean, A.M.; Ritter, E.R. *J. Phys. Chem.* **1995,** *99*(39), 14514-14527.
146. Sumathi, R.; Green, W.H. *Phys. Chem. Chem. Phys.* **2003,** *5*(16), 3402-3417.
147. Carlson, T.R.; Vispute, T.R.; Huber, G.W. *ChemSusChem* **2008,** *1*(5), 397-400.
148. Liu, C.J.; Wang, H.M.; Karim, A.M.; Sun, J.M.; Wang, Y. *Chem. Soc. Rev.* **2014,** *43*(22), 7594-7623.
149. Naqvi, S.R.; Uemura, Y.; Yusup, S.; Sugiur, Y.; Nishiyama, N.; Naqvi, M. *Energ. Proced.* **2015,** *75*, 793-800.
150. Derouane, E.G.; Nagy, J.B.; Dejaifve, P.; Vanhooff, J.H.C.; Spekman, B.P.; Vedrine, J.C.; Naccache, C. *J. Catal.* **1978,** *53*(1), 40-55.
151. Oasmaa, A.; Czernik, S. *Energy Fuel* **1999,** *13*(4), 914-921.
152. Gayubo, A.G.; Aguayo, A.T.; Atutxa, A.; Aguado, R.; Bilbao, J. *Ind. Eng. Chem. Res.* **2004,** *43*(11), 2610-2618.
153. Gayubo, A.G.; Aguayo, A.T.; Atutxa, A.; Aguado, R.; Olazar, M.; Bilbao, J. *Ind. Eng. Chem. Res.* **2004,** *43*(11), 2619-2626.
154. Gayubo, A.G.; Tarrio, A.M.; Aguayo, A.T.; Olazar, M.; Bilbao, J. *Ind. Eng. Chem. Res.* **2001,** *40*(16), 3467-3474.
155. Engtrakul, C.; Mukarakate, C.; Starace, A.K.; Magrini, K.A.; Rogers, A.K.; Yung, M.M. *Catal. Today* **2016,** *269*, 175-181.
156. Bhan, A.; Joshi, Y.V.; Delgass, W.N.; Thomson, K.T. *J. Phys. Chem. B* **2003,** *107*(38), 10476-10487.
157. Boronat, M.; Corma, A. *Appl. Catal. A-Gen.* **2008,** *336*(1-2), 2-10.
158. Boronat, M.; Viruela, P.M.; Corma, A. *J. Am. Chem. Soc.* **2004,** *126*(10), 3300-3309.
159. Tuma, C.; Kerber, T.; Sauer, J. *Angew. Chem. Int. Edit.* **2010,** *49*(27), 4678-4680.
160. Tuma, C.; Sauer, J. *Angew. Chem. Int. Edit.* **2005,** *44*(30), 4769-4771.
161. Cnudde, P.; De Wispelaere, K.; Van Der Mynsbrugge, J.; Waroquier, M.; Van Speybroeck, V. *J. Catal.* **2017,** *345*, 53-69.
162. Nguyen, C.M.; De Moor, B.A.; Reyniers, M.F.; Marin, G.B. *J. Phys. Chem. C* **2012,** *116*(34), 18236-18249.
163. Kazansky, V.B.; Senchenya, I.N. *J. Catal.* **1989,** *119*(1), 108-120.

164. Rigby, A.M.; Kramer, G.J.; Vansanten, R.A. *J. Catal.* **1997,** *170*(1), 1-10.
165. Fang, H.J.; Zheng, A.M.; Xu, J.; Li, S.H.; Chu, Y.Y.; Chen, L.; Deng, F. *J. Phys. Chem. C* **2011,** *115*(15), 7429-7439.
166. Nieminen, V.; Sierka, M.; Murzin, D.Y.; Sauer, J. *J. Catal.* **2005,** *231*(2), 393-404.
167. Feng, W.; Vynckier, E.; Froment, G.F. *Ind. Eng. Chem. Res.* **1993,** *32*(12), 2997-3005.
168. Watson, B.A.; Klein, M.T.; Harding, R.H. *Ind. Eng. Chem. Res.* **1996,** *35*(5), 1506-1516.
169. Gates, B.C.; Katzer, J.R.; Schuit, G.C.A. *Chemistry of Catalytic Processes*, New York, 1979.
170. Gumidyala, A.; Sooknoi, T.; Crossley, S. *J. Catal.* **2016,** *340*, 76-84.
171. Pham, T.N.; Sooknoi, T.; Crossley, S.P.; Resasco, D.E. *ACS Catal.* **2013,** *3*(11), 2456-2473.
172. Herrmann, S.; Iglesia, E. *J. Catal.* **2017,** *346*, 134-153.
173. Diaz, C.D.C.; Locatelli, S.; Gonzo, E.E. *Zeolites* **1992,** *12*(7), 851-857.
174. Wang, S.; Iglesia, E. *ACS Catal.* **2016,** *6*(11), 7664-7684.
175. Wang, S.; Iglesia, E. *J. Catal.* **2017,** *352*, 415-435.
176. Broadbelt, L.J.; Stark, S.M.; Klein, M.T. *Chem. Eng. Sci.* **1994,** *49*(24b), 4991-5010.
177. Broadbelt, L.J.; Stark, S.M.; Klein, M.T. *Ind. Eng. Chem. Res.* **1995,** *34*(8), 2566-2573.
178. Broadbelt, L.J.; Stark, S.M.; Klein, M.T. *Comput. Chem. Eng.* **1996,** *20*(2), 113-129.
179. Ugi, I.; Bauer, J.; Brandt, J.; Friedrich, J.; Gasteiger, J.; Jochum, C.; Schubert, W. *Angew. Chem. Int. Edit.* **1979,** *18*(2), 111-123.
180. Joshi, P.V.; Freund, H.; Klein, M.T. *Energy Fuel* **1999,** *13*(4), 877-880.
181. Shen, V.K.; Krekelberg, W.P.; Hatch, H.W., editors. NIST Standard Reference Database Number 173. Gaithersburg MD: National Institute of Standards and Technology.
182. Da Silva, G.; Kim, C.H.; Bozzelli, J.W. *J. Phys. Chem. A* **2006,** *110*(25), 7925-7934.
183. Dellon, L.D.; Sung, C.Y.; Robichaud, D.J.; Broadbelt, L.J. *Ind. Eng. Chem. Res.* **2017,** *56*(37), 10259-10270.
184. De Moor, B.A.; Reyniers, M.F.; Gobin, O.C.; Lercher, J.A.; Marin, G.B. *J. Phys. Chem. C* **2011,** *115*(4), 1204-1219.
185. Nguyen, C.M.; De Moor, B.A.; Reyniers, M.F.; Marin, G.B. *J. Phys. Chem. C* **2011,** *115*(48), 23831-23847.
186. Nguyen, C.M.; Reyniers, M.F.; Marin, G.B. *J. Catal.* **2015,** *322*, 91-103.
187. Li, X.B.; Zhao, Y.; Wang, S.R.; Zhu, Y.Y.; Yang, G.H. *Catal. Lett.* **2016,** *146*(10), 2015-2024.
188. Sepa, J.; Lee, C.; Gorte, R.J.; White, D.; Kassab, E.; Evleth, E.M.; Jessri, H.; Allavena, M. *J. Phys. Chem.* **1996,** *100*(47), 18515-18523.
189. Thamm, H.; Jerschkewitz, H.G.; Stach, H. *Zeolites* **1988,** *8*(2), 151-153.
190. Moses, P.G.; Norskov, J.K. *ACS Catal.* **2013,** *3*(4), 735-745.
191. Fang, Z.T.; Wang, Y.; Dixon, D.A. *J. Phys. Chem. C* **2015,** *119*(41), 23413-23421.
192. Lee, C.C.; Gorte, R.J.; Farneth, W.E. *J. Phys. Chem. B* **1997,** *101*(19), 3811-3817.
193. De Moor, B.A.; Reyniers, M.F.; Sierka, M.; Sauer, J.; Marin, G.B. *J. Phys. Chem. C* **2008,** *112*(31), 11796-11812.
194. Linstrom, P.J.; Mallard, W.G., editors. NIST Chemistry WebBook, NIST Standard Reference Database Number 69, National Institute of Standards and Technology. Gaithersburg MD, 20899.

195. Kouskoulli, M.A.; Smith, D.M.; Nicolaides, A.V. *J. Mol. Struc.-Theochem.* **2007,** *811*(1-3), 355-359.
196. Yan, H.; Feng, X.; Liu, Y.B.; Yang, C.H.; Shan, H.H. *Mol. Catal.* **2017,** *437*, 11-17.
197. Yaluris, G.; Rekoske, J.E.; Aparicio, L.M.; Madon, R.J.; Dumesic, J.A. *J. Catal.* **1995,** *153*(1), 54-64.
198. Wang, I.; Tsai, T.C.; Huang, S.T. *Ind. Eng. Chem. Res.* **1990,** *29*(10), 2005-2012.
199. Salvapati, G.S.; Ramanamurty, K.V.; Janardanarao, M. *J. Mol. Catal.* **1989,** *54*(1), 9-30.
200. Boekfa, B.; Pantu, P.; Probst, M.; Limtrakul, J. *J. Phys. Chem. C* **2010,** *114*(35), 15061-15067.
201. Vermaas, J.V.; Dellon, L.D.; Broadbelt, L.J.; Beckham, G.T.; Crowley, M.F. *ACS Sustain. Chem. Eng.* **2019,** *7*(3), 3443-3453.
202. Yanez, A.J.; Natarajan, P.; Li, W.J.; Mabon, R.; Broadbelt, L.J. *Energy Fuel* **2018,** *32*(2), 1822-1830.
203. Karmanov, A.P.; Monakov, Y.B. *Vysokomol. Soedin.* **1999,** *41*(7), 1200-1205.
204. Flory, P.J. *Principles of Polymer Chemistry*, Ithaca, 1953; 672.
205. Wiberg, K.B.; Crocker, L.S.; Morgan, K.M. *J. Am. Chem. Soc.* **1991,** *113*(9), 3447-3450.
206. Chao, J.; Zwolinski, B.J. *J. Phys. Chem. Ref. Data* **1976,** *5*(2), 319-328.
207. Pedley, J.B.; Naylor, R.D.; Kirby, S.P. *Thermochemical Data of Organic Compounds*, 2nd ed.; London, 1986.
208. Hansen, N.; Kerber, T.; Sauer, J.; Bell, A.T.; Keil, F.J. *J. Am. Chem. Soc.* **2010,** *132*(33), 11525-11538.
209. Denayer, J.F.; Souverijns, W.; Jacobs, P.A.; Martens, J.A.; Baron, G.V. *J. Phys. Chem. B* **1998,** *102*(23), 4588-4597.

Appendices

Appendix A: Supporting Information for "Computational Generation of Lignin Libraries from Diverse Biomass Sources"

A1.1. Hyper-branched Lignin and the Branching Coefficient.

The nature of the different linkage types as well as the multiple bonding locations on a monomer give rise to a heavily branched structure. In this study, we describe this structure as hyper-branched, meaning branches within branches. In the literature, some authors have referred to lignin as a "chaotically branching fractal cluster,"[28,203] which is analogous to the term hyper-branched. Dolgonosov and Gubernatorova[28] provide an excellent explanation and visual representation of the chaotic branching in lignin.

There are two different branching coefficients in this work: the experimental branching coefficient and the simulated branching coefficient. The experimental branching coefficient is a branching probability obtained through a correlation, based on concepts of polymer science,[204] to the molecular weight distribution. A correlation is available for western hemlock[78] and black cottonwood.[79] The simulated branching coefficient is defined as the ratio of branched monomers to total monomers, where a branched monomer is one that is bonded to three or more monomers.

A1.2. The Markov Chain Monte Carlo (MCMC) Method.

Generally, Monte Carlo is the process of estimating properties of a distribution by repeated random sampling of the distribution. In this work, MCMC is the method by which a growing chain of lignin molecules distributed proportionately to a target distribution is formed. In this method, a molecule is evaluated based on its distance metric, a value that represents the deviation between that molecule's simulated properties and the experimental properties. The steps in the MCMC method are as follows. First, an initialization step is performed to select a molecule with a low

distance metric to start the chain. Then, a new molecule, drawn from a previously created reservoir of molecules, is evaluated based on its distance metric. The Metropolis criterion is used to accept this molecule to the Markov chain with a probability defined by the ratio of probabilities of the initial and final states, which are functions of the distance metric. More details are discussed by Yanez and coworkers.[49]

A1.3. The Molecular Weight Distribution Described by the Schultz Distribution.

The molecular weight distribution was described by a truncated Schultz distribution, Equation A.1. *S(x)dx* is the number fraction of polymers having lengths in the range (*x, x+dx*), *x* is the chain length, $\overline{X_n}$ is the number-average chain length, X_{min} is the minimum chain length, Γ is the upper incomplete gamma function, and y and z are real-valued parameters of the distribution. A minimum chain length of two is required, and the maximum chain length can be defined so as to encompass 99.99% of the original distribution.[49]

$$S(x) = \frac{y^{z+1}x^{z-1}\overline{X_n}}{\Gamma(\mathrm{z}+1, \mathrm{y}X_{min})}e^{-yx} \qquad \text{(A.1)}$$

A1.4. The Edge Weight Optimization.

For a given decision tree, each edge has a numerical value, or weight, associated with it between 0 and 1 that represents the conditional probability of its occurrence. There is an additional parameter called the branching propensity. Thus, the optimization problem is stated as follows: "Define the values of [number of edge weights] edge weights plus the branching propensity such that a random walk across the decision tree results in a population of molecules that minimizes *d*."[49] *d* is the distance metric, defined by Yanez and co-workers.[49] The stochastic optimization procedure is as follows. A random set of edge weight values is produced, a population of 100 molecules is generated from these edge weights, and the distance metric for this population is

calculated. This process is repeated for a large number of iterations ($1x10^7$), and the edge weight values corresponding to the population with the lowest distance metric are the optimized values.

Molecule A1. The following is one representative molecule (or chain) of spruce lignin extracted from a library of 100 molecules. The first list is a monomer list specifying the index and type of monomer. The second list is a bond list matrix specifying the connectivity of each monomer to other monomers via positions and bond types.

The monomer list is interpreted as follows:

X:Y "Monomer index **X** is of monomer type **Y**," where **Y** has three possible values:
1 = syringyl
2 = *p*-hydroxyphenyl
3 = guaiacyl

The bond list is interpreted as follows:

A:B:C:D "Monomer index **A** is connected to monomer **B** via bond type **D** at position **C** of monomer **A**," where **C** has four possible values:
1 = β-carbon
2 = 4-oxygen
3 = 5-carbon
4 = α-carbon
and **D** has seven possible values:
1 = β-O-4
2 = β-5
3 = 5-5
4 = 4-O-5
5 = β-1
6 = α-O-4
7 = β- β

Monomer List:
1: 3
2: 3
3: 3
4: 3
5: 3
6: 3
7: 3
8: 3
9: 3
10: 3

11: 3
12: 3
13: 3
14: 3
15: 3
16: 3
17: 2
18: 3
19: 3
20: 3
21: 3
22: 3
23: 3
24: 3
25: 3
26: 3
27: 3
28: 3
29: 3
30: 3
31: 3
32: 3
33: 3
34: 3
35: 3
36: 3
37: 3
38: 3
39: 3
40: 3
41: 3
42: 3
43: 3
44: 3
45: 3
46: 3
47: 3
48: 3
49: 3
50: 1
51: 2
52: 3
53: 3
54: 3

55: 3
56: 3
57: 3
58: 3
59: 3
60: 3
61: 3
62: 3
63: 3
64: 3
65: 3
66: 1
67: 2
68: 3
69: 3
70: 3
71: 3
72: 3
73: 3
74: 3
75: 3
76: 3
77: 3
78: 1
79: 3
80: 1
81: 3
82: 3
83: 3
84: 3
85: 3
86: 3
87: 3
88: 3
89: 3
90: 3
91: 3
92: 3
93: 3
94: 3
95: 3
96: 3
97: 3
98: 3

99: 3
100: 3
101: 3
102: 3
103: 3

Bond List:

Monomer_I	Monomer_II	Pos_Mon_I	Bond_type
1	2	1	5
2	3	3	3
3	4	1	1
4	5	1	8
4	6	4	6
4	7	3	3
5	8	1	1
6	9	1	1
7	10	1	1
8	11	1	1
9	12	1	1
10	13	1	1
11	14	1	8
11	15	4	6
12	16	1	1
13	17	1	1
14	18	1	1
15	19	1	2
16	20	1	5
16	21	3	3
17	22	1	1
18	23	1	8
18	24	4	6
19	25	1	1
20	26	3	3
21	27	1	1
22	28	1	2
23	29	1	1
24	30	1	1
25	31	1	1
26	32	1	2
27	33	1	1
28	34	1	1
29	35	1	8

29	36	4	6
30	37	1	1
31	38	1	1
32	39	1	1
33	40	1	1
34	41	1	5
35	42	1	1
36	43	1	1
37	44	1	1
38	45	1	2
39	46	1	1
40	47	1	2
41	48	3	3
42	49	1	1
43	50	1	1
43	51	3	3
44	52	1	1
45	53	1	5
46	54	1	2
47	55	1	2
48	56	1	1
49	57	1	1
50	58	1	5
51	59	1	5
52	60	1	2
53	61	2	4
54	62	1	1
55	63	1	1
56	64	1	1
57	65	1	1
58	66	2	1
59	67	3	3
60	68	1	1
61	69	1	1
62	70	1	8
62	71	4	6
63	72	1	2
64	73	1	1
65	74	1	5
66	75	2	1
67	76	1	8
67	77	4	6
68	78	1	5
69	79	1	1

70	80	1	1
71	81	1	7
72	82	1	1
73	83	1	2
74	84	3	3
75	85	2	1
76	86	1	2
77	87	1	1
78	88	2	1
79	89	1	1
79	90	3	4
80	91	1	2
81	92	3	3
82	93	1	7
83	94	1	1
84	95	1	1
85	96	3	3
86	97	1	5
87	98	3	3
87	99	1	1
88	100	3	3
89	101	1	2
90	102	1	1
91	103	1	2

Note that in the literature, the bonds may be called by the following names: β-O-4 as arylglycerol-β-aryl ether, α-O-4 as α-aryl ether or noncylic benzyl aryl ether, β-1 as 1,2-diarylpropane, β-5 as phenylcoumaran, β-β as resinol, 4-O-5 as diaryl ether, and 5-5 as biphenyl.

Table A.1. Tabulated experimental values for wheatstraw lignin with corresponding literature source and method of extraction and characterization.

property	details	exp. value	source	method	comments
monomer percentage	syringyl	30	[69]	Bjorkman extraction, 2D-NMR	
	p-hydroxyphenyl	6	[69]	Bjorkman extraction, 2D-NMR	
	guaiacyl	64	[69]	Bjorkman extraction, 2D-NMR	
bond percentage	β-O-4	75 + 2 = 77	[69]	Bjorkman extraction, 2D-NMR	75% β-O-4 + 2% α-oxidized β-O-4
	β-5	11	[69]	Bjorkman extraction, 2D-NMR	
	5-5	0 + 3 = 3	[69]	Bjorkman extraction, 2D-NMR	0% 5-5 + 3% dibenzodioxocins
	4-O-5	0	[69]	Bjorkman extraction, 2D-NMR	
	β-1	0 + 3 = 3	[69]	Bjorkman extraction, 2D-NMR	0% β-1 + 3% spirodienone
	α-O-4	2	[69]	Bjorkman extraction, 2D-NMR	
	β-β	4	[69]	Bjorkman extraction, 2D-NMR	
molecular weight (Da)	number-average	1850	[69]	Bjorkman extraction, gel permeation chromatography	
	weight-average	4210	[69]	Bjorkman extraction, gel permeation chromatography	
branching coefficient		0.225	[78]	Correlation between branching coefficient and weight-average molecular weight for western hemlock wood	

Table A.2. Tabulated experimental values for *miscanthus giganteus* lignin with corresponding literature source and method of extraction and characterization.

property	details	exp. value	source	method	comments
monomer percentage	syringyl	50	[84]	95% dioxane extraction, 2D-NMR	
	p-hydroxyphenyl	4	[84]	95% dioxane extraction, 2D-NMR	
	guaiacyl	46	[84]	95% dioxane extraction, 2D-NMR	
bond percentage	β-O-4	68	[84]	95% dioxane extraction, 2D-NMR	
	β-5	15	[84]	95% dioxane extraction, 2D-NMR	
	5-5	0	[84]	95% dioxane extraction, 2D-NMR	
	4-O-5	0	[84]	95% dioxane extraction, 2D-NMR	
	β-1	0	[84]	95% dioxane extraction, 2D-NMR	
	α-O-4	0	[84]	95% dioxane extraction, 2D-NMR	
	β-β	17	[84]	95% dioxane extraction, 2D-NMR	
molecular weight (Da)	number-average	1240	[84]	95% dioxane extraction, size exclusion chromatography	
	weight-average	2310	[84]	95% dioxane extraction, size exclusion chromatography	
branching coefficient		0		Linear due to bond types	

Table A.3. Tabulated experimental values for spruce lignin with corresponding literature source and method of extraction and characterization.

property	details	exp. value	source	method	comments
monomer percentage	syringyl	1	[61]	Bjorkman extraction, oxidative degradation	
	p-hydroxyphenyl	5	[61]	Bjorkman extraction, oxidative degradation	
	guaiacyl	94	[61]	Bjorkman extraction, oxidative degradation	
bond percentage	β-O-4	51	[61]	Bjorkman extraction, oxidative degradation	includes glyceraldehyde-2-aryl ether
	β-5	12	[61]	Bjorkman extraction, oxidative degradation	
	5-5	13	[63]	Bjorkman extraction, C-NMR	frequency of 5-5 linkage
	4-O-5	4	[61]	Bjorkman extraction, oxidative degradation	
	β-1	7 + AVG(3, 3) = 10	[85], ([77], [61])	acidolysis & (NMR, permanganate oxidative degradation)	7% β-1 + AVG(3% spirodienone, 3% arylisochroman)
	α-O-4	8	[85]	Bjorkman extraction, acidolysis	
	β-β	2	[63]	Bjorkman extraction, C-NMR	
molecular weight (Da)	number-average	6400	[87]	Bjorkman extraction, milled wood lignin	
	weight-average	23500	[87]	Bjorkman extraction, milled wood lignin	
branching coefficient		0.301	[78]	Correlation between branching coefficient and weight-average molecular weight for western hemlock wood	

Table A.4. Experimental values for beech lignin with corresponding literature source and method of extraction and characterization.

property	details	exp. value	source	method	comments
monomer percentage	syringyl	36	[88]	Bjorkman extraction, oxidative degradation	
	p-hydroxyphenyl	0	[88]	Bjorkman extraction, oxidative degradation	
	guaiacyl	64	[88]	Bjorkman extraction, oxidative degradation	
bond percentage	β-O-4	60	[64]	Bjorkman extraction, C-NMR	
	β-5	6	[72]	Acetone/water (9:1) extraction, thioacetic acid degradation	
	5-5	2	[72]	Acetone/water (9:1) extraction, thioacetic acid degradation	Rounded from 2.3%
	4-O-5	2	[72]	Acetone/water (9:1) extraction, thioacetic acid degradation	Rounded from 1.5%
	β-1	15	[72]	Acetone/water (9:1) extraction, thioacetic acid degradation	
	α-O-4	5	[64]	Bjorkman extraction, C-NMR	
	β-β	6 + 2.5 + 2 + 0.5 = 10	[64]	Acetone/water (9:1) extraction, thioacetic acid degradation	6% syringaresinol and pinoresinol units + 2.5% α-β bonds + 2% β-β bonds in dibenzyltetrahydrofuran units + 0.5% β-β and α-6 bonds (tetralin units)
molecular weight (Da)	number-average	3690	[89]	milled wood lignin	
	weight-average	5510	[89]	milled wood lignin	
branching coefficient		0.088	[79]	Correlation between branching coefficient and weight-average molecular weight for black cottonwood wood	

Table A.5. Tabulated experimental values for birch lignin with corresponding literature source and method of extraction and characterization.

property	details	exp. value	source	method	comments
monomer percentage	syringyl	50	[90]	acetone extraction, C-NMR	
	p-hydroxyphenyl	0	[90]	acetone extraction, C-NMR	
	guaiacyl	50	[90]	acetone extraction, C-NMR	
bond percentage	β-O-4	62	[91]	Bjorkman extraction, oxidative degradation	includes glyceraldehyde-2-aryl ether
	β-5	5 + 1 = 6	[91]	Bjorkman extraction, oxidative degradation	5% β-5 + 1% structures condensed at 5-position
	5-5	4.5 + 2.5 = 7	[91]	Bjorkman extraction, oxidative degradation	4.5% 5-5 + 2.5% structures condensed at 2- or 6- positions
	4-O-5	7	[91]	Bjorkman extraction, oxidative degradation	Rounded from 6.5%
	β-1	7	[24], [92]	Bjorkman extraction, acidolysis	reported by [24], estimated from [92]
	α-O-4	8	[24], [85]	Bjorkman extraction, acidolysis	reported by [24], same value from spruce lignin was assumed
	β-β	3	[24]	Bjorkman extraction, acidolysis	
molecular weight (Da)	number-average	1878	[28]	low-molecular lignins of 20-30 links	average taken
	weight-average	4600	[28]	low-molecular lignins of 20-30 links	average taken
branching coefficient		0.055	[79]	Correlation between branching coefficient and weight-average molecular weight for black cottonwood wood	

Table A.6. The 28 possible combinations of bonds that would constitute a branching situation and the corresponding positions that are occupied on the monomer for that combination. Note that several unlisted combinations of bonds are not allowable due to the structure of the linkage types. A single combination cannot consist of both a bond at the 1-C position and a bond at the α-C and/or the β-C position because the propanoid region containing these latter two positions detaches when a 1-β bond forms. Also, the 5-β linkage is not listed and does not occur in the decision trees because it requires two bonding positions – the 5-C and 4-O positions. Finally, combinations including the 4-O-α bond were not included because the decision trees prohibit the bond from growing in this direction.

	Bonds				Positions			
1	α-O-4/β-O-4		4-O-β	5-5	α-C	β-C	4-O	5-C
2	α-O-4/β-O-4		4-O-β	5-O-4	α-C	β-C	4-O	5-C
3	α-O-4/β-O-4		4-O-5	5-5	α-C	β-C	4-O	5-C
4	α-O-4/β-O-4		4-O-5	5-O-4	α-C	β-C	4-O	5-C
5	α-O-4	β-O-4	4-O-β		α-C	β-C	4-O	
6	α-O-4	β-O-4	4-O-5		α-C	β-C	4-O	
7	α-O-4	β-O-4	5-5		α-C	β-C	5-C	
8	α-O-4	β-O-4	5-O-4		α-C	β-C	5-C	
9	β-O-4	4-O-β	5-5		β-C	4-O	5-C	
10	β-O-4	4-O-β	5-O-4		β-C	4-O	5-C	
11	β-O-4	4-O-5	5-5		β-C	4-O	5-C	
12	β-O-4	4-O-5	5-O-4		β-C	4-O	5-C	
13	β-1	4-O-β	5-5		β-C	4-O	5-C	
14	β-1	4-O-β	5-O-4		β-C	4-O	5-C	
15	β-1	4-O-5	5-5		β-C	4-O	5-C	
16	β-1	4-O-5	5-O-4		β-C	4-O	5-C	
17	1-β	4-O-β	5-5		1-C	4-O	5-C	
18	1-β	4-O-β	5-O-4		1-C	4-O	5-C	
19	1-β	4-O-5	5-5		1-C	4-O	5-C	
20	1-β	4-O-5	5-O-4		1-C	4-O	5-C	
21	β-β	4-O-β	5-5		β-C	4-O	5-C	
22	β-β	4-O-β	5-O-4		β-C	4-O	5-C	
23	β-β	4-O-5	5-5		β-C	4-O	5-C	
24	β-β	4-O-5	5-O-4		β-C	4-O	5-C	
25	β-5	4-O-β	5-5		β-C	4-O	5-C	
26	β-5	4-O-β	5-O-4		β-C	4-O	5-C	
27	β-5	4-O-5	5-5		β-C	4-O	5-C	
28	β-5	4-O-5	5-O-4		β-C	4-O	5-C	

Appendix B: Supporting Information for "Group Additivity Determination for Oxygenates, Oxonium Ions, and Oxygen-Containing Carbenium Ions"

Table B.1. Absolute enthalpies at 298 K and experimental enthalpies of formation for atomic species for use in atomization enthalpy and proton affinity calculations. All experimental values are from Chase.[142] Experimental uncertainties are listed where available.

Species	H^{298} (Hartree)	Exp ΔH_f^o (kcal/mol)
H	-0.499	52.103 ± 0.001
H^+	0.002	365.69 [a]
C	-37.832	171.29 ± 0.11
C^+	-37.419	430.984 [a]
O	-75.043	59.55 ± 0.02
O^+	-74.545	373.568 [a]

[a] Experimental uncertainty not reported in the literature.

Table B.2. Absolute enthalpies at 298 K and atomization enthalpies for reference molecules and species with available experimental data. The species numbers refer to Figure 3.3 in Chapter 3.

Molecule	H^{298} (Hartree)	ΔH_a (kcal/mol)	Species	H^{298} (Hartree)	ΔH_a (kcal/mol)
CH_3^+	-39.434	324.57	(1)	-193.335	1076.21
$C_3H_7^+$	-118.080	943.48	(2)	-232.608	1354.37
$C_3H_5^+$	-116.860	804.35	(3)	-232.615	1358.79
$C_4H_9^+$	-157.385	1241.43	(26)	-231.405	1226.01
H_2C^+-OH	-114.718	476.04	(150)	-194.513	1243.30
H_3C-HC^+-OH	-154.030	778.15	(151)	-233.786	1521.12
H_3O^+	-76.654	384.22	(153)	-155.223	954.25
$CH_3O^+H_2$	-115.933	665.64	(154)	-194.498	1233.75
CH_3-O^+H-CH_3	-155.210	1059.68	(25)	-307.457	1286.97
CH_4	-40.461	397.49			
C_2H_6	-79.734	675.01			
C_2H_4	-78.518	538.42			
C_3H_8	-119.010	955.29			
C_3H_6	-117.799	821.43			
C_4H_6	-155.869	970.92			
H_2O	-76.393	221.01			
CH_3OH	-115.647	487.14			
C_2H_5OH	-154.929	770.48			
H_3C-O-CH_3	-154.910	758.74			
H_2C=O	-114.449	361.64			
H_3C-CH=O	-153.739	650.06			
CH_3-CO-CH_3	-193.027	937.40			
H_2C=CH-OH	-153.723	639.82			
H_2C=CH-O-CH=CH_2	-231.053	1059.68			
H_2C=CH-C(CH_3)=O	-231.091	1083.55			
O=CH-HC=O	-227.732	617.28			
O=CH-OH	-189.689	484.86			

Table B.3. Comparison of experimental and calculated enthalpies of formation in kcal/mol at 298 K for species using the method of atomization enthalpies. Experimental uncertainties are listed where available.

Charge	Species	Exp ΔH_f^o	Calc ΔH_f^o	Abs Dev
C^+	(1)	119.36 ± 2.04 [a]	121.63	2.27
	(2)	116.29 ± 1.90 [a]	118.96	2.67
	(3)	110.97 ± 2.10 [a]	114.55	3.58
	(26)	138.69 ± 4.53 [a]	143.11	4.42
O^+	(150)	110.99 ± 2.00 [a]	112.96	1.97
	(151)	107.99 ± 2.10 [a]	110.64	2.65
	(153)	123.99 ± 2.00 [a]	126.52	2.53
	(154)	120.66 ± 2.06 [a]	122.51	1.85
0	(25)	-67.09 ± 0.90 [143]	-65.89	1.20
mean absolute dev				2.57

[a] Calculated from proton affinity calculations.
Abbreviations: Exp = experimental value from literature; Calc = value from this study; Abs Dev = absolute deviation, |Calc-Exp|; C^+ = carbenium ion species; O^+ = oxonium ion species; 0 = neutral species.

Table B.4. Part 1 of 11. List of isodesmic reactions with their respective enthalpies of formation calculated from G4 results. All values are in kcal/mol

Species	Reaction	ΔH_f^o	Avg. ΔH_f^o
(1)	$H_2C^+\text{-}OH + C_3H_7^+ \rightarrow CH_3^+ + (1)$	118.48	119.43
	$H_3C\text{-}HC^+\text{-}OH + C_4H_9^+ \rightarrow C_3H_7^+ + (1)$	120.38	
(2)	$H_2C^+\text{-}OH + C_3H_7^+ + CH_3OH \rightarrow CH_3^+ + H_2O + (2)$	116.24	117.19
	$H_3C\text{-}HC^+\text{-}OH + C_4H_9^+ + CH_3OH \rightarrow C_3H_7^+ + H_2O + (2)$	118.14	
(3)	$H_2C^+\text{-}OH + C_3H_7^+ + C_2H_6 \rightarrow CH_3^+ + CH_4 + (3)$	111.18	112.13
	$H_3C\text{-}HC^+\text{-}OH + C_4H_9^+ + C_2H_6 \rightarrow C_3H_7^+ + CH_4 + (3)$	113.08	
(4)	$CH_3OH + C_3H_7^+ \rightarrow CH_4 + (4)$	151.75	151.73
	$C_2H_5OH + C_3H_7^+ \rightarrow C_2H_6 + (4)$	151.71	
(5)	$CH_3OH + C_4H_9^+ \rightarrow CH_4 + (5)$	134.25	134.23
	$C_2H_5OH + C_4H_9^+ \rightarrow C_2H_6 + (5)$	134.21	
(6)	$2\,CH_3OH + C_4H_9^+ \rightarrow CH_4 + H_2O + (6)$	135.31	135.15
	$H_3C\text{-}O\text{-}CH_3 + C_4H_9^+ \rightarrow CH_4 + (6)$	134.99	
(7)	$CH_3OH + C_4H_9^+ + C_2H_6 \rightarrow 2\,CH_4 + (7)$	124.12	124.10
	$C_2H_5OH + C_4H_9^+ \rightarrow CH_4 + (7)$	124.08	
(8)	$2\,CH_3OH + C_4H_9^+ + C_2H_6 \rightarrow 2\,CH_4 + H_2O + (8)$	121.57	121.41
	$H_3C\text{-}O\text{-}CH_3 + C_4H_9^+ + C_2H_6 \rightarrow 2\,CH_4 + (8)$	121.25	
(9)	$C_2H_5OH + C_4H_9^+ + C_2H_6 \rightarrow 2\,CH_4 + (9)$	116.80	116.84
	$CH_3OH + C_4H_9^+ + C_3H_8 \rightarrow 2\,CH_4 + (9)$	116.87	
(10)	$H_2C^+\text{-}OH + CH_3OH + C_4H_9^+ \rightarrow C_3H_7^+ + H_2O + (10)$	137.58	137.08
	$H_3C\text{-}HC^+\text{-}OH + CH_3OH \rightarrow H_2O + (10)$	136.58	
(11)	$2\,CH_3OH + C_4H_9^+ + C_2H_6 \rightarrow 3\,CH_4 + (11)$	81.48	81.44
	$2\,C_2H_5OH + C_4H_9^+ \rightarrow CH_4 + C_2H_6 + (11)$	81.40	
(12)	$3\,CH_3OH + C_4H_9^+ + C_2H_6 \rightarrow 3\,CH_4 + H_2O + (12)$	85.39	85.21
	$H_3C\text{-}O\text{-}CH_3 + C_2H_5OH + C_4H_9^+ \rightarrow 2\,CH_4 + (12)$	85.04	
(13)	$4\,CH_3OH + C_4H_9^+ + C_2H_6 \rightarrow 3\,CH_4 + 2\,H_2O + (13)$	91.66	91.35
	$2\,H_3C\text{-}O\text{-}CH_3 + C_4H_9^+ + C_2H_6 \rightarrow 3\,CH_4 + (13)$	91.04	
(14)	$2\,CH_3OH + C_3H_6 + C_2H_6 \rightarrow 3\,CH_4 + (14)$	-88.56	-88.65
	$2\,C_2H_5OH + C_2H_4 \rightarrow 2\,CH_4 + (14)$	-88.75	
(15)	$3\,CH_3OH + C_3H_6 + C_2H_6 \rightarrow 3\,CH_4 + H_2O + (15)$	-81.99	-82.22
	$H_3C\text{-}O\text{-}CH_3 + C_2H_5OH + C_2H_6 + C_2H_4 \rightarrow 3\,CH_4 + (15)$	-82.45	
(16)	$4\,CH_3OH + C_3H_6 + C_2H_6 \rightarrow 3\,CH_4 + 2\,H_2O + (16)$	-76.33	-76.69
	$2\,H_3C\text{-}O\text{-}CH_3 + C_2H_4 + C_3H_8 + C_2H_6 \rightarrow 3\,CH_4 + (16)$	-77.04	
(17)	$CH_3OH + C_4H_9^+ + C_3H_8 \rightarrow 2\,CH_4 + (17)$	110.58	110.57
	$CH_3OH + C_4H_9^+ + 2\,C_2H_6 \rightarrow 3\,CH_4 + (17)$	110.55	
(18)	$CH_3OH + C_4H_9^+ + C_3H_8 + C_2H_6 \rightarrow 3\,CH_4 + (18)$	105.08	105.10
	$CH_3OH + C_4H_9^+ + 3\,C_2H_6 \rightarrow 4\,CH_4 + (18)$	105.11	

Table B.4. Part 2 of 11. List of isodesmic reactions with their respective enthalpies of formation.

Species	Reaction	ΔH_f^o	Avg. ΔH_f^o
(19)	$2\,CH_3OH + C_4H_9^+ + 2\,C_2H_6 \rightarrow 3\,CH_4 + H_2O + (19)$	110.96	110.97
	$2\,CH_3OH + C_4H_9^+ + C_3H_8 \rightarrow 2\,CH_4 + H_2O + (19)$	110.98	
(20)	$2\,CH_3OH + C_4H_9^+ \rightarrow 2\,CH_4 + (20)$	95.03	94.99
	$2\,C_2H_5OH + C_4H_9^+ \rightarrow 2\,C_2H_6 + (20)$	94.95	
(21)	$3\,CH_3OH + C_4H_9^+ \rightarrow 2\,CH_4 + H_2O + (21)$	97.68	97.50
	$H_3C\text{-}O\text{-}CH_3 + C_2H_5OH + C_4H_9^+ \rightarrow C_2H_6 + CH_4 + (21)$	97.32	
(22)	$4\,CH_3OH + C_4H_9^+ \rightarrow 2\,CH_4 + 2\,H_2O + (22)$	102.46	102.15
	$2\,H_3C\text{-}O\text{-}CH_3 + C_4H_9^+ \rightarrow 2\,CH_4 + (22)$	101.84	
(23)	$2\,H_2C{=}CH\text{-}OH \rightarrow C_2H_4 + (23)$	-78.40	-77.83
	$H_2C{=}CH\text{-}OH + CH_3OH \rightarrow CH_4 + (23)$	-77.26	
(24)	$2\,H_2C{=}CH\text{-}OH + CH_3OH \rightarrow C_2H_4 + H_2O + (24)$	-74.04	-73.47
	$H_2C{=}CH\text{-}OH + 2\,CH_3OH \rightarrow CH_4 + H_2O + (24)$	-72.90	
(25)	$2\,H_2C{=}CH\text{-}OH + 2\,CH_3OH \rightarrow C_2H_4 + 2\,H_2O + (25)$	-67.59	-67.02
	$H_2C{=}CH\text{-}OH + 3\,CH_3OH \rightarrow CH_4 + 2\,H_2O + (25)$	-66.45	
(26)	$H_2C^+\text{-}OH + C_4H_9^+ + C_3H_5^+ \rightarrow C_3H_7^+ + CH_3^+ + (26)$	140.90	140.40
	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ \rightarrow CH_3^+ + (26)$	139.90	
(27)	$H_2C^+\text{-}OH + CH_3OH + C_4H_9^+ + C_3H_5^+ \rightarrow C_3H_7^+ + CH_3^+ + H_2O + (27)$	140.28	139.79
	$H_3C\text{-}HC^+\text{-}OH + CH_3OH + C_3H_5^+ \rightarrow CH_3^+ + H_2O + (27)$	139.29	
(28)	$H_2C^+\text{-}OH + CH_3OH + C_4H_9^+ + C_3H_5^+ + C_3H_6 \rightarrow C_3H_7^+ + CH_3^+ + H_2O + C_2H_4 + (28)$	125.93	125.43
	$H_3C\text{-}HC^+\text{-}OH + CH_3OH + C_3H_5^+ + C_3H_6 \rightarrow CH_3^+ + H_2O + (28)$	124.93	
(29)	$H_2C{=}CH\text{-}OH + C_3H_5^+ + C_4H_9^+ \rightarrow C_2H_4 + C_3H_7^+ + (29)$	161.20	159.75
	$H_2C{=}CH\text{-}OH + C_3H_5^+ + 2\,C_3H_7^+ \rightarrow C_2H_4 + C_4H_9^+ + CH_3^+ + (29)$	158.30	
(30)	$H_2C{=}CH\text{-}OH + C_3H_5^+ + C_3H_7^+ \rightarrow C_2H_4 + CH_3^+ + (30)$	142.58	144.03
	$H_2C{=}CH\text{-}OH + C_3H_5^+ + 2\,C_4H_9^+ \rightarrow C_2H_4 + 2\,C_3H_7^+ + (30)$	145.48	
(31)	$H_2C{=}CH\text{-}OH + CH_3OH + C_3H_5^+ + C_3H_7^+ \rightarrow C_2H_4 + CH_3^+ + H_2O + (31)$	144.53	145.98
	$H_2C{=}CH\text{-}OH + CH_3OH + C_3H_5^+ + 2\,C_4H_9^+ \rightarrow C_2H_4 + 2\,C_3H_7^+ + H_2O + (31)$	147.43	
(32)	$C_4H_9^+ + CH_3OH + C_3H_6 \rightarrow 2\,CH_4 + (32)$	146.34	146.32
	$C_4H_9^+ + C_2H_5OH + C_3H_6 \rightarrow C_2H_6 + CH_4 + (32)$	146.31	
(33)	$C_4H_9^+ + 2\,CH_3OH + C_3H_6 \rightarrow 2\,CH_4 + H_2O + (33)$	148.37	148.15
	$H_3C\text{-}O\text{-}CH_3 + C_4H_9^+ + C_2H_6 + C_2H_4 \rightarrow 3\,CH_4 + (33)$	147.94	
(34)	$C_4H_9^+ + 2\,CH_3OH + 2\,C_3H_6 \rightarrow 2\,CH_4 + C_2H_4 + H_2O + (34)$	138.65	138.40
	$H_3C\text{-}O\text{-}CH_3 + C_4H_9^+ + C_3H_8 + C_2H_4 \rightarrow 3\,CH_4 + (34)$	138.14	
(35)	$H_2C^+\text{-}OH + C_3H_5^+ \rightarrow CH_3^+ + (35)$	158.44	157.94
	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + C_3H_7^+ \rightarrow CH_3^+ + C_4H_9^+ + (35)$	157.44	
(36)	$H_2C^+\text{-}OH + C_3H_5^+ + CH_3OH \rightarrow CH_3^+ + H_2O + (36)$	154.77	154.27
	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + CH_3OH + C_3H_7^+ \rightarrow CH_3^+ + C_4H_9^+ + H_2O + (36)$	153.77	

Table B.4. Part 3 of 11. List of isodesmic reactions with their respective enthalpies of formation.

Species	Reaction	ΔH_f^o	Avg. ΔH_f^o
(37)	$H_2C^+\text{-}OH + C_3H_5^+ + CH_3OH + C_3H_6 \rightarrow CH_3^+ + C_2H_4 + H_2O$ + (37)	142.78	142.29
	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + CH_3OH + C_3H_6 + C_3H_7^+ \rightarrow CH_3^+ + C_4H_9^+ + C_2H_4 + H_2O$ + (37)	141.79	
(38)	$H_3C\text{-}O\text{-}CH_3 + C_4H_9^+ + 2\,C_3H_6 + C_2H_6 \rightarrow C_2H_4 + 4\,CH_4$ + (38)	150.22	150.37
	$2\,CH_3OH + C_4H_9^+ + 2\,C_3H_6 + C_2H_6 \rightarrow C_2H_4 + 4\,CH_4 + H_2O$ + (38)	150.53	
(39)	$H_3C\text{-}O\text{-}CH_3 + C_4H_9^+ + 3\,C_3H_6 + C_2H_6 \rightarrow 2\,C_2H_4 + 4\,CH_4$ + (39)	134.75	134.90
	$2\,CH_3OH + C_4H_9^+ + 3\,C_3H_6 + C_2H_6 \rightarrow 2\,C_2H_4 + 4\,CH_4 + H_2O$ + (39)	135.06	
(40)	$H_3C\text{-}O\text{-}CH_3 + C_4H_9^+ + 4\,C_3H_6 + C_2H_6 \rightarrow 3\,C_2H_4 + 4\,CH_4$ + (40)	126.00	126.16
	$2\,CH_3OH + C_4H_9^+ + 4\,C_3H_6 + C_2H_6 \rightarrow 3\,C_2H_4 + 4\,CH_4 + H_2O$ + (40)	126.31	
(41)	$CH_3OH + 2\,C_3H_6 + C_2H_6 \rightarrow 3\,CH_4$ + (41)	-22.83	-22.85
	$C_2H_5OH + 2\,C_3H_6 \rightarrow 2\,CH_4$ + (41)	-22.87	
(42)	$H_3C\text{-}O\text{-}CH_3 + 2\,C_3H_6 + C_2H_6 \rightarrow 3\,CH_4$ + (42)	-16.36	-16.20
	$2\,CH_3OH + 2\,C_3H_6 + C_2H_6 \rightarrow 3\,CH_4 + H_2O$ + (42)	-16.04	
(43)	$CH_3OH + 3\,C_3H_6 + C_2H_6 \rightarrow C_2H_4 + 3\,CH_4$ + (43)	-30.68	-30.41
	$2\,CH_3OH + 2\,C_3H_6 + C_2H_6 \rightarrow 3\,CH_4 + H_2O$ + (43)	-30.15	
(44)	$CH_3OH + 2\,C_3H_6 \rightarrow 2\,CH_4$ + (44)	-13.08	-13.10
	$C_2H_5OH + 2\,C_3H_6 \rightarrow C_2H_6 + CH_4$ + (44)	-13.12	
(45)	$H_3C\text{-}O\text{-}CH_3 + 2\,C_3H_6 \rightarrow 2\,CH_4$ + (45)	-8.52	-8.36
	$2\,CH_3OH + 2\,C_3H_6 \rightarrow 2\,CH_4 + H_2O$ + (45)	-8.21	
(46)	$CH_3OH + 3\,C_3H_6 \rightarrow C_2H_4 + 2\,CH_4$ + (46)	-22.08	-22.10
	$C_2H_5OH + 3\,C_3H_6 \rightarrow C_2H_4 + C_2H_6 + CH_4$ + (46)	-22.11	
(47)	$H_3C\text{-}O\text{-}CH_3 + CH_3O^+H_2 + 2\,C_2H_5OH + 2\,C_3H_8 \rightarrow C_2H_6 + 5\,CH_4$ + (47)	-38.63	-38.73
	$2\,H_3C\text{-}O\text{-}CH_3 + CH_3O^+H_2 + 3\,C_3H_8 + H_2O \rightarrow C_2H_6 + 6\,CH_4$ + (47)	-38.83	
(48)	$H_3C\text{-}O\text{-}CH_3 + CH_3O^+H_2 + 3\,CH_3OH + 3\,C_3H_8 \rightarrow C_2H_6 + H_2O + 6\,CH_4$ + (48)	-35.71	-35.86
	$2\,H_3C\text{-}O\text{-}CH_3 + CH_3O^+H_2 + CH_3OH + 3\,C_3H_8 \rightarrow C_2H_6 + 6\,CH_4$ + (48)	-36.02	
(49)	$3\,H_3C\text{-}O\text{-}CH_3 + CH_3O^+H_2 + 3\,C_3H_8 \rightarrow C_2H_6 + 6\,CH_4$ + (49)	-34.24	-34.08
	$2\,H_3C\text{-}O\text{-}CH_3 + CH_3O^+H_2 + 2\,CH_3OH + 3\,C_3H_8 \rightarrow C_2H_6 + 6\,CH_4 + H_2O$ + (49)	-33.93	
(50)	$2\,H_3C\text{-}O\text{-}CH_3 + CH_3O^+H_2 + C_3H_6 + C_3H_8 + H_2O \rightarrow 7\,CH_4$ + (50)	-10.96	-10.85
	$2\,C_2H_5OH + CH_3O^+H_2 + C_3H_6 + C_3H_8 + H_3C\text{-}O\text{-}CH_3 \rightarrow 6\,CH_4$ + (50)	-10.75	
(51)	$3\,H_3C\text{-}O\text{-}CH_3 + CH_3O^+H_2 + C_3H_6 + 2\,C_3H_8 + H_2O \rightarrow CH_3OH + 7\,CH_4$ + (51)	-7.21	-6.74
	$5\,CH_3OH + CH_3O^+H_2 + C_3H_6 + 2\,C_3H_8 \rightarrow 7\,CH_4 + 2\,H_2O$ + (51)	-6.27	
(52)	$3\,H_3C\text{-}O\text{-}CH_3 + CH_3O^+H_2 + C_3H_6 + 2\,C_3H_8 \rightarrow 7\,CH_4$ + (52)	-4.53	-4.06
	$6\,CH_3OH + CH_3O^+H_2 + C_3H_6 + 2\,C_3H_8 \rightarrow 7\,CH_4 + 3\,H_2O$ + (52)	-3.59	
(53)	$2\,H_3C\text{-}O\text{-}CH_3 + CH_3O^+H_2 + 3\,C_3H_6 + C_2H_6 + H_2O \rightarrow 2\,C_2H_4 + 6\,CH_4$ + (53)	-4.38	-4.07
	$4\,H_3C\text{-}O\text{-}CH_3 + CH_3O^+H_2 + 3\,C_3H_6 + C_2H_6 \rightarrow 2\,C_2H_4 + 6\,CH_4$ + (53)	-3.75	
(54)	$3\,H_3C\text{-}O\text{-}CH_3 + CH_3O^+H_2 + 3\,C_3H_6 + C_2H_6 + H_2O \rightarrow CH_3OH + 2\,C_2H_4 + 6\,CH_4$ + (54)	-2.12	-1.65
	$5\,CH_3OH + CH_3O^+H_2 + 3\,C_3H_6 + C_2H_6 \rightarrow 2\,C_2H_4 + 6\,CH_4 + 2\,H_2O$ + (54)	-1.18	

Table B.4. Part 4 of 11. List of isodesmic reactions with their respective enthalpies of formation.

Species	Reaction	ΔH_f^o	Avg. ΔH_f^o
(55)	$3\,H_3C\text{-}O\text{-}CH_3 + CH_3O^+H_2 + 3\,C_3H_6 + C_2H_6 \rightarrow 2\,C_2H_4 + 6\,CH_4$ + (55)	0.44	0.91
	$6\,CH_3OH + CH_3O^+H_2 + 3\,C_3H_6 + C_2H_6 \rightarrow 2\,C_2H_4 + 6\,CH_4 + 3\,H_2O$ + (55)	1.38	
(56)	$H_2C^+\text{-}OH + C_4H_9^+ + H_2C{=}CH\text{-}OH \rightarrow C_3H_7^+ + H_2O$ + (56)	160.73	160.24
	$H_3C\text{-}HC^+\text{-}OH + H_2C{=}CH\text{-}OH \rightarrow H_2O$ + (56)	159.74	
(57)	$H_2C^+\text{-}OH + C_3H_7^+ + H_2C{=}CH\text{-}OH \rightarrow CH_3^+ + H_2O$ + (57)	139.92	140.87
	$H_3C\text{-}HC^+\text{-}OH + H_2C{=}CH\text{-}OH + C_4H_9^+ \rightarrow C_3H_7^+ + H_2O$ + (57)	141.83	
(58)	$H_2C^+\text{-}OH + C_4H_9^+ + H_2C{=}CH\text{-}OH + C_3H_6 \rightarrow C_3H_7^+ + C_2H_4 + H_2O$ + (58)	147.14	146.64
	$H_3C\text{-}HC^+\text{-}OH + H_2C{=}CH\text{-}OH + C_3H_6 \rightarrow C_2H_4 + H_2O$ + (58)	146.14	
(59)	$H_3C\text{-}CO\text{-}CH_3 + H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + H_2C{=}O \rightarrow CH_3^+ + C_3H_6 + H_3C\text{-}CH{=}O$ + (59)	121.50	121.77
	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + 2\,H_3C\text{-}CH{=}O \rightarrow CH_3^+ + C_2H_4 + H_3C\text{-}CO\text{-}CH_3$ + (59)	122.05	
(60)	$H_3C\text{-}CO\text{-}CH_3 + H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + CH_3OH \rightarrow CH_3^+ + C_3H_6 + H_2O$ + (60)	99.51	99.78
	$H_3C\text{-}CH{=}O + H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + C_2H_5OH + CH_4 \rightarrow CH_3^+ + C_2H_4 + C_2H_6 + H_2O$ + (60)	100.05	
(61)	$H_3C\text{-}CO\text{-}CH_3 + H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + H_2C{=}O + C_2H_6 \rightarrow CH_3^+ + C_3H_6 + H_3C\text{-}CH{=}O + CH_4$ + (61)	109.76	110.04
	$2\,H_3C\text{-}CH{=}O + H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + C_2H_6 \rightarrow CH_3^+ + C_2H_4 + H_3C\text{-}CO\text{-}CH_3 + CH_4$ + (61)	110.31	
(62)	$C_4H_9^+ + CH_3OH + H_3C\text{-}CH{=}O \rightarrow 2\,CH_4$ + (62)	111.04	111.02
	$C_4H_9^+ + C_2H_5OH + H_3C\text{-}CH{=}O \rightarrow C_2H_6 + CH_4$ + (62)	111.00	
(63)	$C_4H_9^+ + 2\,CH_3OH + H_3C\text{-}CH{=}O \rightarrow 2\,CH_4 + H_2O$ + (63)	114.31	114.15
	$C_4H_9^+ + H_3C\text{-}O\text{-}CH_3 + H_3C\text{-}CH{=}O \rightarrow 2\,CH_4$ + (63)	114.00	
(64)	$C_4H_9^+ + 2\,CH_3OH + 2\,H_3C\text{-}CH{=}O \rightarrow 2\,CH_4 + H_2O + H_2C{=}O$ + (64)	89.30	89.16
	$C_4H_9^+ + H_3C\text{-}O\text{-}CH_3 + H_3C\text{-}CO\text{-}CH_3 \rightarrow 2\,CH_4$ + (64)	89.02	
(65)	$CH_3O^+H_2 + C_3H_6 \rightarrow CH_4$ + (65)	149.11	148.74
	$CH_3\text{-}O^+H\text{-}CH_3 + C_3H_6 + H_3O^+ \rightarrow CH_3O^+H_2 + CH_4$ + (65)	148.37	
(66)	$2\,CH_3O^+H_2 + C_3H_6 \rightarrow CH_4 + H_3O^+$ + (66)	145.97	145.60
	$CH_3\text{-}O^+H\text{-}CH_3 + C_3H_6 \rightarrow CH_4$ + (66)	145.23	
(67)	$3\,CH_3O^+H_2 + C_3H_6 \rightarrow CH_4 + 2\,H_3O^+$ + (67)	143.71	143.34
	$CH_3\text{-}O^+H\text{-}CH_3 + CH_3O^+H_2 + C_3H_6 \rightarrow CH_4 + H_3O^+$ + (67)	142.97	
(68)	$CH_3O^+H_2 + H_3C\text{-}CH{=}O \rightarrow CH_4$ + (68)	123.22	122.85
	$CH_3\text{-}O^+H\text{-}CH_3 + H_3C\text{-}CH{=}O + H_3O^+ \rightarrow CH_3O^+H_2 + CH_4$ + (68)	122.48	
(69)	$2\,CH_3O^+H_2 + H_3C\text{-}CH{=}O \rightarrow CH_4 + H_3O^+$ + (69)	105.46	105.09
	$CH_3\text{-}O^+H\text{-}CH_3 + H_3C\text{-}CH{=}O \rightarrow CH_4$ + (69)	104.72	
(70)	$3\,CH_3O^+H_2 + H_3C\text{-}CH{=}O \rightarrow CH_4 + 2\,H_3O^+$ + (70)	108.26	107.89
	$CH_3\text{-}O^+H\text{-}CH_3 + CH_3O^+H_2 + H_3C\text{-}CH{=}O \rightarrow CH_4 + H_3O^+$ + (70)	107.52	
(71)	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + H_2C{=}O + C_3H_7^+ \rightarrow CH_3^+ + C_2H_4 + C_4H_9^+$ + (71)	143.92	143.61
	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + H_3C\text{-}CH{=}O + C_3H_7^+ \rightarrow CH_3^+ + C_3H_6 + C_4H_9^+$ + (71)	143.30	
(72)	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + H_2C{=}O + C_3H_7^+ + CH_3OH \rightarrow CH_3^+ + C_2H_4 + C_4H_9^+ + H_2O$ + (72)	142.24	141.93
	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + H_3C\text{-}CH{=}O + C_3H_7^+ + CH_3OH \rightarrow CH_3^+ + C_3H_6 + C_4H_9^+ + H_2O$ + (72)	141.63	

Table B.4. Part 5 of 11. List of isodesmic reactions with their respective enthalpies of formation.

Species	Reaction	ΔH_f^o	Avg. ΔH_f^o
(73)	$H_3C\text{-}CO\text{-}CH_3 + H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + C_3H_7^+ \rightarrow CH_3^+ + C_3H_6 + C_4H_9^+ + (73)$	121.32	123.06
	$H_3C\text{-}CH{=}O + H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + C_4H_9^+ \rightarrow C_2H_4 + 2\ C_3H_7^+ + (73)$	124.81	
(74)	$H_3C\text{-}CO\text{-}CH_3 + H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ \rightarrow C_3H_6 + H_2C^+\text{-}OH + (74)$	165.29	166.08
	$H_3C\text{-}CH{=}O + C_3H_5^+ + C_4H_9^+ \rightarrow C_2H_4 + C_3H_7^+ + (74)$	166.87	
(75)	$H_3C\text{-}CO\text{-}CH_3 + H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + C_4H_9^+ \rightarrow C_3H_6 + H_2C^+\text{-}OH + C_3H_7^+ + (75)$	149.61	148.95
	$H_3C\text{-}CH{=}O + C_3H_5^+ + C_3H_7^+ \rightarrow CH_3^+ + C_2H_4 + (75)$	148.29	
(76)	$H_3C\text{-}CO\text{-}CH_3 + H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + C_4H_9^+ + C_2H_6 \rightarrow C_3H_6 + H_2C^+\text{-}OH + C_3H_7^+ + CH_4 + (76)$	141.90	141.24
	$H_3C\text{-}CH{=}O + C_3H_5^+ + C_3H_7^+ + C_2H_6 \rightarrow CH_3^+ + C_2H_4 + CH_4 + (76)$	140.58	
(77)	$C_3H_7^+ + H_3C\text{-}CH{=}O \rightarrow CH_4 + (77)$	158.93	158.95
	$C_3H_7^+ + H_3C\text{-}CO\text{-}CH_3 + H_2C{=}O \rightarrow H_3C\text{-}CH{=}O + CH_4 + (77)$	158.96	
(78)	$C_4H_9^+ + H_3C\text{-}CH{=}O \rightarrow CH_4 + (78)$	144.50	144.52
	$C_4H_9^+ + H_3C\text{-}CO\text{-}CH_3 + H_2C{=}O \rightarrow H_3C\text{-}CH{=}O + CH_4 + (78)$	144.54	
(79)	$C_3H_7^+ + H_3C\text{-}CO\text{-}CH_3 \rightarrow CH_4 + (79)$	134.13	134.11
	$C_3H_7^+ + 2\ H_3C\text{-}CH{=}O \rightarrow H_2C{=}O + CH_4 + (79)$	134.09	
(80)	$CH_3O^+H_2 + H_2C{=}CH\text{-}C(CH_3){=}O + H_2C{=}O \rightarrow H_3C\text{-}CH{=}O + CH_4 + (80)$	144.70	144.68
	$CH_3O^+H_2 + H_2C{=}CH\text{-}C(CH_3){=}O + H_3C\text{-}CH{=}O \rightarrow H_3C\text{-}CO\text{-}CH_3 + CH_4 + (80)$	144.67	
(81)	$CH_3\text{-}O^+H\text{-}CH_3 + H_2C{=}CH\text{-}C(CH_3){=}O + H_2C{=}O \rightarrow H_3C\text{-}CH{=}O + CH_4 + (81)$	134.45	134.43
	$CH_3\text{-}O^+H\text{-}CH_3 + H_2C{=}CH\text{-}C(CH_3){=}O + H_3C\text{-}CH{=}O \rightarrow H_3C\text{-}CO\text{-}CH_3 + CH_4 + (81)$	134.42	
(82)	$3\ CH_3O^+H_2 + H_2C{=}CH\text{-}C(CH_3){=}O + H_2C{=}O \rightarrow H_3C\text{-}CH{=}O + CH_4 + 2\ H_3O^+ + (82)$	128.48	127.74
	$2\ CH_3\text{-}O^+H\text{-}CH_3 + H_2C{=}CH\text{-}C(CH_3){=}O + H_2C{=}O \rightarrow CH_3O^+H_2 + H_3C\text{-}CH{=}O + CH_4 + (82)$	127.00	
(83)	$CH_3O^+H_2 + C_3H_6 \rightarrow CH_4 + (83)$	148.22	147.85
	$CH_3\text{-}O^+H\text{-}CH_3 + C_3H_6 + H_3O^+ \rightarrow CH_3O^+H_2 + CH_4 + (83)$	147.48	
(84)	$2\ CH_3O^+H_2 + C_3H_6 \rightarrow CH_4 + H_3O^+ + (84)$	144.44	144.81
	$CH_3\text{-}O^+H\text{-}CH_3 + C_3H_6 \rightarrow CH_4 + (84)$	145.18	
(85)	$3\ CH_3O^+H_2 + C_3H_6 \rightarrow CH_4 + 2\ H_3O^+ + (85)$	142.01	141.27
	$2\ CH_3\text{-}O^+H\text{-}CH_3 + C_3H_6 \rightarrow CH_3O^+H_2 + CH_4 + (85)$	140.53	
(86)	$CH_3O^+H_2 + H_3C\text{-}CH{=}O + C_2H_6 \rightarrow 2\ CH_4 + (86)$	94.51	94.14
	$CH_3\text{-}O^+H\text{-}CH_3 + H_3C\text{-}CH{=}O + C_2H_6 + H_3O^+ \rightarrow CH_3O^+H_2 + 2\ CH_4 + (86)$	93.77	
(87)	$2\ CH_3O^+H_2 + H_3C\text{-}CH{=}O + C_2H_6 \rightarrow 2\ CH_4 + H_3O^+ + (87)$	92.19	91.82
	$CH_3\text{-}O^+H\text{-}CH_3 + H_3C\text{-}CH{=}O + C_2H_6 \rightarrow 2\ CH_4 + (87)$	91.45	
(88)	$3\ CH_3O^+H_2 + H_3C\text{-}CH{=}O + C_2H_6 \rightarrow 2\ CH_4 + 2\ H_3O^+ + (88)$	97.09	96.35
	$2\ CH_3\text{-}O^+H\text{-}CH_3 + H_3C\text{-}CH{=}O + C_2H_6 \rightarrow CH_3O^+H_2 + 2\ CH_4 + (88)$	95.61	
(89)	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + H_2C{=}O \rightarrow C_2H_4 + H_2C^+\text{-}OH + (89)$	193.73	193.42
	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + H_3C\text{-}CH{=}O \rightarrow C_3H_6 + H_2C^+\text{-}OH + (89)$	193.11	
(90)	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + H_2C{=}O + C_2H_6 \rightarrow C_2H_4 + H_2C^+\text{-}OH + CH_4 + (90)$	186.23	185.92
	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + H_3C\text{-}CH{=}O + C_2H_6 \rightarrow C_3H_6 + H_2C^+\text{-}OH + CH_4 + (90)$	185.61	

Table B.4. Part 6 of 11. List of isodesmic reactions with their respective enthalpies of formation.

Species	Reaction	ΔH_f^o	Avg. ΔH_f^o
(91)	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + C_4H_9^+ + H_2C{=}O \rightarrow C_2H_4 + H_2C^+\text{-}OH + C_3H_7^+ + (91)$	174.34	174.03
	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + C_4H_9^+ + H_3C\text{-}CH{=}O \rightarrow C_3H_6 + H_2C^+\text{-}OH + C_3H_7^+ + (91)$	173.72	
(92)	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + C_4H_9^+ + H_2C{=}O + C_2H_6 \rightarrow C_2H_4 + H_2C^+\text{-}OH + C_3H_7^+ + CH_4 + (92)$	165.45	165.14
	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + C_4H_9^+ + H_3C\text{-}CH{=}O + C_2H_6 \rightarrow C_3H_6 + H_2C^+\text{-}OH + C_3H_7^+ + CH_4 + (92)$	164.83	
(93)	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + O{=}CH\text{-}HC{=}O \rightarrow C_2H_4 + H_2C^+\text{-}OH + (93)$	159.96	159.65
	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + O{=}CH\text{-}HC{=}O + H_3C\text{-}CH{=}O \rightarrow C_3H_6 + H_2C^+\text{-}OH + H_2C{=}O + (93)$	159.34	
(94)	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + O{=}CH\text{-}HC{=}O + C_4H_9^+ \rightarrow C_2H_4 + H_2C^+\text{-}OH + C_3H_7^+ + (94)$	143.96	143.65
	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + O{=}CH\text{-}HC{=}O + H_3C\text{-}CH{=}O + C_4H_9^+ \rightarrow C_3H_6 + H_2C^+\text{-}OH + H_2C{=}O + C_3H_7^+ + (94)$	143.34	
(95)	$H_3C\text{-}CH{=}O + H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + O{=}CH\text{-}HC{=}O \rightarrow C_2H_4 + H_2C^+\text{-}OH + H_2C{=}O + (95)$	132.58	131.33
	$H_3C\text{-}CO\text{-}CH_3 + 2\ C_3H_7^+ + C_3H_5^+ + O{=}CH\text{-}HC{=}O \rightarrow CH_3^+ + C_3H_6 + H_2C{=}O + C_4H_9^+ + (95)$	130.09	
(96)	$H_3C\text{-}CH{=}O + C_3H_7^+ + C_2H_6 \rightarrow 2\ CH_4 + (96)$	149.21	149.22
	$H_3C\text{-}CH{=}O + C_3H_7^+ + C_3H_8 \rightarrow CH_4 + C_2H_6 + (96)$	149.24	
(97)	$H_3C\text{-}CH{=}O + C_4H_9^+ + C_2H_6 \rightarrow 2\ CH_4 + (97)$	135.62	135.63
	$H_3C\text{-}CH{=}O + C_4H_9^+ + C_3H_8 \rightarrow CH_4 + C_2H_6 + (97)$	135.65	
(98)	$2\ H_3C\text{-}CH{=}O + C_4H_9^+ + C_2H_6 \rightarrow 2\ CH_4 + H_2C{=}O + (98)$	111.58	111.61
	$H_3C\text{-}CO\text{-}CH_3 + C_4H_9^+ + C_3H_8 \rightarrow CH_4 + C_2H_6 + (98)$	111.64	
(99)	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + H_2C{=}CH\text{-}C(CH_3){=}O + H_2C{=}O \rightarrow C_2H_4 + H_2C^+\text{-}OH + H_3C\text{-}CH{=}O + (99)$	186.19	185.88
	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + H_2C{=}CH\text{-}C(CH_3){=}O \rightarrow C_3H_6 + H_2C^+\text{-}OH + (99)$	185.57	
(100)	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + H_2C{=}CH\text{-}C(CH_3){=}O + H_2C{=}O + C_4H_9^+ \rightarrow C_2H_4 + H_2C^+\text{-}OH + H_3C\text{-}CH{=}O + C_3H_7^+ + (100)$	171.15	170.84
	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + H_2C{=}CH\text{-}C(CH_3){=}O + C_4H_9^+ \rightarrow C_3H_6 + H_2C^+\text{-}OH + C_3H_7^+ + (100)$	170.53	
(101)	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + H_2C{=}CH\text{-}C(CH_3){=}O + H_2C{=}O + C_3H_6 \rightarrow 2\ C_2H_4 + H_2C^+\text{-}OH + H_3C\text{-}CH{=}O + (101)$	173.78	173.47
	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + H_2C{=}CH\text{-}C(CH_3){=}O \rightarrow C_2H_4 + H_2C^+\text{-}OH + (101)$	173.16	
(102)	$H_3C\text{-}HC^+\text{-}OH + 2\ C_3H_5^+ + H_2C{=}O + C_3H_7^+ \rightarrow C_2H_4 + 2\ CH_3^+ + C_4H_9^+ + (102)$	135.40	135.09
	$H_3C\text{-}HC^+\text{-}OH + 2\ C_3H_5^+ + H_3C\text{-}CH{=}O + C_3H_7^+ \rightarrow C_3H_6 + 2\ CH_3^+ + C_4H_9^+ + (102)$	134.78	
(103)	$H_3C\text{-}HC^+\text{-}OH + 2\ C_3H_5^+ + H_2C{=}O + C_3H_7^+ + C_3H_6 \rightarrow 2\ C_2H_4 + 2\ CH_3^+ + C_4H_9^+ + (103)$	122.50	122.19
	$H_3C\text{-}HC^+\text{-}OH + 2\ C_3H_5^+ + H_3C\text{-}CH{=}O + C_3H_7^+ \rightarrow C_2H_4 + 2\ CH_3^+ + C_4H_9^+ + (103)$	121.88	
(104)	$H_3C\text{-}HC^+\text{-}OH + 2\ C_3H_5^+ + C_3H_7^+ + H_3C\text{-}CH{=}O \rightarrow C_2H_4 + 2\ CH_3^+ + C_4H_9^+ + (104)$	116.00	117.16
	$H_3C\text{-}HC^+\text{-}OH + 2\ C_3H_5^+ + C_4H_9^+ + H_3C\text{-}CO\text{-}CH_3 \rightarrow C_3H_6 + CH_3^+ + 2\ C_3H_7^+ + (104)$	118.32	
(105)	$C_3H_7^+ + C_2H_5OH + H_3C\text{-}CH{=}O \rightarrow 2\ CH_4 + (105)$	109.68	109.70
	$C_3H_7^+ + CH_3OH + H_3C\text{-}CH{=}O + C_2H_6 \rightarrow 3\ CH_4 + (105)$	109.72	
(106)	$C_4H_9^+ + C_2H_5OH + H_3C\text{-}CH{=}O \rightarrow 2\ CH_4 + (106)$	97.34	97.36
	$C_4H_9^+ + CH_3OH + H_3C\text{-}CH{=}O + C_2H_6 \rightarrow 3\ CH_4 + (106)$	97.38	
(107)	$C_4H_9^+ + 2\ CH_3OH + H_3C\text{-}CH{=}O + C_2H_6 \rightarrow 3\ CH_4 + H_2O + (107)$	103.31	103.16
	$C_4H_9^+ + H_3C\text{-}O\text{-}CH_3 + H_3C\text{-}CH{=}O + C_2H_6 \rightarrow 3\ CH_4 + (107)$	103.00	
(108)	$C_3H_7^+ + H_3C\text{-}CH{=}O + C_3H_8 \rightarrow 2\ CH_4 + (108)$	139.20	139.18
	$C_3H_7^+ + H_3C\text{-}CH{=}O + 2\ C_2H_6 \rightarrow 3\ CH_4 + (108)$	139.17	

Table B.4. Part 7 of 11. List of isodesmic reactions with their respective enthalpies of formation.

Species	Reaction	ΔH_f^o	Avg. ΔH_f^o
(109)	$C_4H_9^+ + H_3C\text{-}CH{=}O + C_3H_8 \rightarrow 2\ CH_4 + (109)$	126.74	126.73
	$C_4H_9^+ + H_3C\text{-}CH{=}O + 2\ C_2H_6 \rightarrow 3\ CH_4 + (109)$	126.71	
(110)	$C_3H_7^+ + 2\ H_3C\text{-}CH{=}O + C_3H_8 \rightarrow 2\ CH_4 + H_2C{=}O + (110)$	114.47	114.45
	$C_3H_7^+ + 2\ H_3C\text{-}CH{=}O + 2\ C_2H_6 \rightarrow 3\ CH_4 + H_2C{=}O + (110)$	114.44	
(111)	$CH_3O^+H_2 + H_3C\text{-}O\text{-}CH_3 + 2\ C_2H_5OH + C_3H_8 \rightarrow 5\ CH_4 + (111)$	-23.20	-23.15
	$CH_3O^+H_2 + H_3C\text{-}O\text{-}CH_3 + 2\ CH_3OH + 2\ C_3H_8 \rightarrow 6\ CH_4 + (111)$	-23.10	
(112)	$CH_3O^+H_2 + 2\ H_3C\text{-}O\text{-}CH_3 + CH_3OH + 2\ C_3H_8 \rightarrow 6\ CH_4 + (112)$	-20.40	-20.45
	$CH_3O^+H_2 + 2\ H_3C\text{-}O\text{-}CH_3 + C_2H_5OH + 3\ C_2H_6 \rightarrow 7\ CH_4 + (112)$	-20.50	
(113)	$CH_3O^+H_2 + 3\ H_3C\text{-}O\text{-}CH_3 + 2\ C_3H_8 \rightarrow 6\ CH_4 + (113)$	-18.89	-18.90
	$CH_3O^+H_2 + 3\ H_3C\text{-}O\text{-}CH_3 + C_3H_8 + 2\ C_2H_6 \rightarrow 7\ CH_4 + (113)$	-18.92	
(114)	$CH_3O^+H_2 + CH_3OH + C_2H_6 + H_3C\text{-}CH{=}O \rightarrow 3\ CH_4 + (114)$	48.25	48.23
	$CH_3O^+H_2 + C_2H_5OH + H_3C\text{-}CH{=}O \rightarrow 2\ CH_4 + (114)$	48.21	
(115)	$CH_3\text{-}O^+H\text{-}CH_3 + CH_3OH + C_2H_6 + H_3C\text{-}CH{=}O \rightarrow 3\ CH_4 + (115)$	48.70	48.68
	$CH_3\text{-}O^+H\text{-}CH_3 + C_2H_5OH + H_3C\text{-}CH{=}O \rightarrow 2\ CH_4 + (115)$	48.66	
(116)	$CH_3\text{-}O^+H\text{-}CH_3 + CH_3O^+H_2 + CH_3OH + C_2H_6 + H_3C\text{-}CH{=}O \rightarrow 3\ CH_4 + H_3O^+ + (116)$	50.97	50.95
	$CH_3\text{-}O^+H\text{-}CH_3 + CH_3O^+H_2 + C_2H_5OH + H_3C\text{-}CH{=}O \rightarrow 2\ CH_4 + H_3O^+ + (116)$	50.93	
(117)	$C_3H_6 + H_3C\text{-}CH{=}O + CH_3OH \rightarrow 2\ CH_4 + (117)$	-58.40	-58.42
	$C_3H_6 + H_3C\text{-}CH{=}O + C_2H_5OH \rightarrow C_2H_6 + CH_4 + (117)$	-58.44	
(118)	$C_3H_6 + H_3C\text{-}CO\text{-}CH_3 + CH_3OH \rightarrow 2\ CH_4 + (118)$	-69.63	-69.67
	$C_3H_6 + 2\ H_3C\text{-}CH{=}O + C_2H_5OH \rightarrow C_2H_6 + CH_4 + H_2C{=}O + (118)$	-69.70	
(119)	$C_3H_6 + H_3C\text{-}CH{=}O + H_3C\text{-}O\text{-}CH_3 \rightarrow 2\ CH_4 + (119)$	-50.78	-50.62
	$C_3H_6 + H_3C\text{-}CH{=}O + 2\ CH_3OH \rightarrow 2\ CH_4 + H_2O + (119)$	-50.47	
(120)	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + H_2C{=}O + C_3H_7^+ \rightarrow C_2H_4 + H_2C^+\text{-}OH + C_4H_9^+ + (120)$	211.61	211.30
	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + H_3C\text{-}CH{=}O + C_3H_7^+ \rightarrow C_3H_6 + H_2C^+\text{-}OH + C_4H_9^+ + (120)$	210.99	
(121)	$H_3C\text{-}CH{=}O + H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + C_3H_7^+ \rightarrow C_2H_4 + H_2C^+\text{-}OH + C_4H_9^+ + (121)$	181.90	183.06
	$H_3C\text{-}CO\text{-}CH_3 + H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + C_4H_9^+ + CH_3^+ \rightarrow C_3H_6 + H_2C^+\text{-}OH + 2\ C_3H_7^+ + (121)$	184.22	
(122)	$H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + O{=}CH\text{-}OH + C_3H_7^+ \rightarrow C_2H_4 + H_2C^+\text{-}OH + C_4H_9^+ + (122)$	146.53	147.69
	$H_3C\text{-}CO\text{-}CH_3 + H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + O{=}CH\text{-}OH + C_4H_9^+ + CH_3^+ \rightarrow C_3H_6 + H_3C\text{-}CH{=}O + H_2C^+\text{-}OH + 2\ C_3H_7^+ + (122)$	148.85	
(123)	$CH_3O^+H_2 + CH_3OH + H_3C\text{-}CH{=}O \rightarrow 2\ CH_4 + (123)$	65.08	65.06
	$CH_3O^+H_2 + C_2H_5OH + H_3C\text{-}CH{=}O \rightarrow C_2H_6 + CH_4 + (123)$	65.04	
(124)	$2\ CH_3O^+H_2 + CH_3OH + H_3C\text{-}CH{=}O \rightarrow 2\ CH_4 + H_3O^+ + (124)$	63.36	62.97
	$CH_3\text{-}O^+H\text{-}CH_3 + C_2H_5OH + H_3C\text{-}CH{=}O \rightarrow C_2H_6 + CH_4 + (124)$	62.58	
(125)	$3\ CH_3O^+H_2 + CH_3OH + H_3C\text{-}CH{=}O \rightarrow 2\ CH_4 + 2\ H_3O^+ + (125)$	64.68	64.29
	$CH_3\text{-}O^+H\text{-}CH_3 + CH_3O^+H_2 + C_2H_5OH + H_3C\text{-}CH{=}O \rightarrow C_2H_6 + CH_4 + H_3O^+ + (125)$	63.90	
(126)	$H_3C\text{-}HC^+\text{-}OH + 2\ C_3H_5^+ + H_2C{=}O + C_3H_7^+ \rightarrow CH_3^+ + H_2C^+\text{-}OH + C_4H_9^+ + (126)$	205.84	205.53
	$H_3C\text{-}HC^+\text{-}OH + 2\ C_3H_5^+ + H_3C\text{-}CH{=}O + C_3H_7^+ \rightarrow CH_3^+ + H_2C^+\text{-}OH + C_4H_9^+ + (126)$	205.22	

Table B.4. Part 8 of 11. List of isodesmic reactions with their respective enthalpies of formation.

Species	Reaction	ΔH_f^o	Avg. ΔH_f^o
(127)	$H_3C\text{-}CH{=}O + H_3C\text{-}HC^+\text{-}OH + 2\ C_3H_5^+ + C_3H_7^+ \rightarrow CH_3^+ + C_2H_4 + H_2C^+\text{-}OH + C_4H_9^+ + (127)$	184.72	185.87
	$H_3C\text{-}CO\text{-}CH_3 + H_3C\text{-}HC^+\text{-}OH + 2\ C_3H_5^+ + C_4H_9^+ \rightarrow C_3H_6 + H_2C^+\text{-}OH + 2\ C_3H_7^+ + (127)$	187.03	
(128)	$H_3C\text{-}HC^+\text{-}OH + 2\ C_3H_5^+ + H_2C{=}O + C_3H_7^+ + C_3H_6 \rightarrow CH_3^+ + 2\ C_2H_4 + H_2C^+\text{-}OH + C_4H_9^+ + (128)$	193.64	193.33
	$H_3C\text{-}HC^+\text{-}OH + 2\ C_3H_5^+ + H_3C\text{-}CH{=}O + C_3H_7^+ \rightarrow CH_3^+ + H_2C^+\text{-}OH + C_4H_9^+ + C_2H_4 + (128)$	193.02	
(129)	$C_3H_6 + H_3C\text{-}CH{=}O + C_2H_6 \rightarrow 2\ CH_4 + (129)$	-25.86	-25.85
	$C_3H_6 + H_3C\text{-}CH{=}O + C_3H_8 \rightarrow C_2H_6 + CH_4 + (129)$	-25.83	
(130)	$2\ C_3H_6 + H_3C\text{-}CH{=}O + C_2H_6 \rightarrow C_2H_4 + 2\ CH_4 + (130)$	-34.24	-34.23
	$2\ C_3H_6 + H_3C\text{-}CH{=}O + C_3H_8 \rightarrow C_2H_4 + C_2H_6 + CH_4 + (130)$	-34.21	
(131)	$C_3H_6 + 2\ H_3C\text{-}CH{=}O + C_2H_6 \rightarrow H_2C{=}O + 2\ CH_4 + (131)$	-37.65	-37.63
	$C_3H_6 + 2\ H_3C\text{-}CH{=}O + C_3H_8 \rightarrow H_2C{=}O + C_2H_6 + CH_4 + (131)$	-37.62	
(132)	$2\ C_3H_5^+ + H_2C^+\text{-}OH \rightarrow 2\ CH_3^+ + (132)$	161.33	160.83
	$2\ C_3H_5^+ + H_3C\text{-}HC^+\text{-}OH + C_3H_7^+ \rightarrow 2\ CH_3^+ + C_4H_9^+ + (132)$	160.33	
(133)	$2\ C_3H_5^+ + H_2C^+\text{-}OH + CH_3OH \rightarrow 2\ CH_3^+ + H_2O + (133)$	161.88	161.39
	$2\ C_3H_5^+ + H_3C\text{-}HC^+\text{-}OH + C_3H_7^+ + CH_3OH \rightarrow 2\ CH_3^+ + C_4H_9^+ + H_2O + (133)$	160.89	
(134)	$2\ C_3H_5^+ + H_2C^+\text{-}OH + C_3H_6 \rightarrow C_2H_4 + 2\ CH_3^+ + (134)$	151.57	151.08
	$2\ C_3H_5^+ + H_3C\text{-}HC^+\text{-}OH + C_3H_7^+ + C_3H_6 \rightarrow C_2H_4 + 2\ CH_3^+ + C_4H_9^+ + (134)$	150.58	
(135)	$H_3C\text{-}HC^+\text{-}OH + 2\ C_3H_5^+ + H_2C{=}O \rightarrow CH_3^+ + C_2H_4 + H_2C^+\text{-}OH + (135)$	186.26	185.95
	$H_3C\text{-}HC^+\text{-}OH + 2\ C_3H_5^+ + H_3C\text{-}CH{=}O \rightarrow CH_3^+ + C_3H_6 + H_2C^+\text{-}OH + (135)$	185.64	
(136)	$H_3C\text{-}CH{=}O + H_3C\text{-}HC^+\text{-}OH + 2\ C_3H_5^+ \rightarrow CH_3^+ + C_2H_4 + H_2C^+\text{-}OH + (136)$	167.89	168.10
	$H_3C\text{-}CO\text{-}CH_3 + 2\ C_3H_5^+ + C_4H_9^+ \rightarrow CH_3^+ + C_3H_6 + C_3H_7^+ + (136)$	168.30	
(137)	$H_3C\text{-}HC^+\text{-}OH + 2\ C_3H_5^+ + H_2C{=}O + C_3H_6 \rightarrow CH_3^+ + 2\ C_2H_4 + H_2C^+\text{-}OH + (137)$	174.89	174.59
	$H_3C\text{-}HC^+\text{-}OH + 2\ C_3H_5^+ + H_3C\text{-}CH{=}O \rightarrow CH_3^+ + C_2H_4 + H_2C^+\text{-}OH + (137)$	174.28	
(138)	$CH_3O^+H_2 + C_3H_8 + H_3C\text{-}CH{=}O \rightarrow 2\ CH_4 + (138)$	80.51	80.50
	$CH_3O^+H_2 + 2\ C_2H_6 + H_3C\text{-}CH{=}O \rightarrow 3\ CH_4 + (138)$	80.49	
(139)	$CH_3\text{-}O^+H\text{-}CH_3 + C_3H_8 + H_3C\text{-}CH{=}O \rightarrow 2\ CH_4 + (139)$	79.34	79.33
	$CH_3\text{-}O^+H\text{-}CH_3 + 2\ C_2H_6 + H_3C\text{-}CH{=}O \rightarrow 3\ CH_4 + (139)$	79.31	
(140)	$CH_3\text{-}O^+H\text{-}CH_3 + CH_3O^+H_2 + C_3H_8 + H_3C\text{-}CH{=}O \rightarrow 2\ CH_4 + H_3O^+ + (140)$	86.34	86.32
	$CH_3\text{-}O^+H\text{-}CH_3 + CH_3O^+H_2 + 2\ C_2H_6 + H_3C\text{-}CH{=}O \rightarrow 3\ CH_4 + H_3O^+ + (140)$	86.31	
(141)	$C_3H_6 + H_3C\text{-}CH{=}O + C_2H_6 + CH_3OH \rightarrow 3\ CH_4 + (141)$	-68.71	-68.73
	$C_3H_6 + H_3C\text{-}CH{=}O + C_2H_5OH \rightarrow 2\ CH_4 + (141)$	-68.75	
(142)	$C_3H_6 + H_3C\text{-}CH{=}O + C_2H_6 + 2\ CH_3OH \rightarrow 3\ CH_4 + H_2O + (142)$	-60.06	-60.22
	$C_3H_6 + H_3C\text{-}CH{=}O + H_3C\text{-}O\text{-}CH_3 + C_2H_6 \rightarrow 3\ CH_4 + (142)$	-60.37	
(143)	$C_3H_6 + 2\ H_3C\text{-}CH{=}O + C_2H_6 + CH_3OH \rightarrow H_2C{=}O + 3\ CH_4 + (143)$	-81.75	-81.75
	$C_3H_6 + H_3C\text{-}CO\text{-}CH_3 + C_2H_5OH \rightarrow 2\ CH_4 + (143)$	-81.75	
(144)	$CH_3O^+H_2 + H_3C\text{-}CH{=}O + 3\ C_3H_6 + C_2H_5OH + CH_3OH + H_3C\text{-}O\text{-}CH_3 \rightarrow 2\ C_2H_4 + 6\ CH_4 + (144)$	-25.83	-25.65
	$CH_3O^+H_2 + H_3C\text{-}CH{=}O + 3\ C_3H_6 + 4\ CH_3OH + C_2H_6 \rightarrow 2\ C_2H_4 + 7\ CH_4 + (144)$	-25.48	

Table B.4. Part 9 of 11. List of isodesmic reactions with their respective enthalpies of formation.

Species	Reaction	ΔH_f^o	Avg. ΔH_f^o
(145)	$CH_3O^+H_2 + H_3C\text{-}CH{=}O + 3\ C_3H_6 + C_2H_5OH + 2\ H_3C\text{-}O\text{-}CH_3 \rightarrow 2\ C_2H_4 + 6\ CH_4 + (145)$	-24.81	-24.48
	$CH_3O^+H_2 + H_3C\text{-}CH{=}O + 3\ C_3H_6 + 5\ CH_3OH + C_2H_6 \rightarrow 2\ C_2H_4 + 7\ CH_4 + 2\ H_2O + (145)$	-24.14	
(146)	$CH_3O^+H_2 + H_3C\text{-}CH{=}O + 3\ C_3H_6 + CH_3OH + 2\ H_3C\text{-}O\text{-}CH_3 + C_2H_6 \rightarrow 2\ C_2H_4 + 7\ CH_4 + (146)$	-24.19	-23.87
	$CH_3O^+H_2 + H_3C\text{-}CH{=}O + 3\ C_3H_6 + 5\ CH_3OH + C_2H_6 \rightarrow 2\ C_2H_4 + 7\ CH_4 + 2\ H_2O + (146)$	-23.56	
(147)	$H_2C{=}CH\text{-}C(CH_3){=}O + C_4H_9^+ + C_3H_5^+ + H_2C{=}O \rightarrow C_3H_7^+ + H_3C\text{-}CH{=}O + C_2H_4 + (147)$	185.70	185.68
	$H_2C{=}CH\text{-}C(CH_3){=}O + C_4H_9^+ + C_3H_5^+ + H_3C\text{-}CH{=}O \rightarrow C_3H_7^+ + H_3C\text{-}CO\text{-}CH_3 + C_2H_4 + (147)$	185.66	
(148)	$H_2C{=}CH\text{-}C(CH_3){=}O + C_3H_7^+ + C_3H_5^+ + H_2C{=}O \rightarrow H_3C\text{-}CH{=}O + C_2H_4 + CH_3^+ + (148)$	166.08	166.06
	$H_2C{=}CH\text{-}C(CH_3){=}O + C_3H_7^+ + C_3H_5^+ + H_3C\text{-}CH{=}O \rightarrow H_3C\text{-}CO\text{-}CH_3 + C_2H_4 + CH_3^+ + (148)$	166.04	
(149)	$H_2C{=}CH\text{-}C(CH_3){=}O + C_4H_9^+ + C_3H_5^+ \rightarrow C_3H_7^+ + C_2H_4 + (149)$	169.75	168.30
	$H_2C{=}CH\text{-}C(CH_3){=}O + 2\ C_3H_7^+ + C_3H_5^+ \rightarrow C_4H_9^+ + C_2H_4 + CH_3^+ + (149)$	166.85	
(150)	$CH_3O^+H_2 + C_3H_8 \rightarrow CH_4 + (150)$	110.31	110.29
	$CH_3O^+H_2 + 2\ C_2H_6 \rightarrow 2\ CH_4 + (150)$	110.28	
(151)	$CH_3\text{-}O^+H\text{-}CH_3 + C_3H_8 \rightarrow CH_4 + (151)$	107.87	107.85
	$CH_3\text{-}O^+H\text{-}CH_3 + 2\ C_2H_6 \rightarrow 2\ CH_4 + (151)$	107.84	
(152)	$CH_3\text{-}O^+H\text{-}CH_3 + CH_3O^+H_2 + C_3H_8 \rightarrow CH_4 + H_3O^+ + (152)$	106.85	106.83
	$CH_3\text{-}O^+H\text{-}CH_3 + CH_3O^+H_2 + 2\ C_2H_6 \rightarrow 2\ CH_4 + H_3O^+ + (152)$	106.82	
(153)	$CH_3O^+H_2 + C_2H_6 \rightarrow CH_4 + (153)$	124.05	124.07
	$CH_3O^+H_2 + C_3H_8 \rightarrow C_2H_6 + (153)$	124.08	
(154)	$2\ CH_3O^+H_2 + C_2H_6 \rightarrow CH_4 + H_3O^+ + (154)$	120.67	120.31
	$CH_3\text{-}O^+H\text{-}CH_3 + C_3H_8 \rightarrow C_2H_6 + (154)$	119.96	
(155)	$3\ CH_3O^+H_2 + C_2H_6 \rightarrow CH_4 + 2\ H_3O^+ + (155)$	118.65	118.30
	$CH_3\text{-}O^+H\text{-}CH_3 + CH_3O^+H_2 + C_3H_8 \rightarrow C_2H_6 + H_3O^+ + (155)$	117.94	
(156)	$CH_3O^+H_2 + C_3H_6 + C_2H_6 \rightarrow 2\ CH_4 + (156)$	135.29	135.30
	$CH_3O^+H_2 + C_3H_6 + C_3H_8 \rightarrow C_2H_6 + CH_4 + (156)$	135.32	
(157)	$2\ CH_3O^+H_2 + C_3H_6 + C_2H_6 \rightarrow 2\ CH_4 + H_3O^+ + (157)$	133.46	133.11
	$CH_3\text{-}O^+H\text{-}CH_3 + C_3H_6 + C_3H_8 \rightarrow C_2H_6 + CH_4 + (157)$	132.75	
(158)	$3\ CH_3O^+H_2 + C_3H_6 + C_2H_6 \rightarrow 2\ CH_4 + 2\ H_3O^+ + (158)$	132.49	132.14
	$CH_3\text{-}O^+H\text{-}CH_3 + CH_3O^+H_2 + C_3H_6 + C_3H_8 \rightarrow C_2H_6 + CH_4 + H_3O^+ + (158)$	131.78	
(159)	$CH_3O^+H_2 + C_2H_4 \rightarrow CH_4 + (159)$	163.80	163.43
	$CH_3\text{-}O^+H\text{-}CH_3 + C_2H_4 + H_3O^+ \rightarrow CH_4 + CH_3O^+H_2 + (159)$	163.06	
(160)	$2\ CH_3O^+H_2 + C_2H_4 \rightarrow CH_4 + H_3O^+ + (160)$	159.39	159.02
	$CH_3\text{-}O^+H\text{-}CH_3 + C_2H_4 \rightarrow CH_4 + (160)$	158.65	
(161)	$3\ CH_3O^+H_2 + C_2H_4 \rightarrow CH_4 + 2\ H_3O^+ + (161)$	155.02	154.65
	$CH_3\text{-}O^+H\text{-}CH_3 + CH_3O^+H_2 + C_2H_4 \rightarrow CH_4 + H_3O^+ + (161)$	154.28	
(162)	$CH_3O^+H_2 + 2\ C_3H_6 \rightarrow C_2H_4 + CH_4 + (162)$	137.93	137.56
	$CH_3\text{-}O^+H\text{-}CH_3 + 2\ C_3H_6 + H_3O^+ \rightarrow CH_3O^+H_2 + C_2H_4 + CH_4 + (162)$	137.19	

Table B.4. Part 10 of 11. List of isodesmic reactions with their respective enthalpies of formation.

Species	Reaction	ΔH_f^o	Avg. ΔH_f^o
(163)	$CH_3O^+H_2 + 2\,C_3H_6 + C_4H_6 \rightarrow 2\,C_2H_4 + 2\,CH_4 + (163)$	177.00	176.63
	$CH_3\text{-}O^+H\text{-}CH_3 + 2\,C_3H_6 + C_4H_6 + H_3O^+ \rightarrow CH_3O^+H_2 + 2\,C_2H_4 + 2\,CH_4 + (163)$	176.26	
(164)	$2\,CH_3O^+H_2 + 2\,C_3H_6 \rightarrow 2\,CH_4 + H_3O^+ + (164)$	159.00	158.63
	$CH_3\text{-}O^+H\text{-}CH_3 + 2\,C_3H_6 \rightarrow 2\,CH_4 + (164)$	158.26	
(165)	$3\,CH_3O^+H_2 + 2\,C_3H_6 \rightarrow 2\,CH_4 + 2\,H_3O^+ + (165)$	158.66	158.29
	$CH_3\text{-}O^+H\text{-}CH_3 + CH_3O^+H_2 + 2\,C_3H_6 \rightarrow 2\,CH_4 + H_3O^+ + (165)$	157.92	
(166)	$H_3C\text{-}CO\text{-}CH_3 + H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + O{=}CH\text{-}OH \rightarrow C_3H_6 + H_3C\text{-}CH{=}O + CH_3O^+H_2 + (166)$	128.47	129.26
	$C_3H_5^+ + O{=}CH\text{-}OH + C_4H_9^+ \rightarrow C_2H_4 + C_3H_7^+ + (166)$	130.05	
(167)	$H_3C\text{-}CO\text{-}CH_3 + H_3C\text{-}HC^+\text{-}OH + C_3H_5^+ + O{=}CH\text{-}OH + C_4H_9^+ \rightarrow C_3H_6 + H_3C\text{-}CH{=}O + CH_3O^+H_2 + C_3H_7^+ + (167)$	108.31	107.65
	$C_3H_5^+ + O{=}CH\text{-}OH + C_3H_7^+ \rightarrow C_2H_4 + CH_3^+ + (167)$	106.98	
(168)	$3\,CH_3O^+H_2 + C_3H_6 \rightarrow CH_4 + 2\,H_3O^+ + (168)$	144.21	143.84
	$CH_3\text{-}O^+H\text{-}CH_3 + CH_3O^+H_2 + C_3H_6 \rightarrow CH_4 + H_3O^+ + (168)$	143.47	
(169)	$H_2C^+\text{-}OH + H_3C\text{-}HC^+\text{-}OH \rightarrow CH_3^+ + (169)$	74.58	75.08
	$2\,H_2C^+\text{-}OH + C_4H_9^+ \rightarrow C_3H_7^+ + CH_3^+ + (169)$	75.58	
(170)	$H_2C^+\text{-}OH + H_3C\text{-}HC^+\text{-}OH + CH_3OH \rightarrow CH_3^+ + H_2O + (170)$	72.87	73.37
	$2\,H_2C^+\text{-}OH + C_4H_9^+ + CH_3OH \rightarrow C_3H_7^+ + CH_3^+ + H_2O + (170)$	73.86	
(171)	$H_2C^+\text{-}OH + H_3C\text{-}HC^+\text{-}OH + H_3C\text{-}O\text{-}CH_3 \rightarrow CH_3^+ + H_2O + (171)$	71.16	71.66
	$2\,H_2C^+\text{-}OH + C_4H_9^+ + H_3C\text{-}O\text{-}CH_3 \rightarrow C_3H_7^+ + CH_3^+ + H_2O + (171)$	72.16	
(172)	$CH_3O^+H_2 + H_3C\text{-}CH{=}O + C_2H_5OH \rightarrow 2\,CH_4 + (172)$	33.14	32.92
	$CH_3O^+H_2 + H_3C\text{-}CO\text{-}CH_3 + CH_3OH \rightarrow 2\,CH_4 + (172)$	32.71	
(173)	$CH_3O^+H_2 + H_3C\text{-}CH{=}O \rightarrow CH_4 + (173)$	85.25	85.01
	$CH_3O^+H_2 + H_3C\text{-}CO\text{-}CH_3 \rightarrow C_2H_6 + (173)$	84.78	
(174)	$CH_3O^+H_2 + H_3C\text{-}CH{=}O + C_3H_8 \rightarrow CH_4 + C_2H_6 + (174)$	76.61	76.36
	$CH_3O^+H_2 + H_3C\text{-}CO\text{-}CH_3 \rightarrow CH_4 + (174)$	76.11	
(175)	$CH_3O^+H_2 + H_3C\text{-}CH{=}O + C_3H_8 \rightarrow 2\,CH_4 + (175)$	68.02	68.27
	$CH_3O^+H_2 + H_2C{=}O + C_3H_8 + C_2H_6 \rightarrow 3\,CH_4 + (175)$	68.52	
(176)	$CH_3O^+H_2 + H_3C\text{-}CH{=}O + C_3H_8 \rightarrow 2\,CH_4 + (176)$	69.67	69.93
	$CH_3O^+H_2 + H_2C{=}O + C_3H_8 + C_2H_6 \rightarrow 3\,CH_4 + (176)$	70.18	
(177)	$CH_3O^+H_2 + H_3C\text{-}CH{=}O + 2\,C_3H_8 \rightarrow 2\,CH_4 + C_2H_6 + (177)$	60.74	61.01
	$CH_3O^+H_2 + H_2C{=}O + 3\,C_3H_8 \rightarrow 2\,CH_4 + 2\,C_2H_6 + (177)$	61.27	
(178)	$CH_3O^+H_2 + H_2C{=}CH\text{-}OH \rightarrow CH_4 + (178)$	118.04	118.66
	$CH_3O^+H_2 + C_3H_6 + CH_3OH \rightarrow CH_4 + C_2H_6 + (178)$	119.29	
(179)	$CH_3\text{-}O^+H\text{-}CH_3 + H_2C{=}CH\text{-}OH \rightarrow CH_4 + (179)$	112.26	112.88
	$CH_3\text{-}O^+H\text{-}CH_3 + C_3H_6 + CH_3OH \rightarrow CH_4 + C_2H_6 + (179)$	113.51	
(180)	$CH_3\text{-}O^+H\text{-}CH_3 + CH_3O^+H_2 + H_2C{=}CH\text{-}OH \rightarrow CH_4 + H_3O^+ + (180)$	107.63	108.00
	$3\,CH_3O^+H_2 + H_2C{=}CH\text{-}OH \rightarrow CH_4 + 2\,H_3O^+ + (180)$	108.37	

Table B.4. Part 11 of 11. List of isodesmic reactions with their respective enthalpies of formation.

Species	Reaction	ΔH_f^o	Avg. ΔH_f^o
(181)	$CH_3O^+H_2 + 2\,CH_3OH + C_2H_6 \rightarrow 3\,CH_4 + (181)$	7.40	7.38
	$CH_3O^+H_2 + C_2H_5OH + CH_3OH \rightarrow 2\,CH_4 + (181)$	7.36	
(182)	$CH_3\text{-}O^+H\text{-}CH_3 + 2\,CH_3OH + C_2H_6 \rightarrow 3\,CH_4 + (182)$	15.08	15.06
	$CH_3\text{-}O^+H\text{-}CH_3 + C_2H_5OH + CH_3OH \rightarrow 2\,CH_4 + (182)$	15.05	
(183)	$CH_3\text{-}O^+H\text{-}CH_3 + CH_3O^+H_2 + 2\,CH_3OH + C_2H_6 \rightarrow 3\,CH_4 + H_3O^+ + (183)$	18.63	18.98
	$3\,CH_3O^+H_2 + C_2H_5OH + CH_3OH \rightarrow 2\,CH_4 + 2\,H_3O^+ + (183)$	19.33	
(184)	$CH_3O^+H_2 + 3\,C_2H_6 \rightarrow 3\,CH_4 + (184)$	96.97	96.99
	$CH_3O^+H_2 + C_3H_8 + C_2H_6 \rightarrow 2\,CH_4 + (184)$	97.00	
(185)	$CH_3\text{-}O^+H\text{-}CH_3 + 3\,C_2H_6 \rightarrow 3\,CH_4 + (185)$	95.94	95.95
	$CH_3\text{-}O^+H\text{-}CH_3 + C_3H_8 + C_2H_6 \rightarrow 2\,CH_4 + (185)$	95.97	
(186)	$CH_3\text{-}O^+H\text{-}CH_3 + CH_3O^+H_2 + 3\,C_2H_6 \rightarrow 3\,CH_4 + H_3O^+ + (186)$	97.00	97.38
	$3\,CH_3O^+H_2 + C_3H_8 + C_2H_6 \rightarrow 2\,CH_4 + 2\,H_3O^+ + (186)$	97.76	
(187)	$C_2H_4 + H_3C\text{-}CH{=}O + 2\,CH_3OH + C_2H_6 \rightarrow 4\,CH_4 + (187)$	-104.15	-103.93
	$C_2H_4 + H_2C{=}O + 2\,C_2H_5OH \rightarrow 3\,CH_4 + (187)$	-103.72	
(188)	$H_3C\text{-}CO\text{-}CH_3 + H_3C\text{-}O\text{-}CH_3 + H_2C{=}CH\text{-}OH \rightarrow 2\,CH_4 + (188)$	-100.41	-99.63
	$C_2H_4 + H_3C\text{-}CH{=}O + H_3C\text{-}O\text{-}CH_3 + C_2H_5OH \rightarrow 3\,CH_4 + (188)$	-98.84	
(189)	$H_2C{=}CH\text{-}C(CH_3){=}O + 2\,H_3C\text{-}O\text{-}CH_3 \rightarrow 2\,CH_4 + (189)$	-92.29	-92.17
	$C_2H_4 + H_3C\text{-}CO\text{-}CH_3 + 2\,H_3C\text{-}O\text{-}CH_3 \rightarrow 3\,CH_4 + (189)$	-92.05	
(190)	$H_3C\text{-}HC^+\text{-}OH + C_2H_4 + 2\,CH_3OH + C_2H_6 \rightarrow 4\,CH_4 + (190)$	83.18	83.14
	$H_3C\text{-}HC^+\text{-}OH + C_2H_4 + 2\,C_2H_5OH \rightarrow 2\,CH_4 + C_2H_6 + (190)$	83.10	
(191)	$2\,H_3C\text{-}HC^+\text{-}OH + C_2H_4 + 2\,CH_3OH + C_2H_6 \rightarrow 4\,CH_4 + H_2C^+\text{-}OH + (191)$	53.42	53.38
	$2\,H_3C\text{-}HC^+\text{-}OH + C_2H_4 + 2\,C_2H_5OH \rightarrow 2\,CH_4 + C_2H_6 + H_2C^+\text{-}OH + (191)$	53.34	
(192)	$H_3C\text{-}HC^+\text{-}OH + H_2C^+\text{-}OH + C_2H_4 + 2\,CH_3OH + C_2H_6 \rightarrow 4\,CH_4 + CH_3^+ + (192)$	11.49	11.45
	$H_3C\text{-}HC^+\text{-}OH + H_2C^+\text{-}OH + C_2H_4 + 2\,C_2H_5OH \rightarrow 2\,CH_4 + C_2H_6 + CH_3^+ + (192)$	11.41	
(193)	$H_3C\text{-}HC^+\text{-}OH + H_2C^+\text{-}OH + C_2H_4 \rightarrow CH_4 + CH_3^+ + (193)$	97.18	97.24
	$H_3C\text{-}HC^+\text{-}OH + H_2C^+\text{-}OH + C_3H_6 \rightarrow C_2H_6 + CH_3^+ + (193)$	97.30	
(194)	$H_3C\text{-}HC^+\text{-}OH + H_2C^+\text{-}OH + C_2H_4 + CH_3OH \rightarrow CH_4 + CH_3^+ + H_2O + (194)$	96.64	96.63
	$H_3C\text{-}HC^+\text{-}OH + H_2C^+\text{-}OH + C_2H_4 + C_2H_5OH \rightarrow C_2H_6 + CH_3^+ + H_2O + (194)$	96.61	
(195)	$H_3C\text{-}HC^+\text{-}OH + H_2C^+\text{-}OH + C_2H_4 + H_3C\text{-}O\text{-}CH_3 \rightarrow CH_4 + CH_3^+ + H_2O + (195)$	95.89	95.94
	$H_3C\text{-}HC^+\text{-}OH + H_2C^+\text{-}OH + C_3H_6 + H_3C\text{-}O\text{-}CH_3 \rightarrow C_2H_6 + CH_3^+ + H_2O + (195)$	96.00	

Table B.5. Proton affinity calculations to calculate experimental enthalpies of formation for species (Species MH^+). PA = proton affinity from Hunter and Lias.[117] Enthalpy of formation of a proton from Chase.[142] All values are in kcal/mol. Experimental uncertainties are listed where available.

M			H^+ →		MH^+
Species M	**ΔH_f^o**	**PA**	**ΔH_f^o**	**Species MH^+**	**ΔH_f^o**
acetone	-52.23 ± 0.14 [205]	194.1 ± 1.90	365.690	(1)	119.36 ± 2.04
2-methoxypropene	-35.5 [132], *a*	213.9 ± 1.90	365.690	(2)	116.29 ± 1.90
2-butanone	-57.02 ± 0.20 [206]	197.7 ± 1.90	365.690	(3)	110.97 ± 2.10
3-buten-2-one	-27.49 ± 2.63[139]	199.5 ± 1.90	365.690	(26)	138.69 ± 4.53
isopropyl alcohol	-65.2 ± 0.1[207]	189.5 ± 1.90	365.690	(150)	110.99 ± 2.00
2-methoxypropane	-60.2 ± 0.2[207]	197.5 ± 1.90	365.690	(151)	107.99 ± 2.10
ethanol	-56.1 ± 0.1[133]	185.6 ± 1.90	365.690	(153)	123.99 ± 2.00
methoxyethane	-51.73 ± 0.16 [134]	193.3 ± 1.90	365.690	(154)	120.66 ± 2.06

[a] Experimental uncertainty not reported in the literature.

Table B.6. Part 1 of 2. List of fixed groups with enthalpy of formation group additivity values, in kcal/mol, from literature. For some groups, an analogous group was proposed, and the enthalpy of formation for these groups is that of the analogous group.

Group	Analogous Group	Analogy Proposed By	ΔH_f^o Lit. Value	Value Regressed By
CO – (C)(O^+)	CO – (C)(O)	this work	-34.86	Khan et al.[109]
CO – (C^+)(C_d)	CO – (C)(C_d)	this work	-32.71	Khan et al.[109]
C_d – (C)(O^+)	C_d – (C)(O)	this work	8.94	Khan et al.[109]
O – (C^+)(C_d)	O – (C)(C_d)	this work	-30.36	da Silva/Bozzelli[144]
C – $(C)_2$(O)(O^+)	C – $(C)_2(O)_2$	this work	-16.2	Cohen[108]
C – (C_d)(H)(O)(O^+)	C – (C_d)(H)$(O)_2$	this work	-0.3	Cohen[108]
C – $(H)_3$(O^+)	C – $(H)_3$(O)	this work	-10	Cohen[108]
C – (C^+)$(H)_3$	C – (C)$(H)_3$	Bjorkman et al.[123]	-10.25	Sabbe et al.[99]
C – (C_d)$(H)_3$	C – (C)$(H)_3$	Benson[107]	-10.25	Sabbe et al.[99]
C – (C)(C^+)$(H)_2$	C – $(C)_2(H)_2$	Bjorkman et al.[123]	-4.90	Sabbe et al.[99]
C_d – (C^+)(C)	C_d – (C_d)(C)	Bjorkman et al.[123]	9.56	Sabbe et al.[99]
C_d – (C^+)(H)	C_d – (C_d)(H)	Bjorkman et al.[123]	7.27	Sabbe et al.[99]
C – $(C)_2$(CO)(H)	–	–	-1.67	Cohen[108]
C – (C)(H)$(O)_2$	–	–	-15.8	Cohen[108]
C – (C)$(H)_3$	–	–	-10.25	Sabbe et al.[99]
C – (C)(C_d)$(H)_2$	–	–	-4.52	Sabbe et al.[99]
C – (C)$(H)_2$(O)	–	–	-7.92	Sumathi/Green[146]
C – (C_d)$(H)_2$(O)	–	–	-6.9	Cohen[108]
C – (C_d)(H)$(O)_2$	–	–	-0.3	Cohen[108]
C – (C)(CO)$(H)_2$	–	–	-5.26	da Silva/Bozzelli[144]
C – (CO)$(H)_3$	–	–	-10.31	da Silva/Bozzelli[144]
C – $(C)_2(H)_2$	–	–	-4.90	Sabbe et al.[99]
C – $(C)_2$(H)(O)	–	–	-7.42	Sumathi/Green[146]
C – $(H)_3$(O)	–	–	-10	Cohen[108]
C^+ – $(C)_3$	–	–	194.64	Bjorkman et al.[123]
C^+ – $(C)_2$(C_d)	–	–	187.43	Bjorkman et al.[123]
C^+ – (C)(C_d)(H)	–	–	187.61	Bjorkman et al.[123]
C^+ – $(C)_2$(H)	–	–	205.10	Bjorkman et al.[123]

Table B.6. Part 2 of 2. List of fixed groups with enthalpy of formation group additivity values, in kcal/mol, from literature.

Group	Analogous Group	Analogy Proposed By	ΔH_f^o Lit. Value	Value Regressed By
$C_d - (C)_2$	–	–	10.83	Sabbe et al.[99]
$C_d - (C)(H)$	–	–	8.87	Sabbe et al.[99]
$C_d - (C)(O)$	–	–	8.94	Khan et al.[109]
$C_d - (C)(CO)$	–	–	11.02	Khan et al.[109]
$C_d - (C_d)(H)$	–	–	7.27	Sabbe et al.[99]
$C_d - (C_d)(O)$	–	–	9.5	Cohen[108]
$C_d - (CO)(H)$	–	–	10.14	Khan et al.[109]
$C_d - (H)_2$	–	–	6.0	Sabbe et al.[99]
$C_d - (H)(O)$	–	–	8.6	Cohen[108]
$CO - (C)_2$	–	–	-31.69	Cohen[108]
$CO - (C_d)(C)$	–	–	-32.71	Khan et al.[109]
$CO - (C)(H)$	–	–	-29.47	da Silva/Bozzelli[144]
$CO - (C_d)(H)$	–	–	-32.15	Khan et al.[109]
$CO - (C)(CO)$	–	–	-29.88	Khan et al.[109]
$CO - (CO)(H)$	–	–	-25.19	Cohen[108]
$O - (C)_2$	–	–	-23.18	Sumathi/Green[146]
$O - (C)(C_d)$	–	–	-30.36	da Silva/Bozzelli[144]
$O - (C)(H)$	–	–	-37.86	Sumathi/Green[146]
$O - (C_d)_2$	–	–	-33	Cohen[108]
$O - (C_d)(H)$	–	–	-44.45	da Silva/Bozzelli[144]
$O - (CO)(H)$	–	–	-57.79	Cohen[108]
$C^+ - C - C_d$	–	–	-7.02	Bjorkman et al.[123]
Allylic *cis*	–	–	1.25	Bjorkman et al.[123]
Cyclopentadiene RSC	–	–	5.02	Sabbe et al.[99]
Alkane *gauche*	–	–	0.69	Sabbe et al.[99]
THF RSC	–	–	5.9	Cohen[108]
2,5-DHF RSC	–	–	4.2	Cohen[108]

Abbreviations: RSC = ring-strain correction; THF = tetrahydrofuran; 2,5-DHF = 2,5-dihydrofuran

Table B.7. Part 1 of 4. Enthalpies of formation from G4 and isodesmic reactions compared to those obtained from group additivity (GA) values. All values are in kcal/mol.

Species	ΔH_f^o from G4	ΔH_f^o from GA	Abs Dev	Species	ΔH_f^o from G4	ΔH_f^o from GA	Abs Dev
(1)	119.43	118.20	1.23	(26)	140.40	138.55	1.85
(2)	117.19	116.87	0.32	(27)	139.79	137.22	2.56
(3)	112.13	113.30	1.17	(28)	125.43	129.84	4.41
(4)	151.73	152.62	0.89	(29)	159.75	154.67	5.07
(5)	134.23	131.91	2.32	(30)	144.03	145.49	1.47
(6)	135.15	136.59	1.44	(31)	145.98	149.58	3.60
(7)	124.10	121.08	3.01	(32)	146.32	143.93	2.39
(8)	121.41	125.07	3.66	(33)	148.15	148.61	0.46
(9)	116.84	116.18	0.65	(34)	138.40	140.32	1.93
(10)	137.08	138.76	1.67	(35)	157.94	155.04	2.90
(11)	81.44	80.86	0.58	(36)	154.27	153.71	0.56
(12)	85.21	86.23	1.02	(37)	142.29	145.75	3.46
(13)	91.35	90.91	0.44	(38)	150.37	145.43	4.94
(14)	-88.65	-88.35	0.30	(39)	134.90	137.14	2.24
(15)	-82.22	-82.98	0.76	(40)	126.16	128.85	2.70
(16)	-76.69	-76.23	0.46	(41)	-22.85	-22.64	0.21
(17)	110.57	108.72	1.85	(42)	-16.20	-16.58	0.38
(18)	105.10	104.51	0.58	(43)	-30.41	-30.24	0.17
(19)	110.97	113.40	2.43	(44)	-13.10	-13.78	0.68
(20)	94.99	93.53	1.46	(45)	-8.36	-8.41	0.04
(21)	97.50	98.21	0.71	(46)	-22.10	-21.38	0.72
(22)	102.15	102.89	0.74	(47)	-38.73	-41.37	2.64
(23)	-77.83	-76.86	0.97	(48)	-35.86	-36.00	0.13
(24)	-73.47	-72.77	0.69	(49)	-34.08	-31.32	2.77
(25)	-67.02	-68.68	1.66	(50)	-10.85	-12.36	1.50

Abbreviations: GA = group additivity; Abs Dev = absolute deviation in ΔH_f^o, |G4-GA|

Table B.7. Part 2 of 4. Enthalpies of formation from G4 and isodesmic reactions compared to those obtained from group additivity (GA) values. All values are in kcal/mol.

Species	ΔH_f^o from G4	ΔH_f^o from GA	Abs Dev	Species	ΔH_f^o from G4	ΔH_f^o from GA	Abs Dev
(51)	-6.74	-6.99	0.25	(76)	141.24	145.04	3.80
(52)	-4.06	-2.31	1.75	(77)	158.95	161.50	2.56
(53)	-4.07	-6.28	2.22	(78)	144.52	140.79	3.73
(54)	-1.65	-1.60	0.05	(79)	134.11	135.28	1.17
(55)	0.91	3.08	2.17	(80)	144.68	140.94	3.75
(56)	160.24	159.85	0.39	(81)	134.43	135.71	1.27
(57)	140.87	137.96	2.91	(82)	127.74	130.21	2.47
(58)	146.64	149.94	3.30	(83)	147.85	148.19	0.34
(59)	121.77	119.45	2.32	(84)	144.81	143.65	1.16
(60)	99.78	97.58	2.20	(85)	141.27	138.84	2.43
(61)	110.04	114.55	4.52	(86)	94.14	96.02	1.87
(62)	111.02	110.40	0.62	(87)	91.82	92.49	0.66
(63)	114.15	115.08	0.93	(88)	96.35	93.82	2.53
(64)	89.16	88.85	0.31	(89)	193.42	189.13	4.29
(65)	148.74	147.80	0.94	(90)	185.92	184.23	1.69
(66)	145.60	144.27	1.33	(91)	174.03	170.78	3.24
(67)	143.34	145.60	2.26	(92)	165.14	165.88	0.75
(68)	122.85	114.08	8.77	(93)	159.65	160.56	0.91
(69)	105.09	110.55	5.46	(94)	143.65	142.21	1.44
(70)	107.89	111.19	3.30	(95)	131.33	131.87	0.53
(71)	143.61	143.49	0.12	(96)	149.22	154.47	5.25
(72)	141.93	142.17	0.23	(97)	135.63	133.76	1.87
(73)	123.06	122.95	0.12	(98)	111.61	108.23	3.38
(74)	166.08	168.59	2.51	(99)	185.88	185.97	0.09
(75)	148.95	150.24	1.29	(100)	170.84	167.62	3.22

Abbreviations: GA = group additivity; Abs Dev = absolute deviation in ΔH_f^o, |G4-GA|

Table B.7. Part 3 of 4. Enthalpies of formation from G4 and isodesmic reactions compared to those obtained from group additivity (GA) values. All values are in kcal/mol.

Species	ΔH_f^o from G4	ΔH_f^o from GA	Abs Dev	Species	ΔH_f^o from G4	ΔH_f^o from GA	Abs Dev
(101)	173.47	176.60	3.13	(126)	205.53	204.00	1.53
(102)	135.09	133.90	1.19	(127)	185.87	183.45	2.42
(103)	122.19	127.19	4.99	(128)	193.33	197.29	3.95
(104)	117.16	113.35	3.80	(129)	-25.85	-25.86	0.01
(105)	109.70	115.65	5.95	(130)	-34.23	-34.15	0.08
(106)	97.36	94.94	2.42	(131)	-37.63	-37.70	0.07
(107)	103.16	99.62	3.53	(132)	160.83	160.44	0.39
(108)	139.18	141.97	2.79	(133)	161.39	159.12	2.27
(109)	126.73	121.95	4.77	(134)	151.08	153.73	2.66
(110)	114.45	116.44	1.99	(135)	185.95	185.30	0.66
(111)	-23.15	-25.52	2.36	(136)	168.10	164.75	3.34
(112)	-20.45	-20.84	0.39	(137)	174.59	178.59	4.00
(113)	-18.90	-16.16	2.75	(138)	80.50	83.73	3.23
(114)	48.23	51.20	2.97	(139)	79.33	80.20	0.87
(115)	48.68	47.67	1.01	(140)	86.32	82.22	4.10
(116)	50.95	49.00	1.95	(141)	-68.73	-68.54	0.19
(117)	-58.42	-56.95	1.47	(142)	-60.22	-62.48	2.26
(118)	-69.67	-69.48	0.18	(143)	-81.75	-79.69	2.07
(119)	-50.62	-52.27	1.65	(144)	-25.65	-27.79	2.13
(120)	211.30	207.89	3.41	(145)	-24.48	-23.11	1.37
(121)	183.06	187.34	4.28	(146)	-23.87	-23.11	0.76
(122)	147.69	146.82	0.87	(147)	185.68	179.80	5.88
(123)	65.06	66.25	1.18	(148)	166.06	171.31	5.25
(124)	62.97	62.72	0.25	(149)	168.30	168.93	0.63
(125)	64.29	63.36	0.93	(150)	110.29	109.55	0.75

Abbreviations: GA = group additivity; Abs Dev = absolute deviation in ΔH_f^o, |G4-GA|

Table B.7. Part 4 of 4. Enthalpies of formation from G4 and isodesmic reactions compared to those obtained from group additivity (GA) values. All values are in kcal/mol.

Species	ΔH_f^o from G4	ΔH_f^o from GA	Abs Dev	Species	ΔH_f^o from G4	ΔH_f^o from GA	Abs Dev
(151)	107.85	106.71	1.15	(176)	69.93	69.48	0.45
(152)	106.83	108.73	1.89	(177)	61.01	63.51	2.50
(153)	124.07	122.80	1.26	(178)	118.66	118.50	0.16
(154)	120.31	119.27	1.04	(179)	112.88	113.27	0.39
(155)	118.30	120.60	2.30	(180)	108.00	107.77	0.22
(156)	135.30	134.74	0.57	(181)	7.38	15.49	8.11
(157)	133.11	131.90	1.21	(182)	15.06	12.65	2.42
(158)	132.14	133.92	1.78	(183)	18.98	13.29	5.69
(159)	163.43	163.75	0.32	(184)	96.99	97.53	0.54
(160)	159.02	158.52	0.50	(185)	95.95	95.38	0.57
(161)	154.65	153.02	1.62	(186)	97.38	97.40	0.02
(162)	137.56	140.81	3.25	(187)	-103.93	-103.25	0.68
(163)	176.63	171.01	5.63	(188)	-99.63	-98.57	1.05
(164)	158.63	160.61	1.98	(189)	-92.17	-93.89	1.73
(165)	158.29	161.94	3.65	(190)	83.14	78.56	4.58
(166)	129.26	128.07	1.20	(191)	53.38	56.67	3.29
(167)	107.65	109.71	2.07	(192)	11.45	12.74	1.29
(168)	143.84	145.64	1.80	(193)	97.24	97.93	0.69
(169)	75.08	74.26	0.82	(194)	96.63	96.60	0.02
(170)	73.37	72.94	0.43	(195)	95.94	95.28	0.67
(171)	71.66	71.61	0.05				
(172)	32.92	38.85	5.92	mean absolute deviation			1.95
(173)	85.01	79.58	5.44	minimum absolute deviation			0.01
(174)	76.36	74.38	1.98	maximum absolute deviation			8.77
(175)	68.27	67.72	0.55				

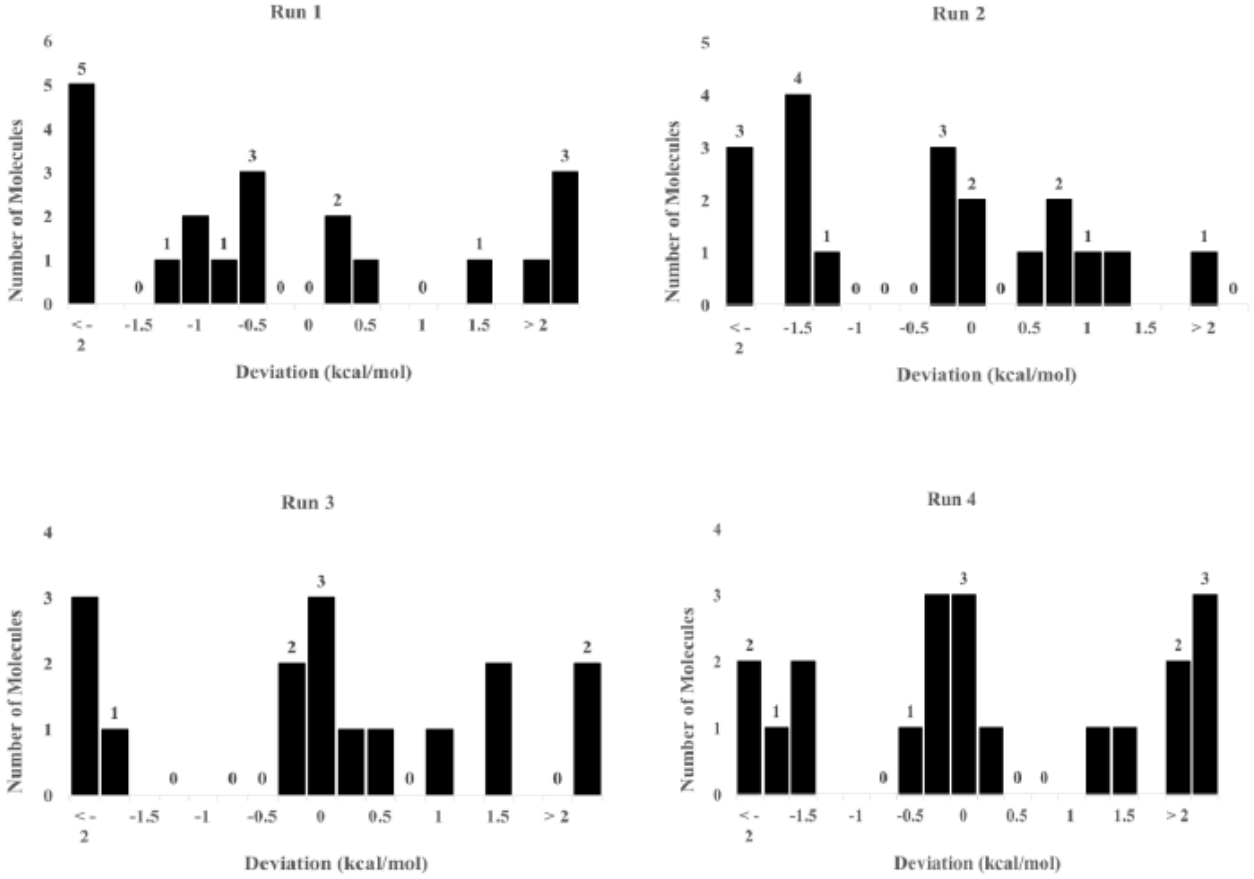


Figure B.1. Histograms of the deviations between the enthalpies of formation of 20 randomly selected molecules when all molecules were included in the matrix for regression versus when the 20 selected molecules were excluded from the matrix. Four different trial runs are shown above.

Appendix C: Supporting Information for "Microkinetic Modeling of the Vapor Phase Upgrading of Biomass-Derived Oxygenates"

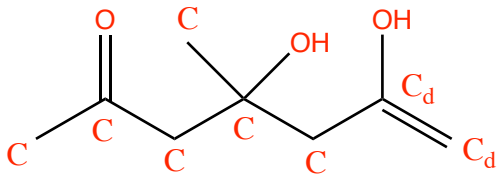
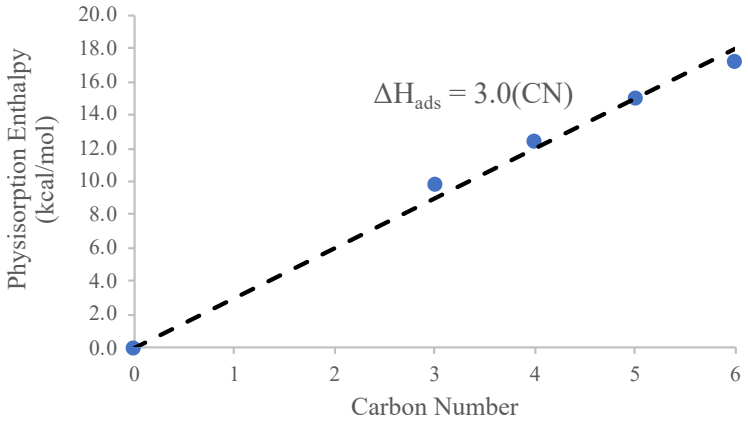


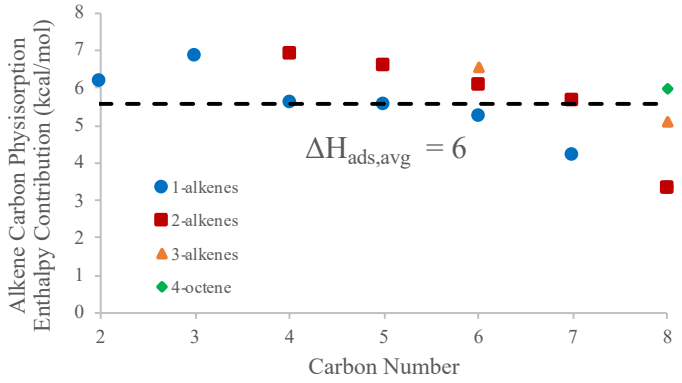
Table C.1. Example calculation: physisorption value for molecule pictured above.

Group	Amount	Contribution (kcal/mol)
Alkane Carbon	6	3.0(6) = 18.0
Alkene Carbon	2	6(2) = 12
Aromatic Carbon	0	2.5(0) = 0.0
-OH Group	2	23(2) = 46
=O Group	1	22.1(1) = 22.1
-OOH Group	0	23.1(0) = 0.0
Total	11	98
Final Enthalpy		98(0.6) = 59

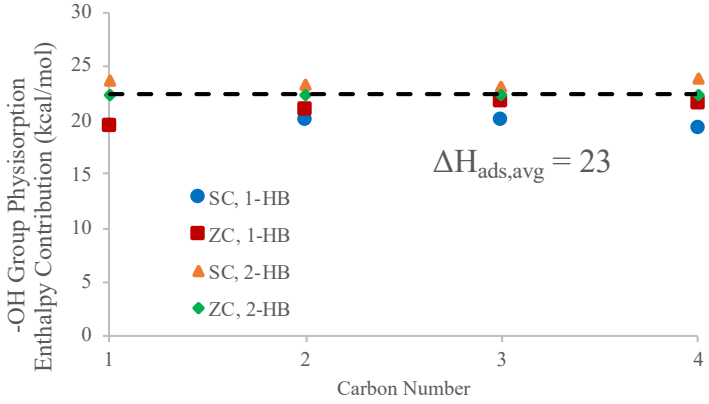
(a) Alkane Carbon Contribution



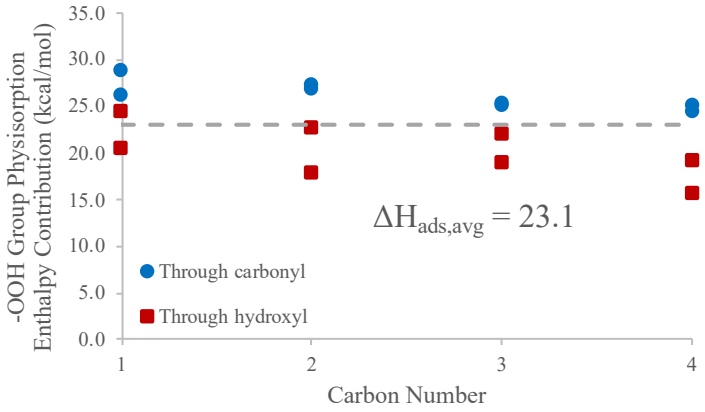
(b) Alkene Carbon Contribution



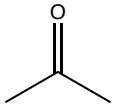
(c) -OH Group Contribution



(d) -OOH Group Contribution

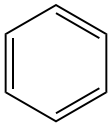


(e) Physisorption Enthalpy Calculation for the =O Group



$$31.07 - 3(3.0) = 22.1\ kcal/mol\ per$$ =O Group

(f) Physisorption Enthalpy Calculation for an Aromatic Carbon



$$15.2/6 = 2.5\ kcal/mol\ per\ aromatic\ carbon$$

Figure C.1. Determination of physisorption group contribution value for (a) alkane carbon atoms[184]; (b) alkene carbon atoms[185]; (c) -OH groups[186]; (d) -OOH groups[187]; (e) =O groups[188]; (f) aromatic carbon atoms[189,208].

Table C.2. Experimental physisorption enthalpies for alkanes C3-C6.[184]

Alkane	Carbon Number	ΔH_{ads} (kJ/mol)	ΔH_{ads} (kcal/mol)
propane	3	41.0	9.8
n-butane	4	52.0	12.4
n-pentane	5	62.5	14.9
n-hexane	6	72.0	17.2

Table C.3. Theoretical physisorption energies for 1/2/3/4-alkenes.[185] Note that the physisorption enthalpies can be taken equal to the calculated physisorption energies and can be considered independent of temperature in the range 300-800 K.

Alkene	Carbon Number	ΔE_{phys} (kJ/mol)	ΔE_{phys} (kcal/mol)	CD Contibution (kcal/mol)
ethene	2	52	12	6
propene	3	70	17	7
1-butene	4	72	17	6
1-pentene	5	84	20	6
1-hexene	6	94	22	5
1-heptene	7	98	23	4
1-octene	8	103	25	3
2-butene	4	83	20	7
2-pentene	5	93	22	7
2-hexene	6	101	24	6
2-heptene	7	110	26	6
2-octene	8	103	25	3
3-hexene	6	105	25	7
3-octene	8	118	28	5
4-octene	8	125	30	6
mean contribution				6

Table C.4. Theoretical physisorption enthalpies for alcohols.[186] Note that 1-HB is 1-hydrogen-bonded, 2-HB is 2-hydrogen-bonded, SC is in the straight channel, and ZC is in the zig-zag channel.

Bonding	Channel	Alcohol	Carbon Number	ΔH_{ads} (kJ/mol)	ΔH_{ads} (kcal/mol)	OH Contibution (kcal/mol)
1-HB	SC	methanol	1	94	22	19
		ethanol	2	109	26	20
		n-propanol	3	122	29	20
		n-butanol	4	131	31	19
	ZC	methanol	1	94	22	19
		ethanol	2	113	27	21
		n-propanol	3	129	31	22
		n-butanol	4	141	34	22
2-HB	SC	methanol	1	112	27	24
		ethanol	2	123	29	23
		n-propanol	3	135	32	23
		n-butanol	4	150	36	24
	ZC	methanol	1	116	28	25
		ethanol	2	131	31	25
		n-propanol	3	146	35	26
		n-butanol	4	161	38	27
		mean contribution				23

Table C.5. Theoretical physisorption enthalpies for carboxylic acids.[187] Note that there were two modes of bonding, which are depicted by Li and coworkers. Additionally, physisorption was conducted through the carbonyl oxygen and the hydroxyl oxygen.

Adsorption	Bonding	Carboxylic Acid	Carbon Number	ΔE_{phys} (kJ/mol)	ΔE_{phys} (kcal/mol)	OOH Contibution (kcal/mol)
Through carbonyl oxygen	Mode A	formic acid	1	121.8	29.1	26.1
		acetic acid	2	136.8	32.7	26.7
		propionic acid	3	143.0	34.2	25.2
		butyric acid	4	154.8	37.0	25.0
	Mode B	formic acid	1	133.2	31.8	28.8
		acetic acid	2	139.0	33.2	27.2
		propionic acid	3	142.5	34.1	25.1
		butyric acid	4	152.5	36.4	24.5
Through hydroxyl oxygen	Mode A	formic acid	1	114.8	27.4	24.4
		acetic acid	2	119.4	28.5	22.6
		propionic acid	3	129.4	30.9	22.0
		butyric acid	4	129.6	31.0	19.0
	Mode B	formic acid	1	97.7	23.4	20.4
		acetic acid	2	99.3	23.7	17.8
		propionic acid	3	116.2	27.8	18.8
		butyric acid	4	115.5	27.6	15.6
		mean contribution				23.1

Table C.6. Chemisorption energies, proton affinities, and stabilization energies for various 1-alkenes.

Alkene	Carbon Number	ΔE_{chem} (kJ/mol)	ΔE_{chem} (kcal/mol)	PA(alkene) (kJ/mol)	PA(alkene) (kcal/mol)	Δq (kcal/mol)
ethene	2	-130[a]	-31	704[b]	168	152
propene	3	-141[a]	-34	752[c]	180	143
1-butene	4	-147[a]	-35	773[b]	185	140
1-pentene	5	-156[a]	-37	787[b]	188	139
1-hexene	6	-155[a]	-37	805[c]	192	134
1-octene	8	-167[a]	-40	799[b]	191	138
1-nonene	9	-170[d]	-41	799[e]	191	139
isobutene	4	-72[f]	-17	802[c]	192	115

[a]From Nguyen and coworkers.[185] [b]From De Moor and coworkers.[193] [c]From NIST webbook.[194] [d]Extrapolated from a linear trend of propene and other 1-alkenes. [e]Assumed equal to the proton affinity for 1-octene. [f]From Nguyen and coworkers.[162]

Table C.7. Chemisorption energies, proton affinities, and stabilization energies for various n-alcohols.

Alcohol	Carbon Number	ΔE_{chem} (kJ/mol)	ΔE_{chem} (kcal/mol)	PA(alcohol) (kJ/mol)	PA(alcohol) (kcal/mol)	Δq (kcal/mol)
methanol	1	-117[a]	-28	754[b]	180	137
ethanol	2	-135[a]	-32	776[b]	185	136
propanol	3	-150[a]	-36	787[b]	188	137
butanol	4	-166[a]	-40	789[b]	189	141
pentanol	5	-183[c]	-44	795[d]	190	143
hexanol	6	-199[c]	-47	799[d]	191	146
heptanol	7	-215[c]	-51	799[d]	191	150
octanol	8	-231[c]	-55	799[d]	191	154
nonanol	9	-247[c]	-59	799[e]	191	158

[a]From Nguyen and coworkers.[186] Note that the zig-zag channel was chosen. [b]From Denayer and coworkers.[209] [c]Extrapolated from linear trend of n-alcohols with carbon numbers 1-4. [d]From NIST webbook.[194] [e]Assumed equal to the proton affinity for hexanol, heptanol, and octanol.

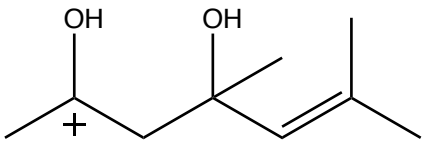


Table C.8. Example calculation for the stabilization energy of the carbenium ion depicted above.

Group	Contribution (kcal/mol)
Local interactions (ion type)	150
Non-local interactions from carbon atoms	$0.6(9)^2 - 8.2(9) = -25.2$
Alcohol local interaction (second neighbor)	16.8
Final stabilization energy	142

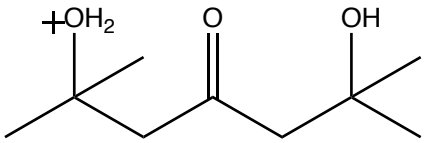


Table C.9. Example calculation for the stabilization energy of the oxonium ion depicted above.

Group	Contribution (kcal/mol)
Local interactions (ion type)	136
Non-local interactions from carbon atoms	$0.3(9)^2 = 24.3$
Ketone local interaction (third neighbor)	-3.8
Alcohol non-local interaction	23
Final stabilization energy	180

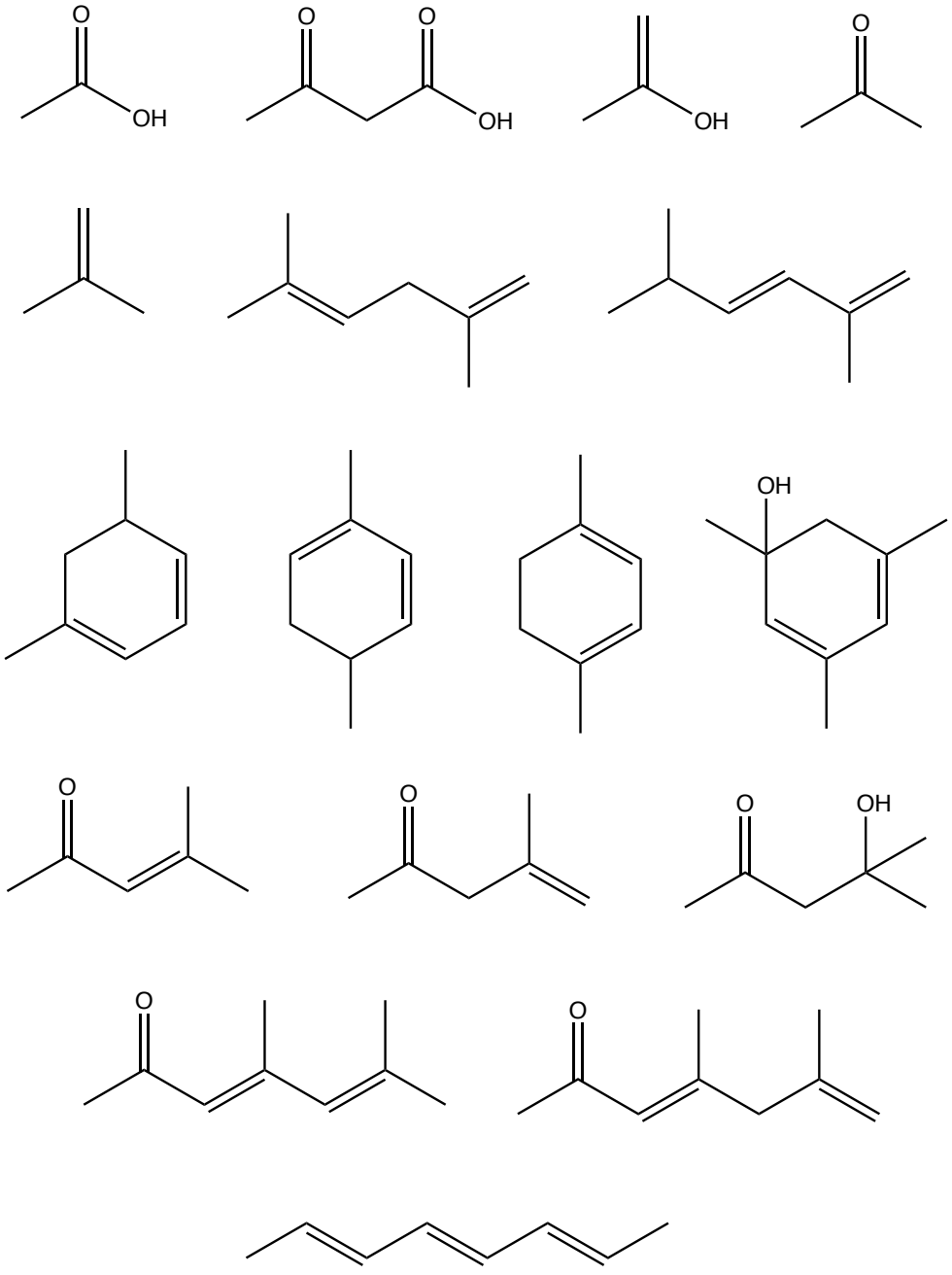


Figure C.2. List of seed molecules.

Table C.10. Inventory of elementary reaction steps.

Reaction Family	#
Protonation (p, s)	143
Protonation (t)	186
Deprotonation (p, s)	143
Deprotonation (t)	186
Alcohol Protonation	18
Oxodeprotonation	18
Keto Protonation	19
Alcohol Deprotonation	19
Hydration	17
Dehydration	17
Aldol Condensation	20
Retro Aldol	19
Acylium Ion Addition	1
Acylium Ion Removal	4
Oligomerization	58
β-Scission	116
Hydride Transfer (p, s)	145
Hydride Transfer (t)	36
Cyclic Hydride Transfer	13
Ring Cyclization (endo)	41
Ring Cyclization (exo)	1
1,2-Hydride Shift	138
1,2-Methyl Shift	106
α PCP Branching	31
β PCP Branching	36
Decarboxylation	1
Physisorption	314
Dephysisorption	314
Total	2160

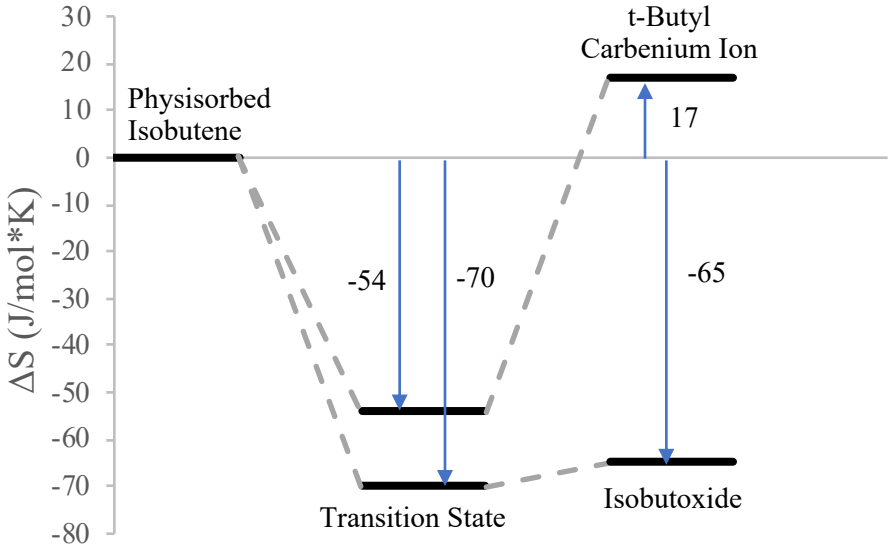


Figure C.3. Entropy changes upon protonation of physisorbed isobutene to isobutoxide and t-butyl carbenium ion.[162]

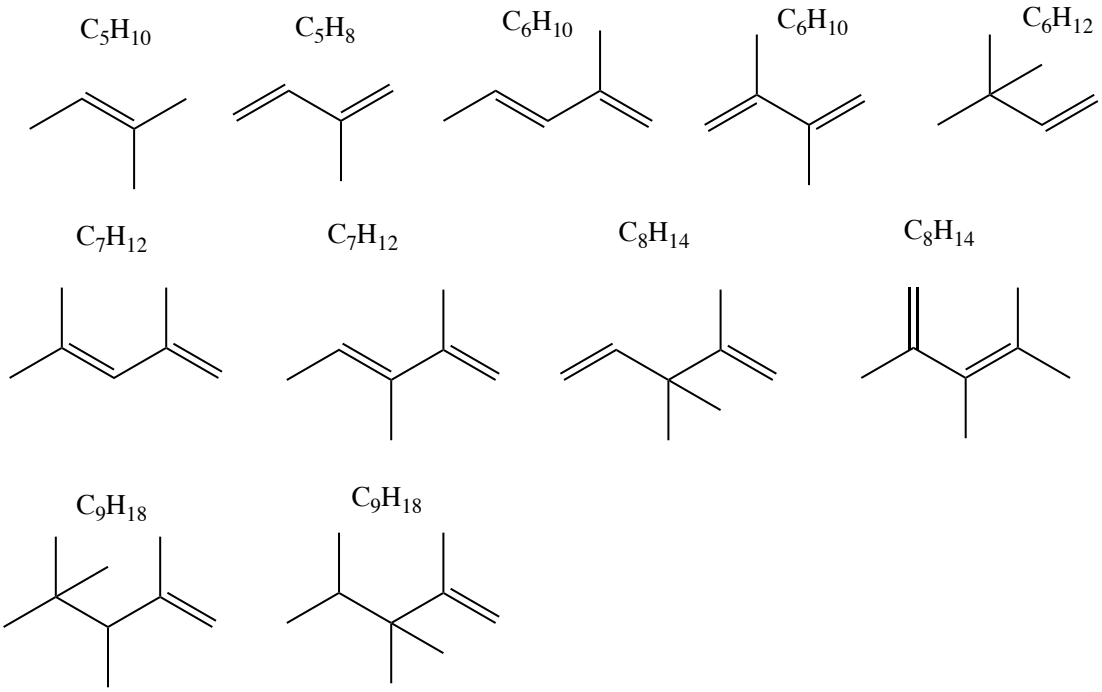


Figure C.4. C5+ olefins with noticeable presence in the product stream in the model.