Generating Detailed Kinetic Models for Large Pyrolysis Systems

by

A. Mark Payne

B.S. Chemical Engineering, Clemson University (2016)
B.S. Chemistry, Clemson University (2016)

Submitted to the Department of Chemical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Chemical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2023

Author . . . .
Department of Chemical Engineering
September 26, 2022

Certified by . . . .
William H. Green
Hoyt C. Hottel Professor in Chemical Engineering
Thesis Supervisor

Accepted by . . . .
Patrick S. Doyle
Robert T. Haslam Professor of Chemical Engineering
Singapore Research Professor
Chairman, Committee for Graduate Students

Generating Detailed Kinetic Models for Large Pyrolysis Systems

by

A. Mark Payne

Submitted to the Department of Chemical Engineering
on September 26, 2022, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Chemical Engineering

Abstract

Detailed kinetic models have been able to accurately the predict the behavior of many complex chemical systems. The benefits of such models is numerous, ranging from being able to predict system behavior under conditions not amenable to experiments to the fact that the mere process of generating such models often leads to the discovery of new reaction pathways. Despite this utility, to date these models have mostly been applied to smaller systems of 10 heavy atoms or less. This is because as the size of the molecules grows, the number of possible isomers and thus reactive pairs grows combinatorially. Furthermore, refining these models often involves high-accuracy quantum chemistry calculations that are expensive for larger species. If these challenges can be overcome, though, generating detailed kinetic models for larger systems aims to provide valuable insights into complex systems, such as the pyrolysis of heavy oil or biomass. In this work, we show that advances in automatic mechanism generation software, quantum chemistry methods, and ever increasing amounts of computational power have made the prospect of generating detailed models for larger systems possible. We were able to generate a detailed kinetic model for the pyrolysis a 3-component hydrocarbon mixture with the largest species containing 18 heavy atoms. Despite the size of the molecules, the generated model was able to predict experimental data for this system. We also discuss aspects of refining these models with quantum chemistry calculations, specifically calculating species thermochemistry. We showed that many of the methods for correcting these calculations, including bond-additivity corrections and isodesmic reaction approaches yield similar results, despite some claims to the contrary. Finally, we collected experimental data necessary to validate detailed kinetic models for the pyrolysis of kerogen. As part of this work, we discussed the challenges of collecting such data, and showed the suitability of modern methods and instrumentation towards this task. With this, it is likely that detailed kinetic models will be increasingly used to study larger systems, though this work will likely involve fully-detailed model compound studies in tandem with approaches to reduce the combinatorial complexity of large systems without much loss in accuracy.

Thesis Supervisor: William H. Green
Title: Hoyt C. Hottel Professor in Chemical Engineering

Acknowledgements

This dissertation would not have been possible without the tremendous amount of support I have received from countless individuals, from project collaborators many of whom helped co-author many of the publications based on this work, to mentors who helped me grow as a scientist and engineer over the years, to the friends and family who supported me along the way. The list of people who have helped me throughout this journey and made an impact on my life is large, and there are many individuals that I want to thank here in particular.

First, I thank my Ph.D. advisor, Dr. William H. Green, for being a fantastic advisor and role model to me. Despite many struggles throughout my Ph.D., Dr. Green never gave up on me and believed that I could achieve my goals and grow as a scientist and engineer. I always felt energized to keep going after talking with Dr. Green, no matter how bleak my projects seemed at the time. He also encouraged me to continue pursuing my teaching interests while at MIT, and always gave me great career advice in how to navigate both the academic and industrial landscape. Beyond all of this, though, Dr. Green showed me how to be the scientist and engineer that I want to become one day, from the way he treated everyone with kindness, to the way he fostered a collaborative environment in his group, to always trying to do the ethically right thing, to his genuine interest in wanting to discover the truth, wherever that may lie. While I have much in the way of growing, it is because of Dr. Green's guidance that I know one day I will make it as the engineer that I want to become, and I will have Dr. Green to thank for this.

I also thank Dr. Richard H. West and Dr. Heather J. Kulik, for serving on my thesis committee and for all of the advice they gave me these past years. Their questions and ideas were invaluable, and often helped me focus my projects to make sure they were achievable and in line with my broader goals. I especially appreciated learning as much as I could about quantum chemsitry from Dr. Kulik, and learning about RMG and software development from Dr. West. I also thank Dr. Tisdale for agreeing to preside over my defense at the last minute, and for serving as a reference for me in getting my next job after MIT.

In addition, I had many other mentors over the years that really helped me grow as a person and as an engineer. I thank Dr. Alon Grinberg Dana for taking me under his wing early on in my Ph.D., and for always being someone I could turn to when I needed advice. I also thank Dr. Mengjie Liu for teaching me so much of what I know as a software developer,

teaching 10.213. I also thank Dr. Daphne Chan for being such a great mentor to Helen and I while she served as Instructor G. During my time as Instructor G I also got to teach alongside Katie Steinberg, Dr. SK Ha, and Spencer Zhu. I thank them for being such great TAs, especially in dealing with remote teaching. Finally, I thank Dr. Joey Gu for all of his help showing me how to make, deliver, and edit high-quality light-board lectures and showing me the ropes to digital learning.

In addition to all of the people who have helped me at MIT, I also had great support from my family and friends. I thank my mom and dad, Matt and Susanne Payne, for everything they have given me over the years, helping me get as far as I did, and for encouraging me to chase after my dreams. Thank you, Dad, for being such a great role model in my life, for being my best man, and for always being there to guide me through starting a career in chemistry and chemical engineering. Additionally, thank you, Mom, for also being such a great role model in my life, for all of the love and support you have given me over the years, and for helping me become the person I am today. I also thank my sister and brother in-law, Kathryn and Nick Pastore, for being there for Blakeley and I throughout the years.

I also thank my parents-in-law, Scott and Maya Hoffman, for welcoming me into their family, supporting me from a far, and for introducing me to so many things that helped keep me going throughout my Ph.D., including photography and also a bit of teaching. Thank you both for all of the love and support you have given me.

I thank my extended family as well, grandparents, aunts, uncles and cousins, who supported us from a far, and who I am looking forward to seeing more of now that I am through this Ph.D. journey. I love and miss you all so much.

I also thank our dog, Roy, for keeping my spirits up through the last bit of my Ph.D., and for being so eager to help write all of my papers.

In addition to my family, I thank many of my friends, including Dr. Kindle Williams, Dr. Caroline Nielson, Dr. AJ Fention, Dr. Sharon Lin, and Suzane Cavalcanti, for supporting me throughout my time at MIT. I also thank my good friends from undergrad, Tony Scavuzzo, along with Alex and Teague Albenesius, who also supported me from a far.

Last but not least, I thank the love of my life, my wife Blakeley Payne, for always being by my side throughout this journey, and for soon-to-be 5 wonderful years of marriage along the way. Working towards a Ph.D. is arduous, and without having Blakeley supporting me in so many ways I would not have made it this for. I cannot thank you for all of your

support during all of the late nights, during all of the times when I got in my own way, for being there to celebrate all of the successes, and keeping my spirits up during all of the failures. I know it wasn't always easy. You sacrificed a lot to get me this far and to make sure that we could always be together throughout all of this. I cannot tell you how much this all means to me. I am looking forward to many more joyous years of marriage, and for the many adventures we have to look forward to together. As you start your own Ph.D. journey, I sincerely hope that I can be a fraction of the loving and support spouse that you were to me.

Contents

Chapter 1

Introduction

Developing kinetic models is key to understanding how many reactive chemical systems behave. With a kinetic model, we are able to predict how a system will behave under different conditions, gain new insights into underlying chemistry, and even discover how to control reactive system to achieve favorable outcomes, such as maximizing the yield of a desired product. While a good kinetic model can have a tremendous amount of utility, generating these models can be a time-intensive and laborious task riddled with challenges that must be overcome. For example, while some systems are as simple as a few reactants forming a desired product and a handful of undesired products (like those utilized for organic synthesis), many other systems involve a reactive network of hundreds to thousands of relevant species and hundreds of thousands of relevant reactions. This is especially common in systems that involve free radical chemistry such as pyrolysis and combustion, as almost any of the radicals formed in these networks can react is countless abstraction and recombination reactions that result in a combinatorial explosion of reactions to consider.

For complex chemical systems such as these, there are two common approaches to generating accurate kinetic models. The first approach is to perform experiments on bulk material of the system of interest and try to fit a kinetic model to the observed data. For example, to study the pyrolysis of a certain feedstock of biomass, one might place a sample of this biomass in a retort, pyrolyze the sample under a set of conditions, and then categorize and quantify the products that formed. One might then assume that this transformation of the starting feedstock to the various products obeys first-order kinetics, and thus generating a kinetic model is as simple as fitting parameters (activation energies and frequency factors) such that the discrepancy between the first-order kinetic model and the observed experimen-

tal data is minimal. There are many advantages to this approach. For one, this approach sidesteps the issue of having to enumerate the seemingly endless number of reactions and reactive intermediates involved in the underlying chemistry while still being able to capture the bulk behavior of the system. A related point here is that the products are usually categorized in terms of observables of interest to the person who generated the model. For example, the amount of light hydrocarbons formed might be the relevant observable that a particular party is interested, with distinctions between various molecules that could make up these lighter hydrocarbons being more detailed than necessary for the desired analysis. Finally, fitting kinetic parameters to match experimental data is a fairly straightforward task.

There are some notable downsides to this approach, though. For one, it is easy to overfit a kinetic model to experimental data. While this would result in a model with minimal errors at the same conditions as the experimental data, it is possible that simulating the model at different conditions will unknowingly lead to large errors. Furthermore, these models don't give much more in terms of insight beyond what was gained from the experimental data alone. It is possible to speculate what sort of processes might be involved based on the fitted activation energies, for example, but these models don't shed insight into key reactive intermediates or new reaction pathways.

The second approach that can be used to generate accurate kinetic models is to generate detailed kinetic models *ab initio*. In this approach, the generated model does not rely on bulk experimental data but rather tries to predict the experimental data *a priori*, with bulk experiments only being performed to validate the model. In this approach, a chemical mechanism is constructed containing a set of elementary reactions that are relevant at the desired conditions. The rate parameters for these reactions can be obtained through a variety of methods, including quantum chemistry calculations, careful experimental measurements, or various approximation methods. In fully-detailed models, the reaction networks contain distinct molecules as species. However, there are other techniques that follow a similar approach but don't track individual species. For example, it is possible to reduce some complexity by only tacking the relevant functional groups, or lumping types of species or isomers together.

There are numerous advantages to this approach. For one, these models don't carry the risk of being overfit to match experimental data. In fact, if the generated model is able to

predict the bulk experimental data well, that gives confidence that the model is actually capturing the underlying chemistry well. Because of this, it is much more likely that the model is able to make accurate predictions across a wide range of conditions. These detailed kinetic models also provide much more insight into the underlying chemistry, making it possible to determine key reaction intermediates and pathways. No detailed kinetic model will ever be able to minimize errors with bulk experimental errors in the same way that directly fitting to this data is able to achieve. However, these discrepancies often hint at missing pathways, allowing us to discover new chemistry.

Unlike fitting model parameters to bulk experimental data, the task of generating a detailed kinetic model is quite complex. It can be a difficult task to determine which reactions should be included in the model. It is not possible to include all conceivable reactions into the model even if they could be enumerated, as the resulting model would be too large to simulate. Therefore, or reactions that seem relevant should be included. Further challenges arise when previously unknown reaction pathways are key to obtain accurate predictions. Even if all of the relevant reactions are included, errors in the rate parameters for these reactions might be large enough to cause significant errors in the final predictions. Because of this, a key step in generating a detailed kinetic model is refining the model parameters, which usually involves performing high-accuracy quantum chemistry calculations.

Obviously, for all but the smallest of systems, generating such a model manually is a daunting task. Thankfully, over the past few decades a great volume of work has gone in to producing automatic mechanism generation software, which is capable of computer-generated a detailed kinetic model. One example such software is the Reaction Mechanism Generator (RMG). As input, RMG requires the initial conditions, including the composition of the starting material and the desired operating conditions, and outputs a detailed kinetic mechanism that can be simulated to predict the behavior of the real chemical system. Through its flux based algorithm, RMG aims to only include the relevant species and reaction based at the conditions of interest. RMG is able to use reaction templates to enumerate possible reactions in the system, and has access to a database of known rate parameters to estimate the rate parameters of the found reactions.

To date, many systems have been successfully modeled using detailed kinetic model generated from software like RMG. However, most of the systems have been smaller systems with 10 heavy atoms or fewer. Part of this is just due to the fact that larger systems simply

have more possible reactions and intermediate species. Additionally, modeling systems with larger species also requires performing quantum chemistry calculations on larger species in order to refine the model. With the availability of faster computers and quantum chemistry methods, though, it should be possible to study larger and larger systems. If possible, this would allow for developing kinetic models that can accurately extrapolate to conditions previously not possible, as to date developing kinetic models for large systems has almost exclusively relied on fitting models to bulk experimental data.

There are many challenges unique to large systems though that must be overcome. In this work, we generate detailed kinetic model containing relatively large species to show that this work can be done, but also to highlight some of the challenges of this work. Furthermore, we discuss some aspects of refining these models given that the species in these models are large. Finally, we discuss methods for obtaining experimental data for validating these models.

1.1 Thesis overview

In this work, we not only develop a detailed kinetic model for a large chemical system, but we also discuss some important aspects of working to develop such models. Chapter 2 shows the development of a detailed kinetic model for the pyrolysis of a mixture containing an alkane, aromatic, and alkylaromatic using RMG. The model generated shows reasonable accuracy to experimental data collected for model validation. Chapter 3 and Chapter 4 then discuss additional aspects relating to work in developing such models. Chapter 3 discusses obtaining accurate species thermochemistry for large molecules using modern quantum chemical methods, which is an important step in refining detailed kinetic models. Chapter 4 instead focuses on the experimental side of studying large systems, studying the low temperature pyrolysis of kerogen samples, and discussing methods for characterizing the initial feedstock and pyrolysis products. Finally, Chapter 5 summarizes the major findings and discusses future work that should follow from this thesis.

Chapter 2

Detailed Reaction Mechanism for 350-400 °C Pyrolysis of an Alkane, Aromatic, and Long-Chain Alkylaromatic Mixture

Chapter Abstract

Many technologically important systems involve mixtures of fairly large molecules and relatively unselective chemistry, leading to complex product mixtures. These corresponding reaction networks are quite complex since each molecule in the feed can form many isometric intermediates and a variety of byproducts in addition to its major product. A variety of modeling methods have been developed to attempt to deal with this, but building accurate reaction mechanisms for these complicated systems is challenging and the methodology is still under development. To showcase the advancements that have been made in automatic

generation of large mechanisms, we constructed such a model for a three-component mixture containing species with up to 18 carbon atoms. The generated model is able to predict many of the major and minor products with relatively high accuracy against gold-tube batch pyrolysis data collected for this system. The high fidelity between the predicted species profiles and the experimental data is notable given the low temperature pyrolysis conditions studied, as any errors in *ab initio* rate parameters become more significant at lower temperatures.

2.1   Introduction

Detailed kinetic mechanisms are valuable engineering tools for understanding and predicting chemical processes. Elementary-step reaction mechanisms have been very helpful for understanding atmospheric chemistry, the pyrolysis of alkanes to form alkenes, and the combustion of simple fuels such as natural gas.[1,2] It would be useful if the same sorts of models could be applied to more complicated reacting systems of larger molecules. Several software packages have been developed to automate mechanism generation by systematically exploring the numerous reaction pathways required to describe these complex systems.[3–6] These software packages have been used to construct many reaction mechanisms, gradually progressing from four-carbon fuels, such as butanol[7] and methyl propyl ether combustion,[8] to larger starting molecules that result in more complex models, e.g., combustion of a mixture of butane with six substituted phenols.[9] However, constructing accurate and detailed models for mixtures of still larger molecules remains challenging due to the combinatorial explosion in the number of possible reaction sites and reaction intermediates. Most existing models for combustion or the pyrolysis of fuel molecules with more than 10 carbon atoms therefore use lumping and other approximations to keep the number of reactions and species manageable. From the software perspective, we recently parallelized the Reaction Mechanism Generator (RMG) software[10] and implemented methods to reduce its demand for memory;[11] together, these improvements make it possible to build models for larger systems before being forced to make these approximations of the chemistry.

Possible strategies for modeling important, yet complex, systems such as heavy oils include structure-oriented lumping[12] and fragment-based modeling.[13] Another promising strategy is to generate a kinetic mechanism for model compounds. Surrogate models are

becoming more common, with examples including models for rocket propellant fuel,[14,15] jet fuel,[16,17] and biofuel.[18] A particularly interesting case study for surrogate modeling is heavy oil pyrolysis. According to the 2020 World Oil Outlook, oil is expected to remain the largest contributor to the global energy mix until 2045.[19] In the long-term, oil demand per day is projected to increase by nearly 10 million barrels, rising to approximately 109 million barrels per day in 2045. As oil demand increases, traditional oil compositions will inevitably be consumed, requiring the use of heavy oils with higher molecular weight hydrocarbons. Thus, there is significant motivation for developing models that improve our understanding and prediction capabilities of heavy oil pyrolysis.

Because of its importance in the conversion of heavy oils into valuable transportation fuels, the pyrolysis of alkylaromatics has been studied extensively in the past.[20–28] Additionally, there are many works in the literature that have used computer-aided mechanistic simulations to try and understand pyrolysis systems, dating as far back as at least 1979 to work by Dente and Ranzi with the SPYRO model.[29] However, to date, little work has been done in autogeneration of surrogate models for these systems. Here, we build on the existing literature and explore the current capabilities of RMG[6] to autogenerate a surrogate model for high-pressure pyrolysis of heavy oil. Our surrogate system is composed of a three-component mixture of undecane, toluene, and dodecylbenzene. We choose this specific surrogate system for several reasons. First, it has relevant aliphatic, aromatic, and alkylaromatic functional groups. The resulting products should be large enough to be easily measured experimentally, yet not too large as to make detailed kinetic modeling of the relevant chemistry intractable. We recently performed a series of quantum chemistry calculations on types of reactions thought to be important in alkylaromatic free radical chemistry; so, we have improved confidence in our ability to estimate the rates and the thermochemistry of various reaction intermediates.[30,31] Finally, dodecylbenzene has been studied previously,[20–22] providing additional experimental checks on our results beyond our own measurements reported here. Although the most obvious practical application of this work is to heavy oil pyrolysis, we note that a recently published process for up-cycling waste polyethylene involves a mixture with a similar composition,[32] so the understanding gained here may also be helpful for addressing that important challenge.

2.2 Methods

2.2.1 Materials

Undecane, toluene, and dodecylbenzene were obtained from Sigma-Aldrich and used as reagents. 3-Chlorothiophene (Sigma-Aldrich) was used as an internal standard for quantification with gas chromatography with flame ionization detection/mass spectrometry (GC-FID/MS) and GC×GC-FID.

2.2.2 Pyrolysis

Gold Tubes

All pyrolysis experiments were performed inside gold tubes (XRF Scientific, 99.99%, 5 mm OD, 0.2 mm thick) of 30 ± 10 mm in length. Because gold is malleable and the tubes are thin, the tube itself will be compressed under pressure such that the contents of the tube are at the same pressure as the surroundings. This allows for the contents of the tube to be in thermal and mechanical equilibrium with its surroundings while isolating the contents from the steel batch reactor, which is known to catalyze wall reactions.[33]

Prior to loading the sample in the tubes, the tubes were cleaned by placing them in a bath of 10 M HCl (trace metal analytical grade) for 24 h. The tubes were then rinsed with deionized water (Millipore) and then annealed using a butane torch. After the tubes were allowed to cool, one end of the tube was crimped and arc welded shut. Then, 0.3 ± 0.1 g of the starting material (dodecylbenzene, toluene, and undecane mixture) was loaded into the gold tube. Prior to crimping the remaining end of the gold tube, argon gas was blown into the tube to achieve an argon atmosphere inside the tube. After crimping, the remaining end was arc welded shut.

Experiments at 400 °C

One gold tube at a time was placed into a stainless steel batch reactor (SITEC 740.8036, 22.6 mL volume) using steel wool to hold the gold tube in place. The batch reactor was then sealed, hung from a gantry crane, and then pressurized with deionized water to 300 bar using a 100 mL ISCO 100DM syringe pump. The pump controller was set to maintain a pressure of 300 bar throughout the run and as the contents of the batch reactor heat up

and expand. The batch reactor was then lowered on the gantry crane into a sand bath with a temperature controller set to 400 °C. At the end of the sample run time, the reactor was raised out of the sand bath and lowered into a large barrel of water to cool off the reactor and quench the reactions. After the reactor had cooled, the pressure of the reactor was lowered to atmospheric pressure before opening the reactor to retrieve the gold tube. A diagram of this setup is shown in Figure 2.1.

Although the temperature inside the reactor was not monitored, it took the reactor about 15 min to reach thermal equilibrium, based on how long it took the syringe pump to equalize the volume delivered to the reactor at a constant pressure of 300 bar (see Model Simulation and Analysis). Samples were run at separate time points of 3 h and 6 h, with one additional replicate run at the 3 h time point, and two additional replicate data points were acquired for the 6 h time point.

It is worth noting that, for every experimental run, the resulting pyrolysis liquid was clear of solids and the walls of the gold tube were clean, showing no signs of coke formation (this was also true for the later runs at 350 °C). This coupled with the fact that 300 bar is above the critical pressure of all expected species gives us confidence that the sample remained a one-phase solution throughout the runs.
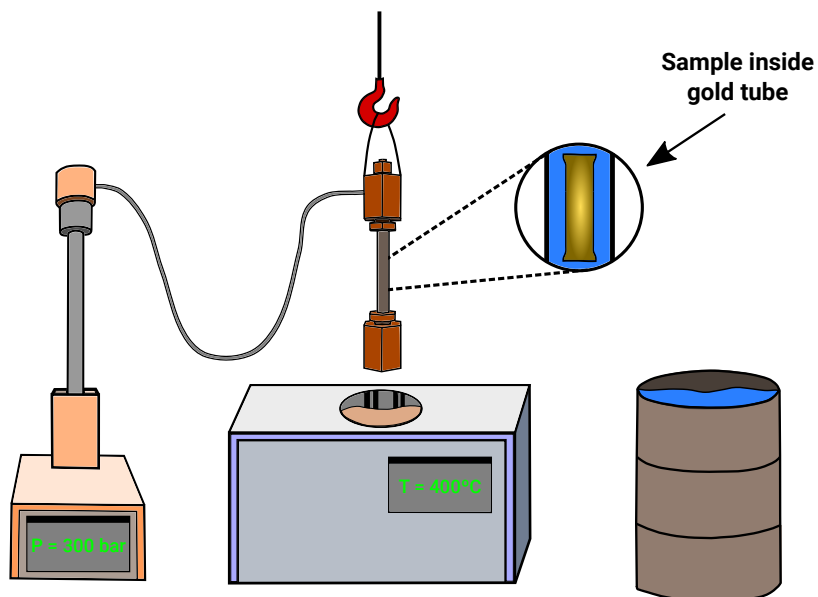
Figure 2.1. Experimental apparatus for gold-tube pyrolysis experiments, consisting of an ISCO 100DM filled with water (left), a SITEC stainless steel batch reactor attached to a gantry crane above a sand bath (middle), and a barrel of water (right). At the end of a run, the reactor can quickly be raised out of the sand bath and lowered into the water barrel to quench the reaction.

Experiments at 350 °C

Pyrolysis experiments at 350 °C were performed at the Woods Hole Oceanographic Institute (WHOI) using the same high-pressure gold tube technique inside of a tube furnace at two separate time points of 116 h and 239 h. Only one data point for each time point was available for these conditions. The tube furnace was then pressurized to 300 bar and then heated to 350 °C. At the end of the run, the tube furnace was allowed to cool before releasing the pressure to atmospheric and retrieving the gold tubes.

2.2.3 Error Bars for Experimental Data

Due to limited replicate data (as some planned replicate runs suffered loss of sample during pyrolysis with rupturing of the gold tube), determining proper error bars was a difficult task. For the available replicate data at 400 °C, relative standard deviations for each species concentrations were determined. However, the relative standard deviation calculated for a single species from two to three points is a poor estimator of the true variability that we can expect from repeat measurements. Instead, we looked at the average relative standard

deviation across all species. For example, the average relative standard deviation was 19% for data at 400 °C and 3 h. On the basis of this, it was estimated that a relative error bar of 20% was appropriate.

Here, we assume a 20% relative error for all species across all of the experiments, except for the reactants at the initial time point. The initial sample composition was measured very precisely gravimetrically, and the error due to this alone would be less than 1%. However, we conservatively chose a relative error bar of 1% to account for the possibility that the impurities are larger than we think, though both undecane and toluene have a reported grade of >99% and dodecylbenzene was measured to have a purity of >99%. Right horizontal error bars were added to account for the 15 min temperature ramp-up.

2.2.4 GC-FID/MS

Liquid-phase pyrolysis products were identified and quantified using GC-FID/MS on an Agilent 7890. RXi-5HT (30 m x 0.25 mm ID x 0.25 μm film thickness) was used as the primary column, and helium was used as the carrier gas. The outlet of the primary column was sent to an Agilent splitter connected to both the FID and the MSD. Prior to being analyzed, pyrolysis samples were mixed with a small amount of 3-chlorothiophene for quantification using it as an internal standard. Usual makeup consisted of 0.2 grams of sample mixed with 0.02 grams of standard, with the exact amount determined gravimetrically for each run. The FID was used to quantify species concentrations, whereas the MS chromatogram was used to identify species using the NIST spectral database to perform library searches. Prior to analyzing the pyrolysis samples, response factors of known products relative to the internal standard (3-chlorothiophene) were determined by preparing known solutions of the reagents.

2.2.5 GCxGC-FID

Liquid-phase pyrolysis products were also analyzed using GCxGC-FID on a LECO modified Agilent 7890. RXi-5HT (30 m $\times$ 0.25 mm ID $\times$ 0.25 μm film thickness) was used as the primary column, and RXi-17SIL MS (2m $\times$ 0.150 mm ID $\times$ 0.15 μm film thickness) was used as the secondary column using helium as the carrier gas. The modulation loop in between the primary and secondary column used cold jets cooled with liquid nitrogen, and a hot jet pulse every 16 s for the modulation time. The primary oven was ramped-up from
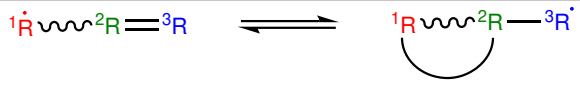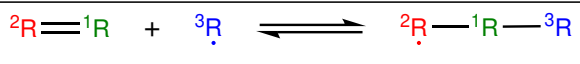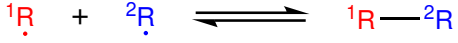
50 to 300 °C over the course of the run (5330 s), and the secondary oven was held at 25 °C above the primary oven. The outlet of the secondary column was then sent to an FID. This analysis was primarily used qualitatively to see what type of compounds were being formed. Without easy identification by mass spectrometry, we were unable to identify most of the peaks visible in the spectrum. Therefore, the resulting GC×GC chromatograms were not used for quantification.

2.2.6 Automatic Mechanism Construction with the Reaction Mechanism Generator

The Reaction Mechanism Generator (RMG) version 3.0[34] was used to automatically generate a surrogate mechanism for modeling heavy oil pyrolysis. More detail about RMG can be accessed in previously published work,[6,34] though the methodology is briefly summarized here. First, initial mole fractions, temperature, and pressure are specified, in addition to termination criteria, such as reactant conversion. The initial reactants are placed in the model core and reacted together using a set of reaction templates to form edge product species. The specified set of templates is shown in Table 2.1; if all reaction templates from RMG-database are allowed, the combinatorial explosion of possible reaction sites and subsequent products quickly becomes expensive. The rate-based algorithm determines whether an edge species and corresponding reactions should be moved to the model core; this process is repeated to iteratively expand the model until the termination criteria is satisfied. RMG is an open-source software project that is frequently updated. To replicate the RMG model generation from this work, one should use exactly the same version of the RMG-Py software and corresponding RMG-database; the GitHub version commit strings are outlined in Appendix A. The input file is also included in Appendix A.

Table 2.1. RMG Reaction Templates.

RMG Reaction Family	Template
Disproportionation	$^1\dot{R}$ + $^3\dot{R}$—^2R—^4H ⇌ ^1R—^4H + ^3R=^2R
H_abstraction	^1R—^2H + $^3\dot{R}$ ⇌ $^1\dot{R}$ + ^2H—^3R
Intra_R_Add_Endocyclic	$^1\dot{R}$⟋^2R=^3R ⇌ ^1R⟋$^2\dot{R}$—^3R

22

Intra_R_Add_Exocyclic	$^1\dot{R} \sim\sim ^2R = ^3R \quad \rightleftharpoons \quad ^1R \sim\sim ^2R - ^3\dot{R}$
R_Addition_MultipleBond	$^2R = ^1R \;+\; ^3\dot{R} \quad \rightleftharpoons \quad ^2\dot{R} - ^1R - ^3R$
R_Recombination	$^1\dot{R} \;+\; ^2\dot{R} \quad \rightleftharpoons \quad ^1R - ^2R$

The initial species placed in RMG's core included equimolar concentrations of undecane, toluene, and dodecylbenzene, the same three components used during the experimental measurements of the surrogate system. Using domain knowledge and results from experimental measurements, some additional aliphatics, aromatics, and alkylaromatic species were added to RMG's initial core to help guide model construction; these species are listed in the input file in Appendix A. RMG built the mechanism by simulating at 400 °C and 300 bar in a constant temperature, constant pressure batch reactor. Note that RMG did not converge with the tolerance specified in the input file. A loose tolerance criteria results in smaller, less detailed reaction networks, while the tighter tolerance used here results in more complete models that require significant resources for mechanism generation to converge. Ideally, several input parameters would be explored, and the resulting models would be compared. Here, mechanism generation was terminated after running for approximately two months on a small 10-year-old server. Up to four cores were utilized during reaction generation. Recent testing suggests the generation time could be cut by an order of magnitude using a modern supercomputer, whose larger memory allows RMG to use more cores in parallel while running on a single node.[10]

The first challenge in building an accurate reaction mechanism is including all of the important reactions to control mechanism truncation error while also excluding unimportant reactions to keep the mechanism at a manageable size. RMG's default reaction generation algorithm explores all possible reactions between all combinations of core and edge species, which can be impractical. Several RMG input options were used to limit both the computational effort and required memory. For example, each molecule was restricted to a maximum of 40 carbon atoms. This is large enough for dodecylbenzene radicals to recombine yet avoids wasting computational resources exploring all reaction sites on enormous hydrocarbons that would otherwise be allowed in the model. The number of unpaired electrons per molecule was also limited to one (i.e., no biradicals in the model) to reduce the num-

ber of edge species generated. Model generation relied on recent RMG advances, including pruning edge species to conserve memory,[11] reaction filtering,[35] improved representations of aromaticity,[36] improved group additivity for thermochemistry estimation of polycyclic compounds and radicals,[31,37] and the new parallelized reaction generation method.[10] In the pruning algorithm, the maximum number of edge species considered at any iteration was limited to 300 000 to avoid excessive memory requirements and avoid spending compute power to consider reactions of unimportant edge species with very low flux.

The second challenge in kinetic modeling is coming up with accurate numerical values for all of the thermochemical and rate parameters to control parameter error. While the kinetic model was being constructed, several contributions were made to RMG's open-source database used to estimate thermodynamic and kinetic parameters. After mechanism generation, these parameters in the reaction network were updated to use the more recent values, which significantly improved the model's predictions. The GitHub commit string for the version of RMG-database used to update the model parameters is listed in Appendix A.

2.2.7 Model Simulation and Analysis

The surrogate mechanism generated by RMG was simulated in Cantera version 2.4.0[38] to predict species concentration as a function of reaction time. Since the goal is to validate model predictions against experimental observations, the species trajectories were generated using the same experimental conditions. Thus, a constant temperature, constant pressure ideal gas batch reactor was used to simulate the mechanism at 400 °C and 300 bar. The experiments were run at 300 bar to achieve a density close to that of real heavy oil, which is still a liquid at 400 °C. The initial mole fractions when simulating the model were identical with those measured experimentally (see Appendix A) and included trace amounts of decane, decylbenzene, and undecylbenzene. The mechanism was simulated out to nine hours. Rate-of-production (ROP) analysis was also performed with Cantera, using the same conditions. The model was also simulated at 350 °C and 300 bar out to 239 h to compare against the corresponding experimental results. The initial conditions can be found in Appendix A.

2.2.8 Error Analysis for Model Simulation

There are three notable sources of error in the Cantera simulation due to simplifications made in the simulation: using an isothermal reactor, treating the species as ideal gases, and

ignoring possible diffusion limitations. However, as discussed below, none of these sources of error are expected to be significant compared to the two dominant sources of model error: parameter error from inaccurate estimates of reaction rate coefficients and thermochemistry as well as mechanism truncation error.

Isothermal Reactor Assumption

For the assumption that the reactor is isothermal, it has been observed using this experimental apparatus that it takes the contents of the reactor about 15 min to reach thermal equilibrium. This was determined by observing how long it takes the volume of water delivered by the syringe pump to equalize during the initial temperature ramp-up; the syringe pump removes water from the reactor as the water in the reactor expands to maintain a pressure of 300 bar. Since the 15 min temperature ramp-up is small compared to the six hour experiment, it was neglected during simulation. However, right-only horizontal error bars were added to all plots to account for this temperature ramp-up.

Ideal Gas Assumption

At 400 °C and 300 bar, the molar density of an ideal gas is only 187 cm^3/mol. For the starting material under these conditions, this equates to a density of 0.9 g/cm^3 if it behaved as an ideal gas. While the system pressure is above the critical pressure of every species in the mixture and the temperature is above most of the species critical temperatures, the mixture is definitely closer to a condensed phase than a gas.

It would be better to simulate this system using an equation of state that is more accurate at high pressures, like the Peng-Robinson equation of state (PREOS), which handles hydrocarbon species particularly well. Unfortunately, this requires knowing the critical properties of every species in the mixture (Tc, Pc, and the acentric factor, ω), which is not possible for all of the 1 326 species in our model, and even then, performing such a simulation can be quite difficult. Despite this, we can still make reasonable estimations for how much the assumption of ideal gas behavior will affect the simulation results.

For this analysis, we derived the relevant kinetic and thermodynamic equations in full rigor to determine how much assuming ideal gas behavior would affect the results. Since this analysis is long, we will only summarize the final results here and include the full detailed analysis in Appendix A. For more details about real gas effects, Kogekar et al. provided

an excellent discussion and derivation of these effects in their work and arrived at similar results to the equations we derived.[39] In short, when plotting concentration profiles in terms of mass fractions (as opposed to molar concentrations), accounting for nonideal contributions is entirely contained in one term: the ratio of fugacity coefficients of the transition state to those of the reactants (e.g., $\hat{\phi}^{\ddagger}/\hat{\phi}_{reactant}$). This is similar to accounting for solvation effects in liquid phase by including activity coefficients (e.g., $\gamma^{\ddagger}/\gamma_{reactant}$). Under the conditions in this study, PREOS predicts that the fugacity coefficients are mostly a function of the size of the molecules, with large molecules having coefficients much less than 1. For unimolecular reactions, the reactant and the transition state are similar in size, so this effect largely cancels out. Even for bimolecular reactions though, there is still some cancellation between the reactants and transition state, as the number of atoms must balance between the two.

Given all of this, it is unlikely that our ideal gas assumption will induce errors larger than even a factor of 2 or 3, as we show in more detail in Appendix A. While these errors are significant, factor of 10 errors or larger are not uncommon when calculating rates *ab initio*, even when high accuracy quantum chemical methods are used (and many of the model parameters are estimated instead by lower accuracy methods). Furthermore, errors in estimating the barrier height become more pronounced at lower temperatures, such as those used in this work.

Diffusion Limitations

The simulations in this work use ideal gas rate constants as determined by RMG. While RMG does cap these rate constants to prevent them from exceeding the collision limit (if necessary), diffusion limitations were not taken into consideration when this model was generated (though RMG has the ability to do so for a selection of known solvents through its "LiquidReactor" functionality[6]). Therefore, it is important to verify that the rate constants used in our model do not exceed diffusion limitations of the system.

It would be difficult to calculate the diffusion rate constant for every bimolecular reaction in the system. Instead, we estimated a lower bound for diffusion rates for this system and showed that this rate is above every rate in our model (i.e. that diffusion is not rate limiting). To get a lower bound, we considered the recombination reaction of two dodecylbenzene radicals in a solution of dodecylbenzene. This reaction should have a lower diffusion rate than any reactions in our system, as the molecules are large, and using pure dodecylbenzene

as the solvent will result in a lower diffusivity than for the mixture in our system.

We assumed that the diffusivity of dodecylbenzene radicals would be similar to the diffusivity of dodecylbenzene, and thus calculated the self diffusivity of dodecylbenzene from the Mathur-Thodos equations,[40] which were shown to be applicable for dense gases and liquids within a factor of 2.[41] With this, we calculated a diffusion limited rate constant of $4.68 * 10^8 \frac{m^3}{mol*s}$ at 400 °C and 300 bar. However, the highest bimolecular rate constant (in the forward or reverse direction) in our model at 400 °C is only $1.94 * 10^8 \frac{m^3}{mol*s}$. Even if the calculated diffusivity was too high by a factor of 2, this lower bound for diffusion limited rate constants in our system would still be higher than even the fastest bimolecular reactions in the model. Furthermore, only a handful of reactions have rate constants near $1.94 * 10^8 \frac{m^3}{mol*s}$, all involving H+ species, which likely have a higher diffusivity than dodecylbenzene radicals. Therefore, even in the worst case, neglecting diffusion limitations will not result in significant errors.

2.3   Results and Discussion

The final reaction network contains 1 326 species and 116 045 reactions; to our knowledge, this is one of the largest reaction mechanisms published in the literature. The kinetic model is provided in CHEMKIN[42] format with an RMG species dictionary in Appendix A. As is typical of detailed pyrolysis mechanisms, the model includes reactions such as hydrogen abstractions, β-scissions, radical recombinations, disproportionations, and cyclizations. Briefly, we first compare our model's predictions against previously published experimental results on the pyrolysis of pure dodecylbenzene at 400°C. Simulating with dodecylbenzene as the only starting material yields major products such as toluene, ethylbenzene, pentane, hexane, decane, undecane, and undecene, which is consistent with previously published work.[20–22,26] The predicted trajectories are shown in Appendix A along with comparisons with some of the prior literature data.
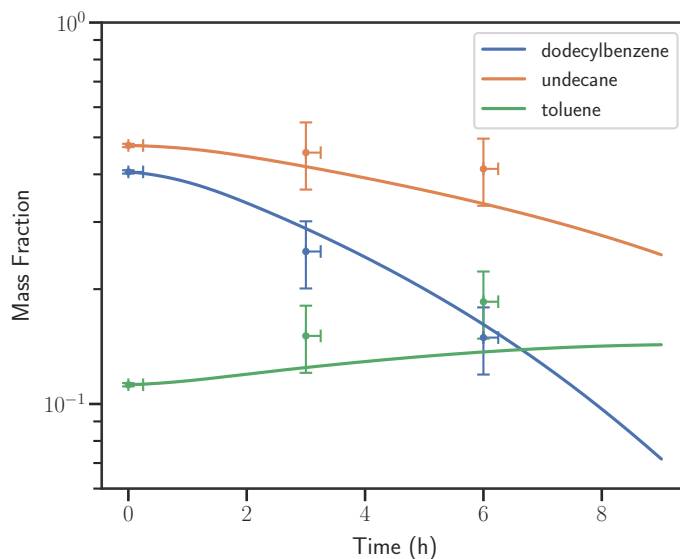
Figure 2.2. Conversion of main reactants at 400 °C. Model predictions show strong agreement with experimental measurements.

Our primary focus is to validate the model against the experimental pyrolysis of the surrogate mixture performed in this work at 300 bar and 400 °C, as well as analyze the relevant pathways. The measured initial conditions, as well as the data at the remaining time points, can be found in Appendix A. The conversion of the main reactants is shown in Figure 2.2. The model predictions are generally consistent with the experimentally observed concentrations, with most predictions falling within the error bars. Previous work claimed the main reaction pathway for the alkylbenzene reactant was either homolysis to form benzyl and undecyl radicals[22,25] or a four-membered "retro-ene" reaction.[27] As discussed in more recent work, the former reaction is quite slow, and the transition state for the latter reaction is highly strained, leading to extremely slow kinetics.[28] Instead, rate-of-production (ROP) analysis shows hydrogen abstraction as the main reaction family consuming dodecyl-benzene. As shown in Figure 2.3, in this surrogate mixture, dodecylbenzene mainly reacts with various undecyl radicals to form undecane and a dodecylbenzene radical. While the reaction to form the resonantly-stabilized phenyldodec-1-yl radical has the highest rate coefficient, that reaction quickly becomes partially equilibrated so its net rate is small. Instead, the fastest effective reaction sequence is hydrogen abstraction from dodecylbenzene to form phenyldodec-3-yl, which we will subsequently refer to as RAD3 since the radical is on the third carbon on the 12-carbon alkyl chain. RAD3 then undergoes β-scission, decomposing to 1-undecene and benzyl. There are many reactions, each with relatively high flux, in which

the benzyl radicals react with undecane or dodecylbenzene – the two main components in the mixture – to produce toluene and the corresponding alkyl or phenylalkyl radicals. The reaction pathways are summarized in Figure 2.4.
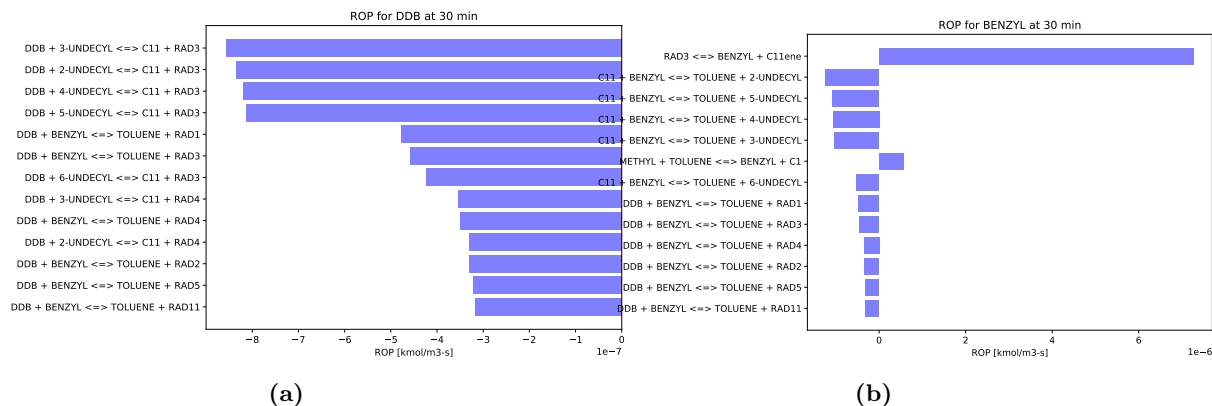
Figure 2.3. Rate-of-production of the highest flux reactions at 400 °C at 30 min for (a) dodecylbenzene (DDB) and (b) benzyl radicals.
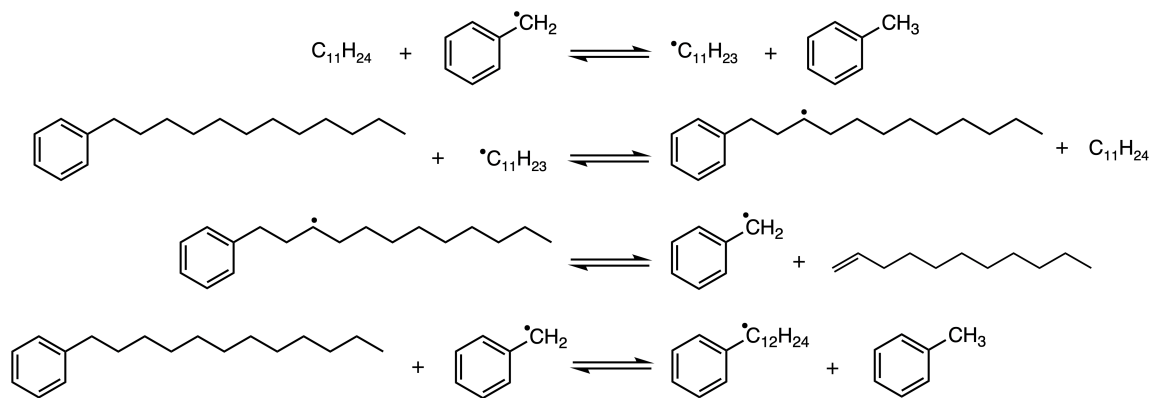
Figure 2.4. Dodecylbenzene is primarily consumed by hydrogen abstraction reactions.

When inspecting Figure 2.3a, it may seem unexpected for RAD3 to be highly preferred over RAD4, RAD5, etc., during the decomposition of dodecylbenzene. Indeed, the model's rate coefficients for producing RAD2 through RAD11 are identical; there is nothing special about hydrogen abstraction at the third carbon. Instead, decomposition of dodecylbenzene through RAD3 has the highest flux since it has a uniquely fast decomposition channel, forming resonantly stabilized benzyl radicals along with undecene, as shown in Figure 2.3b. None of the other radicals has a fast decay channel, so all are partially equilibrated with low flux. The mass fraction of these radicals at 30 min is shown in Figure 2.5 as a representative example. As expected, the concentration of RAD1 is over an order of magnitude larger

29

than the other radicals due to its favorable resonance stabilization with the aromatic ring. This is consistent with previously published work.[28] RAD2 through RAD11 are present in roughly equal amounts, likely because they are all secondary radicals, which do not have access to resonance stabilization. ROP analysis confirms that RAD12 is produced in smaller quantities than the other corresponding dodecylbenzene radicals. This is also expected since the radical is less stabilized when it is on the primary carbon on the alkyl chain, making this pathway less favorable.
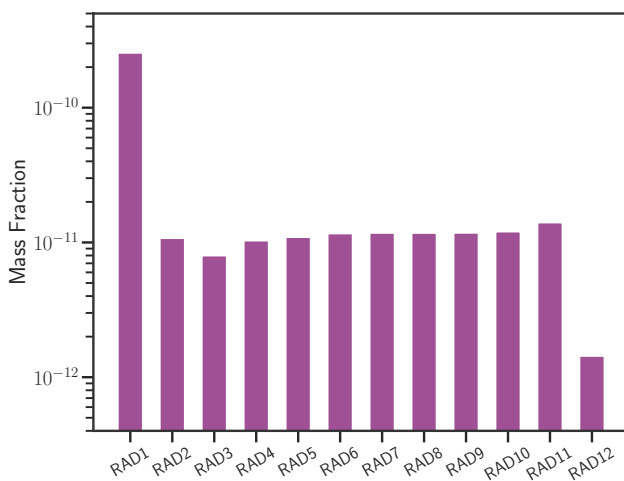
Figure 2.5. Mass fraction of dodecylbenzene radicals at 400 °C at 30 min. As expected, RAD1 is the most stable due to resonance with the aromatic ring.

As shown in Figure 2.6, undecane is primarily consumed by reacting with benzyl radicals to produce toluene and a set of undecyl radicals as shown in Figure 2.7; this is the same as the first reaction shown in Figure 2.4. Note that this is the main production pathway for toluene, though the last reaction template from Figure 2.4 also produces significant amounts of toluene.
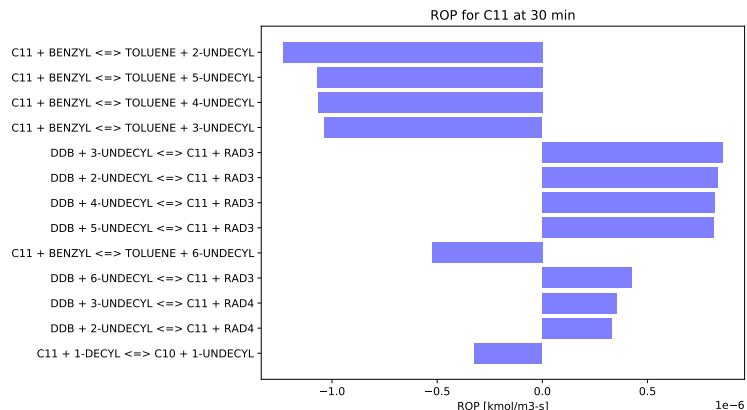
Figure 2.6. Rate-of-production of the highest flux reactions at 400 °C at 30 min for undecane.
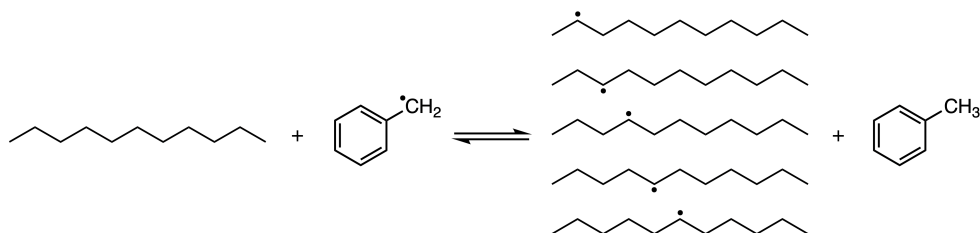
Figure 2.7. Undecane is primarily consumed by hydrogen abstraction reactions with benzyl.

The major products predicted by the model are shown in Figure 2.8. 10-Methylhenicosane is formed when an undecyl radical attacks undecene, whose product then abstracts a hydrogen from DDB or undecane. 4-Methyltridecane is formed when an undecyl radical attacks propene, whose resulting radical then abstracts a hydrogen from DDB or undecane. 2-Pentene is formed when an undecyl radical attacks pentene to form undecane and a pentyl radical, which then abstracts a hydrogen from DDB or undecane. The remaining major products and corresponding reaction pathways are explained in the following analysis.
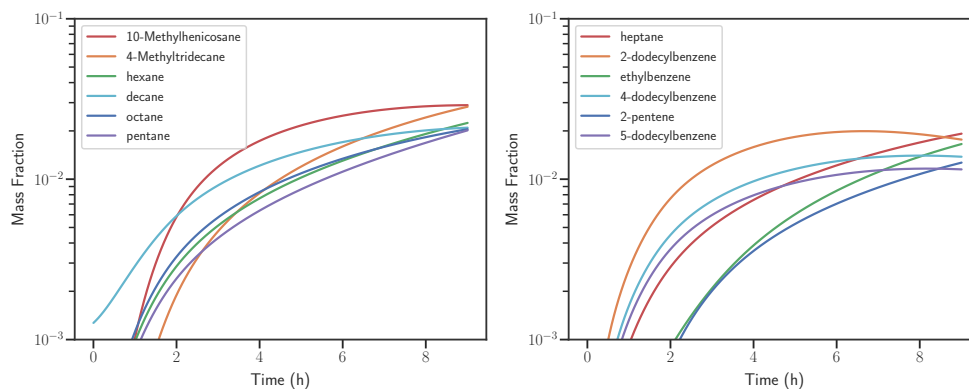
Figure 2.8. Major products predicted by the model when simulated at 400 °C and 300 bar.

Next, we validate model predictions against experimental results for different types of products and then explain the reaction pathways responsible for producing these products. We first compare straight-chain alkanes. As shown in Figure 2.9, the model predictions are generally within a factor of 2 of the experimental observations. The main reaction pathway producing alkane minor products in the model is shown in Figure 2.10 with representative ROP plots provided in Appendix A. It starts with undecane reacting with benzyl to produce various undecyl radicals; this first step is the same as shown in Figure 2.7. Next, undecyl radicals decompose to alkenes and primary alkyl radicals; since ethene is the smallest possible alkene, the alkyl radicals produced from this pathway contain up to nine carbons (ROP analysis shows that the decyl radical is primarily produced by the decomposition of RAD1 as opposed to the decomposition of undecane radicals). Finally, the alkyl radicals extract a hydrogen, primarily from undecane but also to a lesser extent from dodecylbenzene and toluene, to produce the normal alkane products.
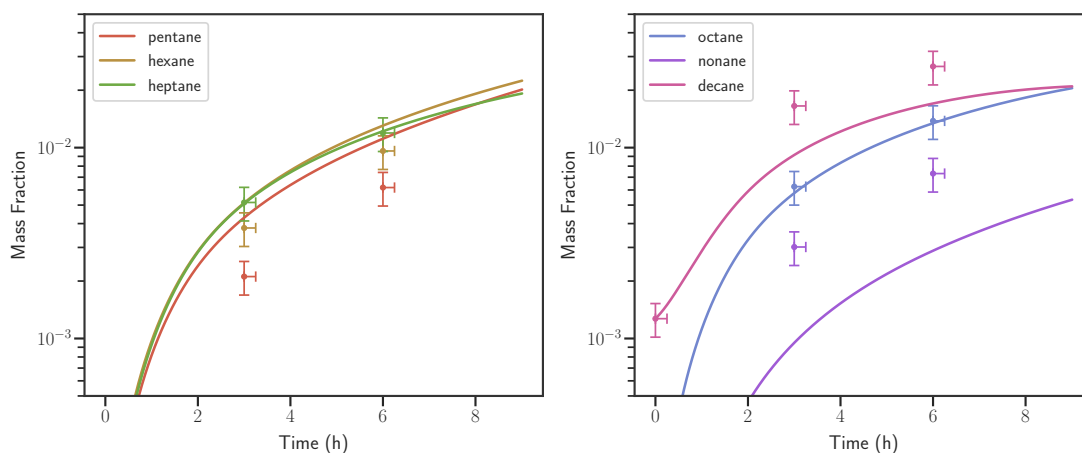
Figure 2.9. Model predictions and experimental measurements at 400 °C for C_5 through C_{10} alkanes.
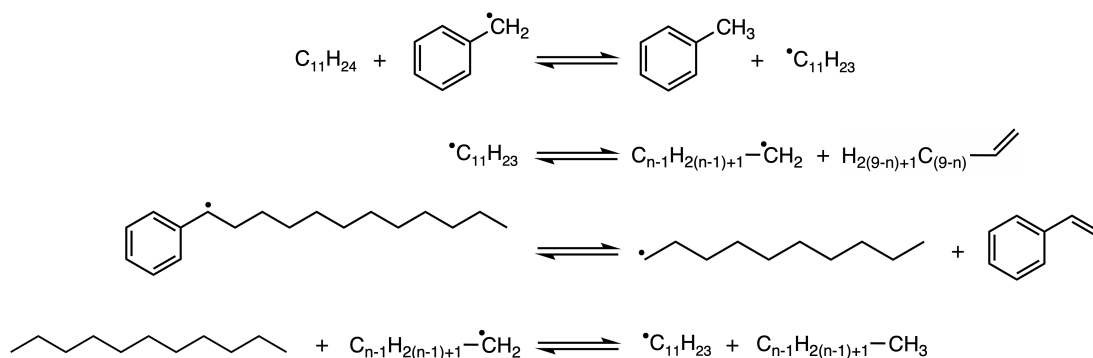
Figure 2.10. Primary production pathways for alkane minor products. Note that $2 \leq n \leq 9$.

Model predictions show that these saturated hydrocarbons are present in a higher fraction than their unsaturated counter parts; for example, there is more octane than octene, more heptane than heptene, etc, which is supported by other published work.[21] As large hydrocarbons pyrolyze to smaller products, one might expect alkanes and alkenes to be produced in equal amounts since that balances the hydrogens. Since this is not what we observe, the hydrogen atoms needed to produce the excess alkanes over alkenes must come from somewhere, and it is important to understand where these hydrogen atoms are coming from. Analyzing the RMG mechanism, this discrepancy is explained by multiple observations. First, the RMG mechanism predicts a sizable formation of alkanes larger than undecane, which would yield hydrogen atoms to the system for forming the smaller alkanes. This alone is not a sufficient source though, as the total amount of hydrogen atoms that could be generated from these larger alkanes is roughly 10% of the hydrogen atoms needed by the small n-alkanes. The other source of hydrogen atoms comes from producing various alkene and cyclic species. Although none of these species are as concentrated as any of the n-alkanes, there are simply many more alkenes, cyclic, and polycyclic species produced.
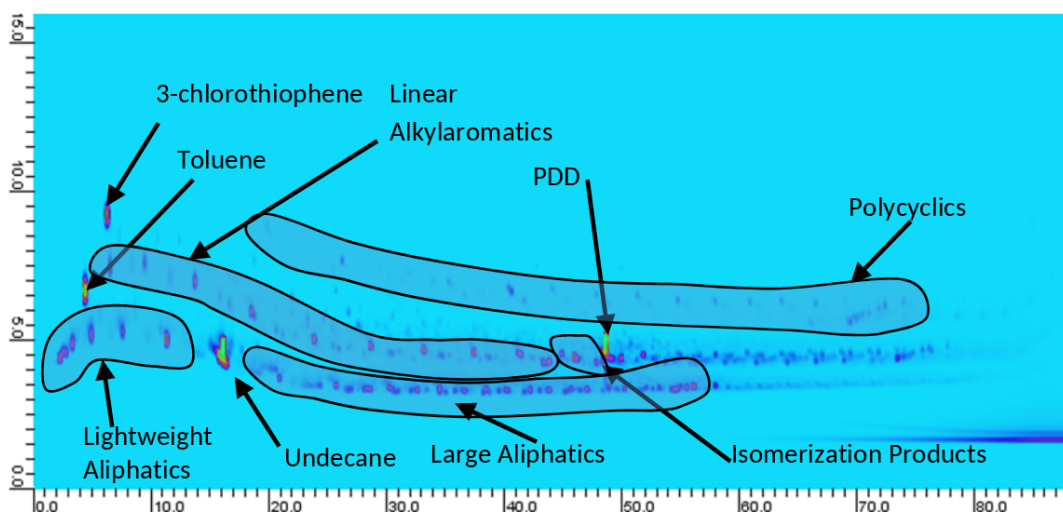
Figure 2.11. GC×GC-FID chromatogram of pyrolysis products after 6 h at 400 °C. In general, compounds with larger boiling points are farther to the right on the x-axis, and compounds that are more polar or have more aromatic rings are higher up on the y-axis. Regions for several different types of compounds are denoted on the plot. The smear in the bottom right-hand corner is column bleed.

There is some experimental evidence to support these predictions from the model. As seen in Figure 2.11, there are indeed many alkanes larger than undecane, and there are

some faint peaks above the lightweight n-alkanes corresponding to the lightweight alkenes predicted by the model. There is even a noticeable amount of multiring species, including polycyclic aromatic species. While it is difficult to quantify, Figure 2.11 suggests that the contribution to the hydrogen atom balance from large aliphatics and polycyclic species is even more important than the model predicts. The experimental data also suggests a much wider diversity of large aliphatic compounds than the RMG model predicts. Different systems have observed similar effects regarding this hydrogen balance. For example, Patwardhan et al.[43] observed low alkene/alkane ratios while adding sulfides during alkane cracking. The authors noticed that hexadecane lacks sufficient hydrogen atoms to produce these many small alkanes and hypothesized that the remaining hydrogens must come from aromatization of other species. While studying catalytic polyethylene up-cycling, Zhang et al.[32] observed hydrogen atoms being released by the formation of cyclic species, such as cycloalkanes and tetralins.

Next, we looked at the alkylaromatic minor products. The model does a good job of predicting radical-induced isomerizations of dodecylbenzene. As shown in Figure 2.12, the model generally predicts the correct trend for which isomer is produced more than others. An example phenyl migration pathway to produce one of these dodecylbenzene isomers is shown in Figure 2.13. Phenyl migrations are primarily responsible for producing all species in Figure 2.12; this pathway is also corroborated by recently published work,[30] whose quantum calculations enabled RMG to automatically identify these as the important reactions. At first, it may seem unintuitive why both the model and experimental data show 2-dodecylbenzene as more concentrated than the other dodecylbenzene isomers since the molecule must pass through a strained three-membered ring as an intermediate. However, RAD2 cannot participate in the reactions from Figure 2.15. Instead, nearly all RAD2 radicals are consumed via the isomerization pathway in Figure 2.13, leading to larger amounts of the corresponding isomer.
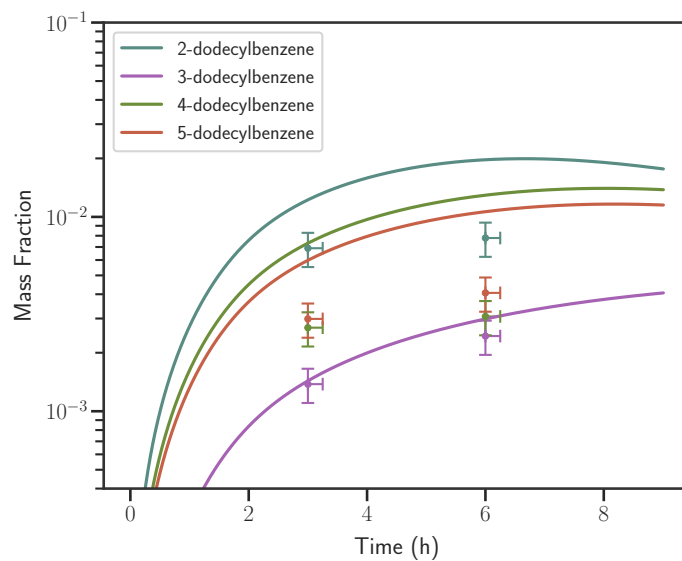
Figure 2.12. Model predictions and experimental measurements at 400 °C for branched alkylaromatic minor products, primarily formed from phenyl migrations in dodecylbenzyl radicals.
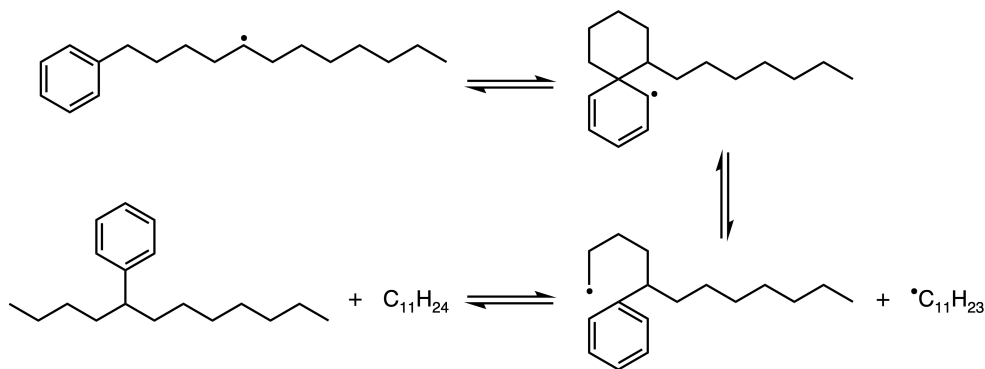
Figure 2.13. One of the phenyl migration pathways responsible for producing dodecylbenzene isomers.
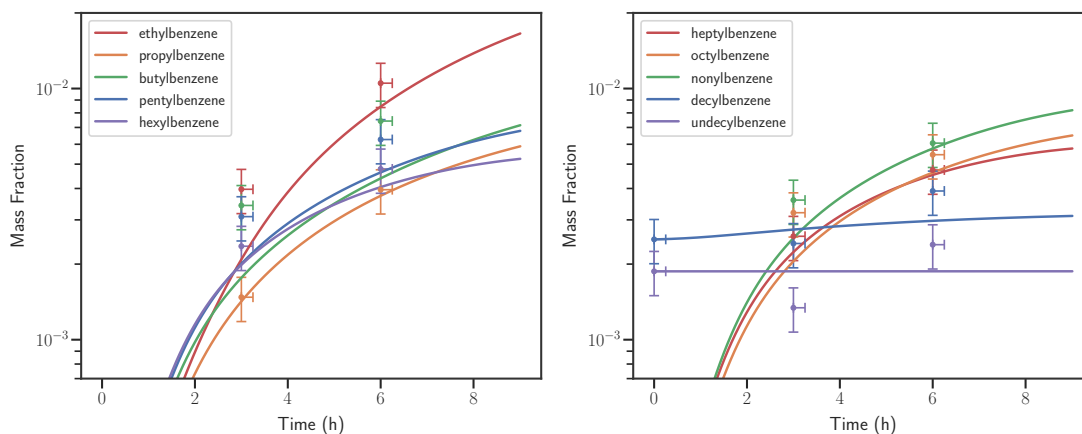
Figure 2.14. Model predictions and experimental measurements at 400 °C for alkylaromatic minor products, ethylbenzene through undecylbenzene.
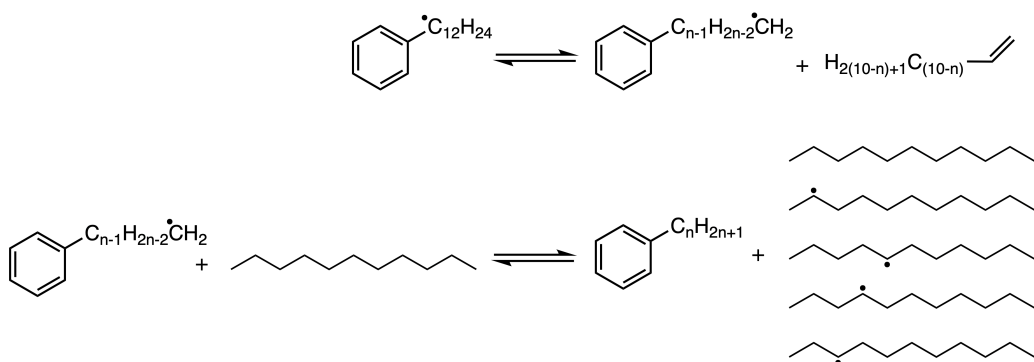
Figure 2.15. Main reactions responsible for producing straight-chain alkylaromatic products. Note that n is bounded by $1 \leq n \leq 10$ and represents the length of the alkyl chain e.g. n=2 corresponds to ethylbenzene.

As shown in Figure 2.14, straight-chain alkylaromatics are also predicted well by the model. Some predictions are within the experimental error bars, and all predictions are within a factor of 2. The dominant reaction pathway producing these straight-chain alkylaromatics products is shown in Figure 2.15, with representative ROP plots shown in Appendix A. Recall that Figure 2.5 showed that RADX, such that $3 \leq X \leq 11$, had similar steady state concentration. Each of these can β-scission to form an alkene along with a primary phenyl alkyl radical. Many of these radicals abstract a hydrogen atom to form the alkylbenzene following the template in Figure 2.15, which is simply a generalized form of the RAD3 decomposition shown as the third reaction in Figure 2.4; note that n = X − 2, so for example, RAD10 is responsible for producing octylbenzene and RAD3 is responsible for producing toluene. RAD2 and RAD1 cannot participate in these reactions. As

noted above, RAD2 is instead primarily consumed via the phenyl migration pathway to produce 2-dodecylbenzene while RAD1 mostly undergoes β-scission to produce decyl and styrene. Recall from Figure 2.5 that RAD12 is present in smaller quantities than the other corresponding dodecylbenzene radicals. Since the steady-state concentration of RAD12 is an order of magnitude smaller, the RAD12 flux through the pathway outlined in Figure 2.15 is also smaller and much less decylbenzene is formed compared with the shorter-chain alkylbenzenes. There is also no fast reaction from the starting materials to undecylbenzene since RAD13 cannot be formed from hydrogen abstraction of dodecylbenzene. Thus, much of the measured decylbenzene and undecylbenzene is due to impurities in the initial dodecylbenzene.

Briefly, some experiments were also performed at 350 °C to validate the model at lower temperatures. As observed in Figure 2.16, the model's predictions strongly agree with the experimental observations for the conversion of main reactants. When comparing the normal alkanes in Figure 2.17, model predictions are generally within a factor of 2 of the experimental observations. As shown in Figure 2.18, the model generally predicts the correct trends for straight-chain alkylaromatics. ROP analysis shows that the highest flux reactions at 400 °C remain important at 350 °C.
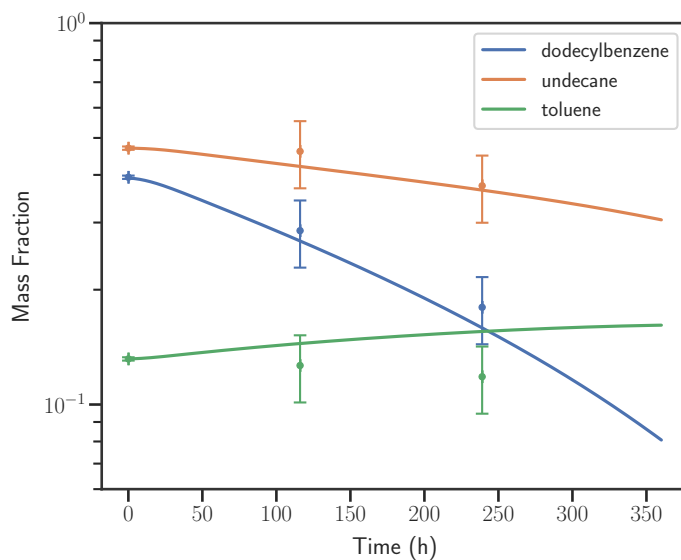
Figure 2.16. Conversion of main reactants over time at 350 °C.
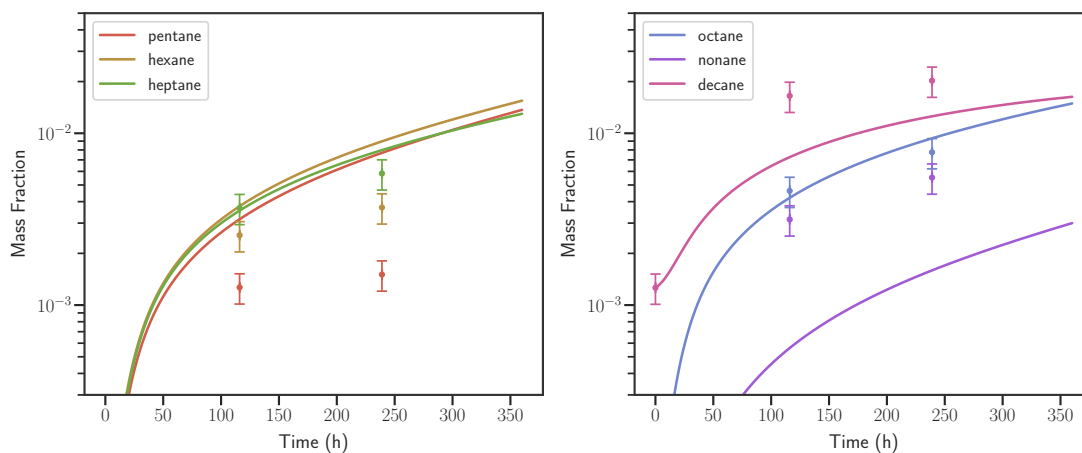
Figure 2.17. Model predictions and experimental measurements at 350 °C for C_5 through C_{10} alkanes.
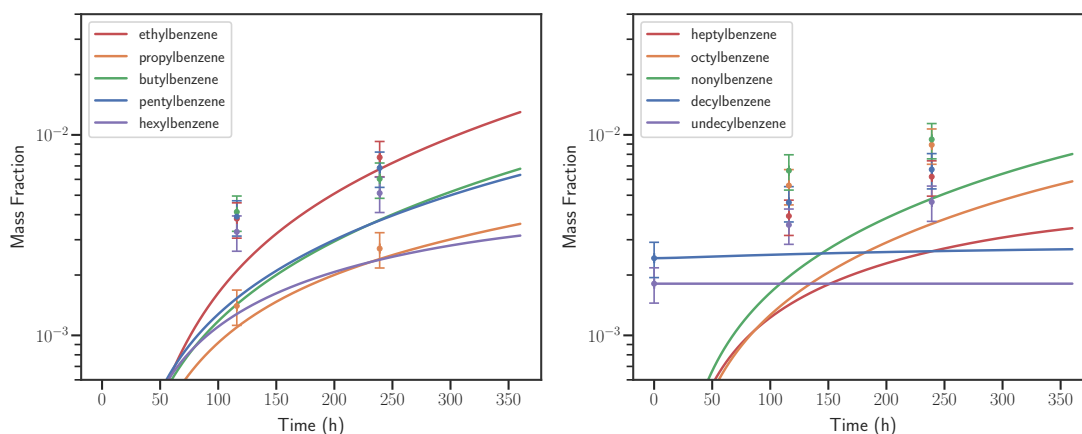
Figure 2.18. Model predictions and experimental measurements at 350 °C for alkylaromatic minor products, ethylbenzene through undecylbenzene.

It is important to emphasize that all parameters were obtained from *ab initio* calculations or group additivity using the RMG-database. No parameters were fit to the experimental data from this work or from any other work, yet nearly all products are predicted within a factor of 2, and many predictions fall within the experimental error bars. Recent work on low-temperature pyrolysis of alkylbenzenes, such as hexylbenzene,[44] almost exclusively plots results against conversion of the starting material; our model shows high-fidelity quantitative predictions for both conversion and selectivity (shown in Appendix A). More broadly, our results show the progress of the RMG software, which is capable of automatically creating detailed mechanisms for complex systems that have remained challenging, if not impossible, due to the combinatorial explosion in the number of possible reaction sites and reaction

intermediates when reacting large species.

2.4   Conclusions

We used the Reaction Mechanism Generator (RMG) to automatically develop a detailed kinetic model for the pyrolysis of a mixture of an alkylbenzene, an alkane, and an aromatic, with a feed molecule as large as C_{18}. The resulting model had 1 326 species and 116 045 reactions. The model was validated against pyrolysis experiments at 300 bar and both 350 °C and 400 °C with the same three-component surrogate mixture. Nearly all model predictions are within a factor of 2 of the measured value, and many predictions fall within the experimental error bars. The high-accuracy predictions, particularly the dodecylbenzene isomerization pathways, were greatly aided by recent quantum chemical calculations.[30,31] Recent work on low-temperature, high-pressure pyrolysis of alkylbenzenes, such as hexylbenzene,[44] almost exclusively plot selectivity. However, by accurately predicting the conversion, our model demonstrates significant improvement. We anticipate that our model will further the understanding of heavy oil pyrolysis and may be relevant to other systems, such as up-cycling polymer waste.[32] Importantly, no parameters were fit to experimental data; instead, this work highlights the increasing capability of the RMG software, which can automatically create detailed, high-fidelity mechanisms for complex chemical systems.

2.5   References

(1)   Frenklach, M.; Wang, H.; Yu, C.-L.; Goldenberg, M.; Bowman, C.; Hanson, R.; Davidson, D.; Chang, E.; Smith, G.; Golden, D.; Gardiner, W.; Lissianski, V. GRI-Mech 1.2, http://combustion.berkeley.edu/gri-mech/new21/version12/text12.html (accessed on 2021-02-03), 1995.

(2)   Frenklach, M.; Wang, H.; Goldenberg, M.; Smith, G.; Golden, D.; CT, B.; Hanson, R.; Gardiner, W.; Lissianski, V. *GRI-MECH – An optimized detailed chemical reaction mechanism for methane combustion*; tech. rep.; 1995.

(3)   Broadbelt, L. J.; Stark, S. M.; Klein, M. T. Computer generated pyrolysis modeling: on-the-fly generation of species, reactions, and rates. *Industrial & Engineering Chemistry Research* **1994**, *33*, 790–799.

(4)   Warth, V.; Battin-Leclerc, F.; Fournet, R.; Glaude, P.-A.; Côme, G.-M.; Scacchi, G. Computer based generation of reaction mechanisms for gas-phase oxidation. *Computers & chemistry* **2000**, *24*, 541–560.

(5) Vandewiele, N. M.; Van Geem, K. M.; Reyniers, M.-F.; Marin, G. B. Genesys: Kinetic model construction using chemo-informatics. *Chemical engineering journal* **2012**, *207*, 526–538.

(6) Gao, C. W.; Allen, J. W.; Green, W. H.; West, R. H. Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms. *Computer Physics Communications* **2016**, *203*, 212–225.

(7) Harper, M. R.; Van Geem, K. M.; Pyl, S. P.; Marin, G. B.; Green, W. H. Comprehensive reaction mechanism for n-butanol pyrolysis and combustion. *Combustion and Flame* **2011**, *158*, 16–41.

(8) Johnson, M. S.; Nimlos, M. R.; Ninnemann, E.; Laich, A.; Fioroni, G. M.; Kang, D.; Bu, L.; Ranasinghe, D.; Khanniche, S.; Goldsborough, S. S., et al. Oxidation and pyrolysis of methyl propyl ether. *International Journal of Chemical Kinetics* **2021**, *53*, 915–938.

(9) Zhang, P.; Yee, N. W.; Filip, S. V.; Hetrick, C. E.; Yang, B.; Green, W. H. Modeling study of the anti-knock tendency of substituted phenols as additives: an application of the reaction mechanism generator (RMG). *Physical Chemistry Chemical Physics* **2018**, *20*, 10637–10649.

(10) Jocher, A.; Vandewiele, N. M.; Han, K.; Liu, M.; Gao, C. W.; Gillis, R. J.; Green, W. H. Scalability strategies for automated reaction mechanism generation. *Computers & Chemical Engineering* **2019**, *131*, 106578.

(11) Han, K.; Green, W. H.; West, R. H. On-the-fly pruning for rate-based reaction mechanism generation. *Computers & Chemical Engineering* **2017**, *100*, 1–8.

(12) Quann, R. J.; Jaffe, S. B. Structure-oriented lumping: describing the chemistry of complex hydrocarbon mixtures. *Industrial & engineering chemistry research* **1992**, *31*, 2483–2497.

(13) Han, K.; Green, W. H. A fragment-based mechanistic kinetic modeling framework for complex systems. *Industrial & Engineering Chemistry Research* **2018**, *57*, 14022–14030.

(14) Malewicki, T.; Brezinsky, K. Experimental and modeling study on the pyrolysis and oxidation of n-decane and n-dodecane. *Proceedings of the combustion institute* **2013**, *34*, 361–368.

(15) Xu, R.; Wang, K.; Banerjee, S.; Shao, J.; Parise, T.; Zhu, Y.; Wang, S.; Movaghar, A.; Lee, D. J.; Zhao, R., et al. A physics-based approach to modeling real-fuel combustion chemistry–II. Reaction kinetic models of jet and rocket fuels. *Combustion and Flame* **2018**, *193*, 520–537.

(16) Han, S.; Li, X.; Zheng, M.; Guo, L. Initial reactivity differences between a 3-component surrogate model and a 24-component model for RP-1 fuel pyrolysis evaluated by ReaxFF MD. *Fuel* **2018**, *222*, 753–765.

(17) Liu, Y.-X.; Richter, S.; Naumann, C.; Braun-Unkhoff, M.; Tian, Z.-Y. Combustion study of a surrogate jet fuel. *Combustion and Flame* **2019**, *202*, 252–261.

(18) Liu, W.; Sivaramakrishnan, R.; Davis, M. J.; Som, S.; Longman, D.; Lu, T. Development of a reduced biodiesel surrogate model for compression ignition engine modeling. *Proceedings of the Combustion Institute* **2013**, *34*, 401–409.

(19)   Organization of the Petroleum Exporting Countries. *World Oil Outlook: 2045* **2020**, 1–145.

(20)   Blouri, B.; Hamdan, F.; Herault, D. Mild cracking of high-molecular-weight hydrocarbons. *Industrial & Engineering Chemistry Process Design and Development* **1985**, *24*, 30–37.

(21)   Savage, P. E.; Klein, M. T. Discrimination between molecular and free-radical models of 1-phenyldodecane pyrolysis. *Industrial & engineering chemistry research* **1987**, *26*, 374–376.

(22)   Behar, F.; Lorant, F.; Budzinski, H.; Desavis, E. Thermal stability of alkylaromatics in natural systems: kinetics of thermal decomposition of dodecylbenzene. *Energy & fuels* **2002**, *16*, 831–841.

(23)   Savage, P. E.; Korotney, D. J. Pyrolysis kinetics for long-chain n-alkylbenzenes: experimental and mechanistic modeling results. *Industrial & engineering chemistry research* **1990**, *29*, 499–502.

(24)   Savage, P. E.; Jacobs, G. E.; Javanmardian, M. Autocatalysis and aryl-alkyl bond cleavage in 1-dodecylpyrene pyrolysis. *Industrial & engineering chemistry research* **1989**, *28*, 645–654.

(25)   Billaud, F.; Chaverot, P.; Berthelin, M.; Freund, E. Thermal decomposition of aromatics substituted by a long aliphatic chain. *Industrial & engineering chemistry research* **1988**, *27*, 1529–1536.

(26)   Burklé-Vitzthum, V.; Michels, R.; Scacchi, G.; Marquaire, P.-M. Mechanistic modeling of the thermal cracking of decylbenzene. Application to the prediction of its thermal stability at geological temperatures. *Industrial & engineering chemistry research* **2003**, *42*, 5791–5808.

(27)   Carr, A. G.; Class, C. A.; Lai, L.; Kida, Y.; Monrose, T.; Green, W. H. Supercritical water treatment of crude oil and hexylbenzene: an experimental and mechanistic study on alkylbenzene decomposition. *Energy & Fuels* **2015**, *29*, 5290–5302.

(28)   Lai, L.; Gudiyella, S.; Liu, M.; Green, W. H. Chemistry of Alkylaromatics Reconsidered. *Energy & Fuels* **2018**, *32*, 5489–5500.

(29)   Dente, M.; Ranzi, E.; Goossens, A. Detailed prediction of olefin yields from hydrocarbon pyrolysis through a fundamental simulation model (SPYRO). *Computers & chemical engineering* **1979**, *3*, 61–75.

(30)   Khanniche, S.; Lai, L.; Green, W. H. Kinetics of Intramolecular Phenyl Migration and Fused Ring Formation in Hexylbenzene Radicals. *The Journal of Physical Chemistry A* **2018**, *122*, 9778–9791.

(31)   Lai, L.; Khanniche, S.; Green, W. H. Thermochemistry and group additivity values for fused two-ring species and radicals. *The Journal of Physical Chemistry A* **2019**, *123*, 3418–3428.

(32)   Zhang, F.; Zeng, M.; Yappert, R. D.; Sun, J.; Lee, Y.-H.; LaPointe, A. M.; Peters, B.; Abu-Omar, M. M.; Scott, S. L. Polyethylene upcycling to long-chain alkylaromatics by tandem hydrogenolysis/aromatization. *Science* **2020**, *370*, 437–441.

(33)   Lazar, C.; Cody, G. D.; Davis, J. M. A kinetic pressure effect on the experimental abiotic reduction of aqueous CO.sub.2 to methane from 1 to 3.5 kbar at 300 [degrees]C. *Geochimica et cosmochimica acta* **2015**, *151*, 34–.

(34) Liu, M.; Grinberg Dana, A.; Johnson, M. S.; Goldman, M. J.; Jocher, A.; Payne, A. M.; Grambow, C. A.; Han, K.; Yee, N. W.; Mazeau, E. J., et al. Reaction Mechanism Generator v3.0: Advances in Automatic Mechanism Generation. *Journal of Chemical Information and Modeling* **2021**, *61*, 2686–2696.

(35) Gao, C. W. Automatic reaction mechanism generation, Ph.D. Thesis, Massachusetts Institute of Technology, 2016.

(36) Liu, M.; Green, W. H. Capturing aromaticity in automatic mechanism generation software. *Proceedings of the Combustion Institute* **2019**, *37*, 575–581.

(37) Han, K.; Jamal, A.; Grambow, C. A.; Buras, Z. J.; Green, W. H. An Extended Group Additivity Method for Polycyclic Thermochemistry Estimation. *International Journal of Chemical Kinetics* **2018**, *50*, 294–303.

(38) Goodwin, D. G.; Speth, R. L.; Moffat, H. K.; Weber, B. W. Cantera: An Object-oriented Software Toolkit for Chemical Kinetics, Thermodynamics, and Transport Processes, https://www.cantera.org, Version 2.4.0, 2018.

(39) Kogekar, G.; Karakaya, C.; Liskovich, G. J.; Oehlschlaeger, M. A.; DeCaluwe, S. C.; Kee, R. J. Impact of non-ideal behavior on ignition delay and chemical kinetics in high-pressure shock tube reactors. *Combustion and flame* **2018**, *189*, 1–11.

(40) Mathur, G. P.; Thodos, G. The self-diffusivity of substances in the gaseous and liquid states. *AIChE journal* **1965**, *11*, 613–616.

(41) Kutney, M. C. (C. Thermodynamic and transport property modeling in super critical water, eng, Ph.D. Thesis, 2005, pp 311–405.

(42) CHEMKIN-PRO, R. 15112, Reaction Design. *Inc., San Diego, CA* **2011**.

(43) Patwardhan, P. R.; Timko, M. T.; Class, C. A.; Bonomi, R. E.; Kida, Y.; Hernandez, H. H.; Tester, J. W.; Green, W. H. Supercritical water desulfurization of organic sulfides is consistent with free-radical kinetics. *Energy & fuels* **2013**, *27*, 6108–6117.

(44) Lai, L.; Pang, H.-W.; Green, W. H. Formation of Two-Ring Aromatics in Hexylbenzene Pyrolysis. *Energy & Fuels* **2020**, *34*, 1365–1377.

Chapter 3

Implementation and Comparison of Bond Additivity Corrections and Isodesmic Reactions for Thermochemistry Calculations

Chapter Abstract

It has long been possible to obtain accurate species thermochemistry, including $\Delta_f H^\circ_{298}$, from quantum chemistry calculations. While quantum chemical methods continue to improve, accuracy gains from these methods are only realized if the calculations for heats of formation are properly corrected. One such approach is to use bond-additivity corrections (BACs), such as those defined by Petersson et al.[2] or Anantharaman and Melius[3] to reduce residual errors present, while a different approach is to utilize isodesmic reactions to calculate heats of formation through these error canceling reactions. In this work, we implement both of these approaches into ARKANE, an open-source software that can calculate species thermochemistry calculations from a wide variety of quantum chemistry outputs files. As

part of this work, we compiled a diverse database of over 400 reference species from the available literature to use in fitting BACs or to participate in isodesmic reactions. In comparing the approaches, we found that both BAC types yield results of similar accuracy, though the type of BACs of Anantharaman and Melius seem to generalize better. Isodesmic reaction approaches can be calculated automatically as work of Buerger et al.,[4] but we found that this approach only yields similar accuracy to BACs if care is taken when choosing the participating reference species, and that this approach has the distinct disadvantage that it does not work if suitable reference species are not available.

3.1 Introduction

Obtaining high-accuracy species thermochemistry is important to solving many different problems in the chemical domain. High-accuracy data can be obtained from experimental data, which can yield data with accuracy within 0.1 kcal/mol or less. Similarly, many resources for this purpose exist such as the Active Thermochemical Tables (ATcT),[5,6] which aims to produce a self-consistent set of species thermochemistry with continually improving precision. If experimental or ATcT data is available for the species of interest this is the way to go to achieve highly accuracte data. However, these species represent only a tiny fraction of possible species under study. Given that experiments can be time consuming, and don't allow for high-throughput predictions, calculating species thermochemistry through quantum chemistry calculations can prove quite valuable.

As quantum chemistry methods continue to improve, higher accuracy data can be obtained for larger and larger species. For example, the domain-based local pairwise natural orbital (DLPNO) coupled cluster methods[7,8] promise to deliver coupled-cluster accuracy of less than 1 kcal/mol to molecules much larger than what has previously been possible. To realize the accuracy gains, though, it is well known in the literature that additional corrections are needed outside of the quantum chemistry calculations to reduce residual errors present from the given level of theory. As we show later in this work, if no additional corrections are used, the resulting heats of formation tend to have a fairly wide distribution of errors, often not centered around zero. In fact, such corrections have already been derived for the DLPNO methods mentioned here.[9]

There are two commmonly used approaches to these corrections. The first approach

involves the use of bond additivity corrections. These methods assume that the residual error from the quantum chemistry calculations can be tied to residual errors that are roughly the same for given types of bonds present in a molecule. A corrective value can then be fitted for each of these types of bonds, and the total correction is simply the sum of these corrections for every bond in the molecule. In this work we focus on two popular implementations of this: those of Petersson et al.[2] and those of Anantharaman and Melius.[3]

A separate approach involves the use of isodesmic reactions, where species participating in the reaction other than the target are calculated at the same level of theory but also have precisely known heats of formation. This approach has been used to calculate accurate heats of formation for large molecules such as C60 and corannulene.[10] Recent breakthroughs have been made in this approach, with the work of Buerger et al. showing how isodesmic reactions can be found using a computer rather than having them be conceived by hand.[4] This allows this approach to be used in high-throughput calculations when implemented into statistical mechanics or thermochemistry software.

One problem we have often encountered is that the choice of reference species used in these approaches can vary from study to study. This can cause issues when BACs fitted to a small set of species are used for species that are not well represented by the original set (for example, using BACs fitted to only stable species for use on radical species). Finding the exact species used in the fitting can also be difficult. When we decided to implement these approaches in order to compare them, we also wanted the implementation to address these concerns.

In this work, we implemented Petersson-type and Melius-type BAC approaches, along with an isodesmic reaction approach based on the work of Buerger et al. into the open-source statistical mechanics, kinetics, and thermochemistry software ARKANE.[11] With this implementation, we compiled a database of reference species from the literature for use in fitting BACs and for participating in isodesmic reactions. We also added scripts for simplifying the process of adding in new BACs for additional levels of theory into ARKANE. Finally, we compared these approaches to see which approach is preferred.

Table 3.1. Properties of Reference Species

Property	Number of Species	Percent of Species [%]
Contains H	350	83.1
Contains C	356	84.6
Contains N	98	23.2
Contains O	169	40.1
Contains F	39	9.3
Contains S	72	17.1
Contains Cl	42	10.0
Contains Br	15	3.6
Anions	24	5.7
Cations	11	2.6
Zwitterions	17	4.0
Multiplicity > 1	73	17.3

3.2 Theoretical Methods

3.2.1 Reference Data

Enthalpy of formation at 298K data for 421 species was compiled from the available literature, taken from the Active Thermochemical Tables (ATcT),[5,6,12] Cioslowski et al.,[13] Benson,[14] National Institute of Standards and Technology Chemistry WebBook (NIST),[15] the Computer Analysis of Thermo Chemical data (CATCH) search and retrieval system of Pedley et al.,[16] Third Millennium Ideal Gas and Condensed Phase Thermochemical Database (3rd Mil.),[17] and the NIST Computational Chemistry Comparison and Benchmark Database (CCCBDB).[18] The species have a wide variety of chemical features, including a diversity of elements, charges, and multiplicities, as summarized in Table 3.1. The species are mostly smaller molecules, although about 100 species have five or more heavy atoms, with the largest species being adamantane (C10H16).

In addition to encompassing a diverse of chemical features, the species chosen all have very precisely known heats of formation. All species have an uncertainty of 1 kJ/mol or less, with the exception of a few sulfur containing species. The average uncertainty is 0.63 kJ/mol (0.15 kcal/mol), and the largest uncertainty is 2.3 kJ/mol for tetramethylthiourea, a larger sulfur containing species.

The data was compiled into a database of reference species (hereforth referred to as the "reference database") located within the larger RMG-database, available on GitHub at https://github.com/ReactionMechanismGenerator/RMG-database. The reference database

consists of YAML text files, one for each species, that contains information about the species structure (including an RMG adjacency list representation of the species) along with the heat of formation data from the literature. Some species have data from multiple sources, which are all included in the database, but by default the source with the lowest uncertainty is used for any computations.

In addition to reference data, each YAML file stores calculated heats of formation (298K, including calculated atom energy corrections) for various levels of theory, including those discussed in this work, with the plan of extending to additional levels of theory going forward. In this way, the YAML files contain enough information to fit/re-fit bond additivity corrections or to perform isodesmic reaction calculations using any of the species in the reference database.

The YAML format was chosen as a way to store the data in a human readable/writable manner while also allowing the data to be read in as native RMG objects in python. Adding data to these files therefore can also be done using the RMG/Arkane API, which contains simple functions for common tasks such as adding in data from quantum chemistry calculations for a new level of theory.

3.2.2 Quantum Chemistry Calculations

While there are many affordable quantum chemistry methods that can used to calculate thermochemistry for large molecules, this work focuses on a subset of methods that can approach or even reach chemical accuracy of 1 kcal/mol for molecules up to (and perhaps even surpassing) C40 in size. In particular, the ωB97M-V,[19] ωB97X-D,[20] and B3LYP-D3[21] DFT functionals were considered for their recent popularity in the literature as well as their purported accuracy. Additionally, the domain-based local pairwise natural orbital (DLPNO) coupled cluster methods were considered, specifically DLPNO-CCSD(T)[7] and DLPNO-CCSD(T)-F12.[8] These methods are an approximation on traditional coupled cluster calculations that allow for a near linear scaling in the coupled cluster iterations with minimal loss of accuracy. This promises to extend the "gold standard" chemical accuracy of coupled cluster methods to much larger molecules, as traditional coupled cluster methods usually become prohibitively expensive for even moderately sized molecules.

ωB97M-V calculations were performed using QChem 5.2[22] using the default grid and an SCF convergence tolerance of 10^{-8}. ωB97X-D, and B3LYP-D3 calculations were performed

Table 3.2. Levels of theory used in this work

Label	Level of Theory
A	ωB97M-V/def2-TZVPD
B	B3LYP-D3BJ (Gaussian)/def2-TZVP
C	ωB97X-D/def2-TZVP
D	DLPNO-CCSD(T)/def2-TZVP // ωB97X-D/def2-TZVP
E	DLPNO-CCSD(T)-F12/cc-pvdz-F12 // ωB97M-V/def2-TZVPD
F	DLPNO-CCSD(T)-F12/cc-pvtz-F12 // ωB97M-V/def2-TZVPD

using Gaussian 16,[23] using tight settings for optimization and SCF convergence, along with an ultrafine grid. For B3LYP-D3, the empirical d3bj[24] dispersion model in Gaussian was employed. Orca 4.2.1[25,26] was used for calculating single point energies with DLPNO-CCSD(T) and DLPNO-CCSD(T)-F12. In all cases a tight SCF convergence was used (tightSCF), along with tightPNO tolerances for DLPNO. For DLPNO-CCSD(T) def2-TZVP was used as the basis set with a def2-TZVP/c auxillary basis set. For DLPNO-CCSD(T)-F12, cc-pVTZ-F12 was used as the basis set with an aug-cc-pVTZ/c auxillary basis sets and a cc-pVTZ-F12-CABS complimentary auxillary basis set. UHF was used for open-shell species such as radicals.

3.2.3 Statistical Mechanics Calculations

Heats of formation were calculated from single point energies and scaled[27] vibrational frequencies via Arkane[11] using the the rigid rotor harmonic oscillator approximation. While the details of performing these calculations is well covered in the literature (we direct the interested reader to the Gaussian thermochemistry white paper[28] for a great introduction), we will cover some of the details here to shed some light on particular decisions made that affect how we define terms like atom energy corrections.

To calculate heats of formation, some care is needed, as the energies usually outputted from quantum chemistry calculations have a different zero of energy than what is needed. For a given molecule, "M", the heat of formation at 298 K, $\Delta_f H(M, 298K)$, is defined as the difference in enthalpy between the molecule and its constituents elemental states (e.g. H_2, carbon (graphite), etc.). However, single point energies obtained from quantum chemistry calculation usually have charges separated at infinity as the zero of energy. Figure 3.1 shows how this difference in the zero of energy can be resolved to calculate heats of formation. To emphasize the conceptual meaning for each leg of the cycle, we have left out notation that

serves to balance the number of each type of element for each step to avoid cluttering the figure.
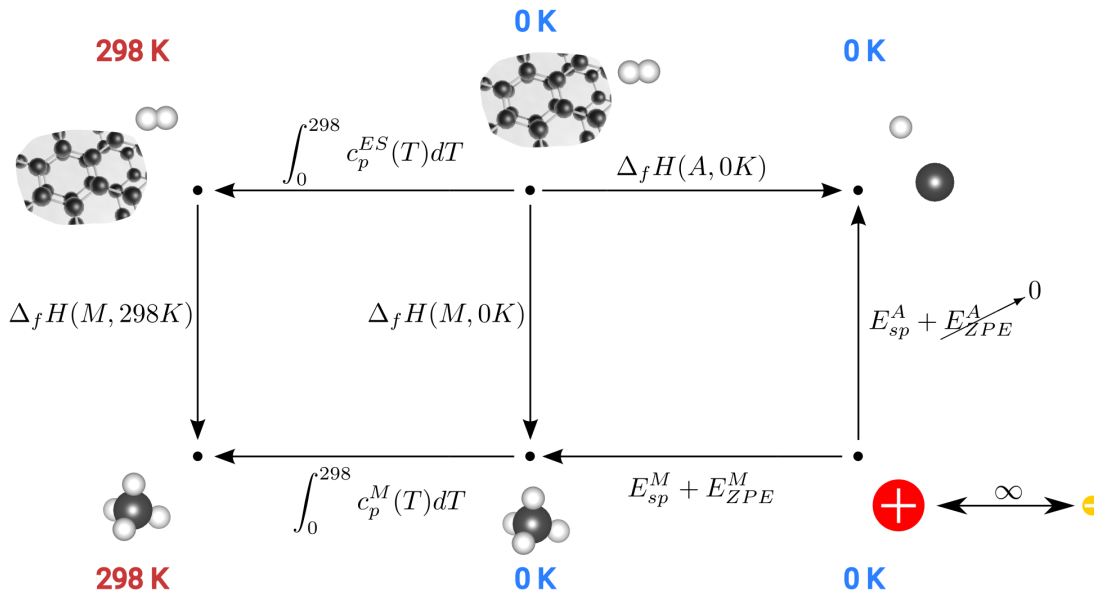
Figure 3.1. Conceptual diagram outlining the thermocycle used to calculate $\Delta_f H(M, 298K)$ from information that can be derived from a quantum chemistry output file or is known experimentally. In calculations performed in this work, E^A_{SP} is fully replaced by AECs. $\Delta_f H(M, 298K)$ can be further corrected by applying BACs, or use of isodesmic reactions schemes.

On the left-hand side of Figure 3.1, we have the desired quantity, $\Delta_f H(M, 298K)$, and in the bottom right-hand step we have the single-point energy of the molecule, E^M_{SP}, plus the zero point energy of the molecule, E^M_{ZPE}. $E^M_{SP} + E^M_{ZPE}$ is equal to the change in internal energy at 0 K of the molecule from charges separated at infinity. Because this is at 0 k, this is equal to the corresponding change in enthalpy (for consistency with the other steps of Figure 3.1). To deal with the difference in the zero of energy, we simply need to find the energy of the elemental states relative to charges at infinite separation. While this could be done by calculating single-point and zero-point energies for the elemental states, in practice this is not done, as some of the elemental states are difficult to calculate (most notably graphite) from quantum chemistry software. Instead, experimental values for the heats of formation of the atoms at 0 K, $\Delta_f H(A, 0K)$ are used to convert from the elemental states to atoms, which are much easier to calculate in quantum chemistry software. Finally, dealing with the remaining temperature difference is just a matter of accounting for the difference in heat capacities between the elemental states and the molecule from 0 K to 298 K. With

this, all legs of the cycle are either known experimental quantities or can be calculated from the quantum chemistry output file.

Figure 3.1 requires values for the single point energies of atoms, E_{SP}^A, which could be calculated separately using quantum chemistry calculations. In practice, though, we have found that this can lead to inaccurate values. Instead of using E_{SP}^A, fitted values known as atom energy corrections (AECs) are used, with different values for every single level of theory. It our use case the term "atom energy corrections" is misleading, as these values are not corrections applied on top of E_{SP}^A (an approach some authors use) but instead a full replacement for E_{SP}^A.

In this work, fitting AECs involved a linear least-squares fitting to minimize the error in calculating $\Delta_f H(M, 298K)$ for a small subset of molecules ranging across all elements: H_2, N_2, O_2, S_2, F_2, Cl_2, Br_2, HF, HCl, HBr, H_2S, H_2O, CH_4, NH_3, $ClCH_3$, and methyl radical. These species were chosen because they have precisely known heats of formation (or by definition exactly known) and are small enough such that the risk of additional errors in calculating any of these species at any one level of theory is low (e.g. no risk of converging to the wrong conformer). To further minimize errors from the chosen level of theory, experimental geometries and frequencies were used for this fitting.

Using these AECs allows for calculating $\Delta_f H(M, 298K)$ for molecules at the given level of theory. However, significant residual error usually remains if no further corrections, such as bond additivity corrections, are applied.

3.2.4 Bond Additivity Corrections

In this work we considered two different types of bond additivity corrections (BACs): BACs similar to those used by Petersson et al.[2] ("Petersson-type BACs"), and those used by Anantharaman and Melius[3] ("Melius-type BACs"). In both cases, BAC values are applied per bond as a correction to $\Delta_f H(M, 298K)$ with only AECs applied. BAC parameters for each type were fit for every level of theory used by fitting to all reference species in the database to minimize errors between the corrected $\Delta_f H(M, 298K)$ and the known experimental values.

Petersson-Type BACs

Petersson-type BACs assign a single corrective value for every type of bonds present in the 2D representation of the molecule. These bond types are defined pairwise for every
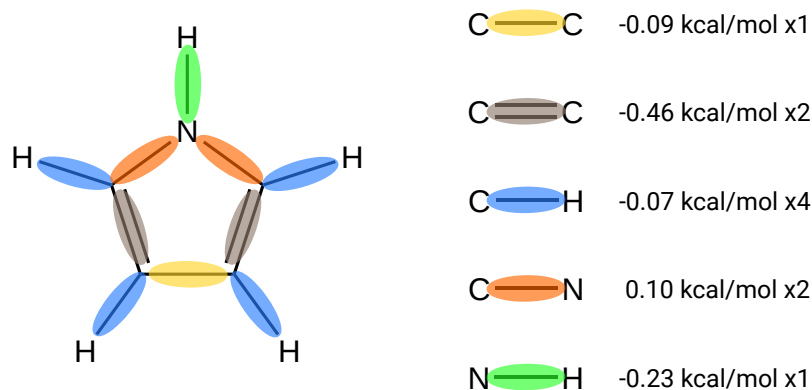
Figure 3.2. Application of Petersson-type BACs to a 2D representation of pyrrole for calculating corrections to $\Delta_f H(\mathrm{pyrrole}, 298K)$.

combinations of elements, including all integer bond orders. For example, common bond types include the carbon-carbon single bond (C-C), the carbon-carbon double bond (C=C), and the carbon-hydrogen bond (C-H). Applying Petersson-type BACs then is just a matter of adding the corresponding corrective value to $\Delta_f H(M, 298K)$ for every single instance of all types of bonds in the molecule. For example, Figure 3.2 shows the process of applying Petersson-type BACs to pyrrole.

Note that because only integer bond length are considered, this scheme considers pyrrole to be made up of 2 standard C=C bonds and 1 standard C-C bond, even though pyrrole is aromatic. The fact that Petersson-type BACs do not treat aromatic bond differently nor consider the chemical environment around the bond is potentially a shortcoming of these types of BACs, though this trade off for the sake of simplicity does reduce the number of parameters needed. As an upper bound the number of parameters grows quadratically with the number of elements considered (for three bond orders the upper bound is $3(N+1)\frac{N}{2}$, where N is the number of elements), though the number of parameters needed is less, as not all bond orders are valid between all pairs of elements.

Melius-Type BACs

Anantharaman and Melius developed an alternative formulation for correcting $\Delta_f H(M, 298K)$ that unlike the approach of Petersson et al. incorporates bond distances from the 3D geometry of the calculated molecule and considers information about neighboring atoms. Equation (3.1) shows the details of calculating Melius-type BACs. We have adapted our own notation

here to help emphasize/distinguish some of the various parameters. For example, variables with subscripts (e.g. α_a) are fitted parameters (with one parameter per element or level of theory), while variables with superscripts (e.g. S^M) are properties of the atoms, bonds, or molecules such as spin and bond distances. ξ is a fixed parameter of $3\,\text{Å}^{-1}$ as recommended by Anantharaman and Melius.

$$\Delta_f H_{298}^{\circ}(M) - \Delta_f^{\text{BAC}} H_{298}^{\circ}(M) = \sum_{a \in M} \text{Cor}^{\text{atom}}(a) + \sum_{b \in M} \text{Cor}^{\text{bond}}(b) + \text{Cor}^{\text{molecule}}(M) \tag{3.1a}$$

$$\text{Cor}^{\text{atom}}(a) = \alpha_a \tag{3.1b}$$

$$\text{Cor}^{\text{bond}}(\text{bond}(x,y)) = \sqrt{\beta_x \beta_y} \cdot e^{-\xi R^{xy}} + \sum_{w \in N(x)\backslash y} [\gamma_w + \gamma_x] + \sum_{z \in N(y)\backslash z} [\gamma_z + \gamma_y] \tag{3.1c}$$

$$\text{Cor}^{\text{molecule}}(a) = K_{\text{LoT}} \cdot \left(S^M - \sum_{a \in M} S^a \right) \tag{3.1d}$$

Equation (3.1a) shows that Melius-type BACs have three additive correction components: one for all of the atoms a in the molecule M, one for all of the bonds b in M, and one final correction based on the molecule as a whole. The atom corrections, given by Equation (3.1b) applies a single corrective parameter (α_a) based on the element for atom a. Equation (3.1c) shows the correction applied to a single bond between atoms x and y. The first term depends on the bond length, R^{xy}, and fitted parameters β_x and β_y based on the elements of atoms x and y. The next two terms in Equation (3.1c) considers contributions from neighboring atoms (atoms that are either bonded to x and y but excluding both x and y), with fitted parameters such as γ_x based per element. Finally, Equation (3.1d) considers the spin of the molecule, S^M and the spins of the atoms S^a, and includes a fitted parameter K_{LoT} that only depends on the level of theory used. Given this, Melius type BACs require fitting $3N + 1$ parameters for a given level of theory considering N elements.

Because Equation (3.1) in non-linear in its fitted parameters, a non-linear least-squares fitting was used to fit these parameters for every level of theory. To do this fitting, a global optimization is performed using different randomly-generated initial values for the parameters in each of 10 iterations. Each iteration uses the `least_squares` function in `scipy.optimize`[29] using the Trust Region Reflective method[30] and a 3-point method for

calculating the Jacobian.

3.2.5   Isodesmic Reactions

Background

An alternative method to using AECs and BACs for calculating corrected heats of forma-
tion is to calculate heats of formation from constructed isodesmic reactions (hypothetical
reactions where the number and type of each atom and bond are conserved). Just like
using AECs and BACs, this approach also uses data from reference species to reduce the
systematic errors from the quantum chemistry calculations at a given level of theory.

In this approach, an isodesmic reaction is constructed containing the target species (we
take the convention that the target species is a reactant of the reaction) along with any
number of reference species as additional reactants and/or products. Both the target and the
included reference species have uncorrected heats of formation calculated at the same level of
theory, $\Delta_f^{\text{calc}} H_{298}^{\circ}(t)$ and $\Delta_f^{\text{calc}} H_{298}^{\circ}(i)$, respectively. Additionally, the reference species have
well known heats of formation from experimental data or high-accuracy quantum chemistry
calculations, $\Delta_f^{\text{ref}} H_{298}^{\circ}(i)$. This is enough information to calculate a value for $\Delta_f^{\text{iso}} H_{298}^{\circ}(t)$,
which is the heat of formation for the target species from isodesmic reactions that in a sense
has been corrected much like $\Delta_f^{\text{BAC}} H_{298}^{\circ}$.

The rationale here is that while quantum chemistry calculations might have systematic
errors in calculating heats of formation for a single species, these systematic errors will at
least partly cancel when calculating heats of reactions, especially when the number of each
bond type is conserved. The assumption of this approach then is that for isodesmic reactions
$\Delta_{rxn}^{\text{ref}} H_{298}^{\circ} \approx \Delta_{rxn}^{\text{cal}} H_{298}^{\circ}$. Because heats of reactions can be calculated from heats of formation,
we can rearrange this approximation to arrive at an approximation for $\Delta_f^{\text{ref}} H_{298}^{\circ}(t)$, which
is unknown. If this assumption holds then this value should be relatively corrected against
systematic errors.

$$\Delta_{rxn}^{\mathrm{ref}} H_{298}^{\circ} = \sum_i \nu_i \Delta_f^{\mathrm{ref}} H_{298}^{\circ}(i) \quad \approx \quad \Delta_{rxn}^{\mathrm{cal}} H_{298}^{\circ} = \sum_i \nu_i \Delta_f^{\mathrm{calc}} H_{298}^{\circ}(i) \tag{3.2a}$$

$$\sum_{i\backslash t} \nu_i \Delta_f^{\mathrm{ref}} H_{298}^{\circ}(i) - \Delta_f^{\mathrm{ref}} H_{298}^{\circ}(t) \approx \sum_{i\backslash t} \nu_i \Delta_f^{\mathrm{calc}} H_{298}^{\circ}(i) - \Delta_f^{\mathrm{calc}} H_{298}^{\circ}(t) \tag{3.2b}$$

$$\Delta_f^{\mathrm{ref}} H_{298}^{\circ}(t) \approx \sum_{i\backslash t} \nu_i \left[\Delta_f^{\mathrm{ref}} H_{298}^{\circ}(i) - \Delta_f^{\mathrm{calc}} H_{298}^{\circ}(i) \right] + \Delta_f^{\mathrm{calc}} H_{298}^{\circ}(t)$$

$$\tag{3.2c}$$

$$\therefore \Delta_f^{\mathrm{iso}} H_{298}^{\circ}(t) \equiv \sum_{i\backslash t} \nu_i \left[\Delta_f^{\mathrm{ref}} H_{298}^{\circ}(i) - \Delta_f^{\mathrm{calc}} H_{298}^{\circ}(i) \right] + \Delta_f^{\mathrm{calc}} H_{298}^{\circ}(t)$$

$$\tag{3.2d}$$

Equation (3.2) shows the calculation scheme starting with this assumption that errors somewhat cancel when calculating heats of reactions to arrive at a value for $\Delta_f^{\mathrm{iso}} H_{298}^{\circ}(t)$. Equation (3.2a) starts with this main assumption, and Equation 3.2b simply expands this by separating out the target species from the remaining reference species of the isodesmic reaction. Equation (3.2c) rearranges this result to yield an approximate value for the unknown $\Delta_f^{\mathrm{ref}} H_{298}^{\circ}(t)$, which Equation (3.2d) uses to define the calculation for $\Delta_f^{\mathrm{iso}} H_{298}^{\circ}(t)$.

A keen observer will note that the summation term in Equation (3.2d) can be viewed as the corrective term to the calculated heat of formation of the target species, and that this term serves a similar role to applying BACs. This connection in fact runs much deeper. Let us suppose that every species in the isodesmic reaction is composed of a set of definable atom and bond types (could be similar to the bond types of Petersson et al. or more complex), and that there is a systematic error from the quantum chemistry calculation associated with each of these atom and bond types. If the reaction under consideration was an isodesmic reaction with respect to these atom and bond types, then these systematic errors exactly cancel. However, since we have excluded the target species from the summation in Equation (3.2d), the atom and bond types present in the target molecule have not cancelled. In this way, the summation in Equation (3.2d) is essentially calculating AECs and BACs for the features present in the target molecule and then applying these corrections. This all to say that if, for example, you construct isodesmic reactions by considering the bond types of Petersson et al., then this scheme is mathematically equivalent to applying Petersson-type BACs, albeit with a much smaller training set.

Given this similarity, it is important to discuss the differences between BACs and isodesmic reaction schemes that can affect their performance. As mentioned previously, isodesmic reactions essentially use a much smaller training set to apply what are essentially another form of BACs. One disadvantage of this is that if a reference species included in the reaction has a large uncharacteristic error, this will effect the final result much more than if a larger reference set was used where this error would have a much smaller effect on the average. One solution to this is to take an average or median of values obtained from multiple isodesmic reactions. On the other hand, the species chosen in the reaction could be more similar in nature to the target than the majority of species in a given reference set, which could be important for better canceling of errors. Finally, constraints could be placed on the species of an isodesmic reaction to ensure that the bonds present in the reaction really are similar in nature to each other, and thus should have similar cancellation of errors. In this work, we considered such constraints, adapting those previously used in the literature and adding some additional constraints of our own.

Isodesmic reaction classes and constraints

As part of their work developing an implementation to apply isodesmic reactions for correcting thermochemistry calculations, Buerger et al.[4] defined four different classes of isodesmic reactions (labeled RC1 - RC4) of increasing rigor (more likely to result in better cancellation of errors). In this work, we only considered classes RC2-RC4, as RC1 is only an "error-canceling" reaction and not an isodesmic reaction (the number of each bond type is not necessarily conserved).

RC2 is what is typically thought of as an isodesmic reaction, where only the number and type of each bond is conserved without regard for the surrounding chemical environment. The bond types are equivalent to those defined by Petersson et al. and do not consider aromatic bond types. The next reaction class up, RC3, adds the further restriction that the total bond order on either side of the bond must also be conserved. For example, the carbon-carbon bond of ethane has a bond order of 3 (the bond in question is excluded from the total bond order) for both carbon atoms in the bond, while the carbon-carbon bond of propylene has a bond order of 3 on one side and 2 on the other. Therefore, these bonds are treated as separate types. Finally, RC4 add the further constraint that the identity of the neighboring atoms on either side of the bond (including the bond orders of those neighbors)

must be conserved. The interested reader should see Buerger et al.[4] for excellent examples and illustrations that make the distinctions between these reaction classes clear.

In this work, we have extended these reaction class definitions into further sub-types to address problems that we observed. For example, because the reference set in this work includes anionic, cationic, and even zwitterionic species, we found that these species would often appear in the calculations for non-charged targets. Given that charged species are more likely to have significant errors from quantum chemistry calculations, this inclusion likely can reduce importance. To solve this issue, we added the charged-constrained sub-class, which limits included species to have a formal charge no more anionic or cationic than what is present in the target species. Reactions that comply with this constraint are labeled with a subscript "c" (e.g. RC2_c, RC3_c, etc.).

Another potential issue we identified was the inclusion of ringed species, especially 3 or 4 member ring species, from the reference set. To solve this issue, the ring-constrained subclass was added (e.g. RC2_r) where the number and type of each sized ring is also conserved (the identity of the atoms in the ring does not matter, but the number of atoms does). Finally, we added a very restrictive sub-class called the "scope" sub-class (e.g. RC2_s) that excludes any reference species that have a feature not found in the target molecule.

Note that these sub-classes are not mutually exclusive (e.g. RC2_cr exists and constrains both charges and rings). Furthermore, RC2_cs, RC2_rs, and RC2_crs are not the same, as the charge and ring features are only included as a feature to exclude as out of scope if the charges or rings are being constrained.

Automatic Isodesmic Reactions

Among others, a great contribution of Buerger et al. was to show that isodesmic reactions can be found automatically using a computer algorithm rather than having to be conceived of by hand. The key insight they had was that finding isodesmic reaction can be view as a constrained optimization problem over the stoichiometric coefficients, ν_i, for the reference species in the reaction. The constraints can ensure that the requirements of the reaction classes are satisfied, while optimizing for ν_i determines which of the reference species should be included in the reaction. Given the size of the reference set, there are many different combinations of reference species that could be included while still satisfying the reaction class constraints. Given this, Buerger et al. chose an optimization scheme to try and find an

isodesmic reaction that includes the least amount of reference species and constraining features. Equation (3.3) shows the mixed-integer linear programming problem (MILP) defined by Buerger et al. to automatically find isodesmic reactions.

$$
\begin{aligned}
\min_{\nu} \quad & \sum_i^{N_{\text{ref}}} \left[|\nu_i| \cdot \sum_j^{N_{\text{constraints}}} c_{ij} \right] \\
\text{s.t.} \quad & \sum_i^{N_{\text{ref}}} \nu_i c_{ij} = 0 \quad \forall j \in \{1, \ldots, N_{\text{constraints}}\}, \\
& \nu_i > 0 \quad \forall i \in \{1, \ldots, N_{\text{ref}}\}
\end{aligned}
\tag{3.3}
$$

In Equation (3.3), ν_i are the integer variables being optimized for over all N_{ref} reference species. c_{ij} are the number of each constraining feature j present in reference species i. For example, if j is the index for a C-H bond in an RC2 scheme and i is for the reference species methane, $c_i j$ is 4. The constraints of this minimization problem ensure that the conditions of the reaction class are met. We should not that the problem as posed in Equation (3.3) is not entirely linear in ν_i given the inclusion of the $|\nu_i|$ term. However, this problem can be recast as a linear problem by assigning separate variables ν_i when the reference species participates as a reactant rather than a product (while taking care of the proper sign of these variables in the constraints). The minimization procedure ensure that no reference species will participate as a both a reactant and a product.

In this work, we implemented the algorithm of Buerger et al. as given in Equation 3.3 inside of Arkane, along with constraints for the various reaction classes described previously, in order to calculate values for $\Delta_f^{\text{iso}} H_{298}^{\circ}(t)$ for a given target species automatically. By default, our implementation takes the median value for $\Delta_f^{\text{iso}} H_{298}^{\circ}(t)$ from 10 isodesmic reactions. To solve the MILP, `lp_solve`[31] was used, although a wrapper to use Pyomo[32,33] was also implemented. In general, we found `lp_solve` to be significantly faster most of the time, though it was also prone to occasionally not converging to a solution. As a final implementation note, although any AECs or BACs applied to the reference species and target will ultimately cancel out, $\Delta_f^{\text{calc}} H_{298}^{\circ}$ does include AECs for both the target and the reference species. This is because the values of $\Delta_f^{\text{calc}} H_{298}^{\circ}$ for the reference species stored in the database include AECs so that the resulting values are atleast somewhat reasonable and close to the reference values.

Figure 3.3. Distribution of errors in heats of formation before and after applying Petersson-type BACs. The RMSE and MAE are leave-one-out cross-validation errors.

3.3 Results and Discussion

In order to compare the performance of the different types of BACs, both the training errors and leave-one-out cross-validation errors were considered for species in the reference set. Figure 3.3 shows the cross-validation errors for Petersson-type BACs, while Figure 3.4 shows the cross-validation errors for Melius-type BACs for all levels of theory (LoT) described previously. Additionally, these plots also show the distribution of errors for the calculated heats of formation prior to BACs being applied (i.e. with only AECs applied). On this note, one of the first thing to notice in these figures is that the distribution of errors prior to applying BACs is both wider than those with BACs applied and also not centered at zero. This shows clearly that systematic errors remain from the quantum chemistry calculations alone, which would result in a substantial loss of accuracy if it were not for applying BACs or other corrective methods. Secondly, as expected the mean absolute error (MAE) and root mean square error (RMSE) decrease as the level of theory is increased.

Melius-type and Petersson-type BACs seem to show similar levels of performance, and at first galnce it is not easy to determine which type of BAC performs better. We gain some insight, though, by considering both the training and cross-validation errors for these BACs. Table 3.3 breaks down both the training and cross-validation errors at each level of theory for easy comparison. Furthermore, the table includes a Δ column, which shows the
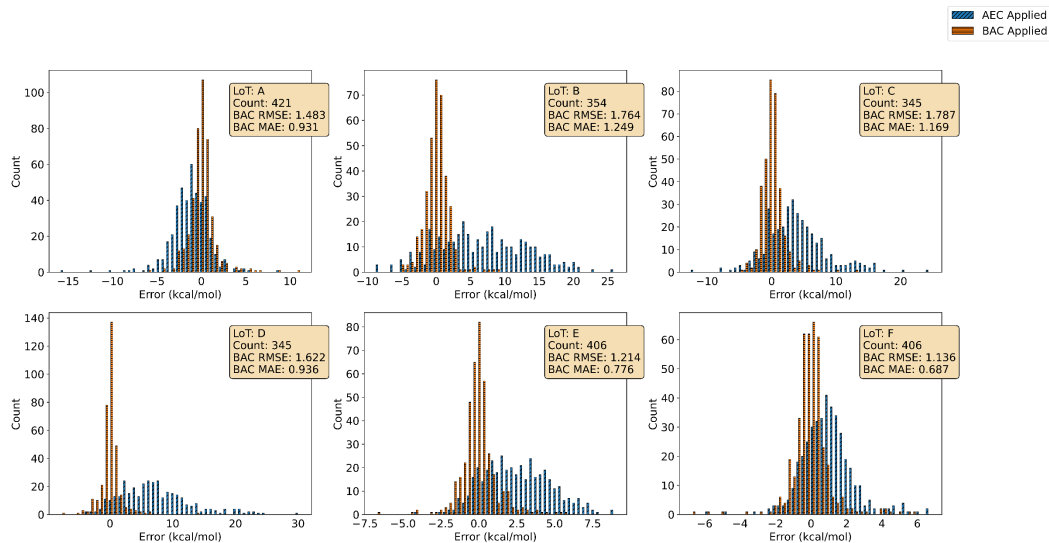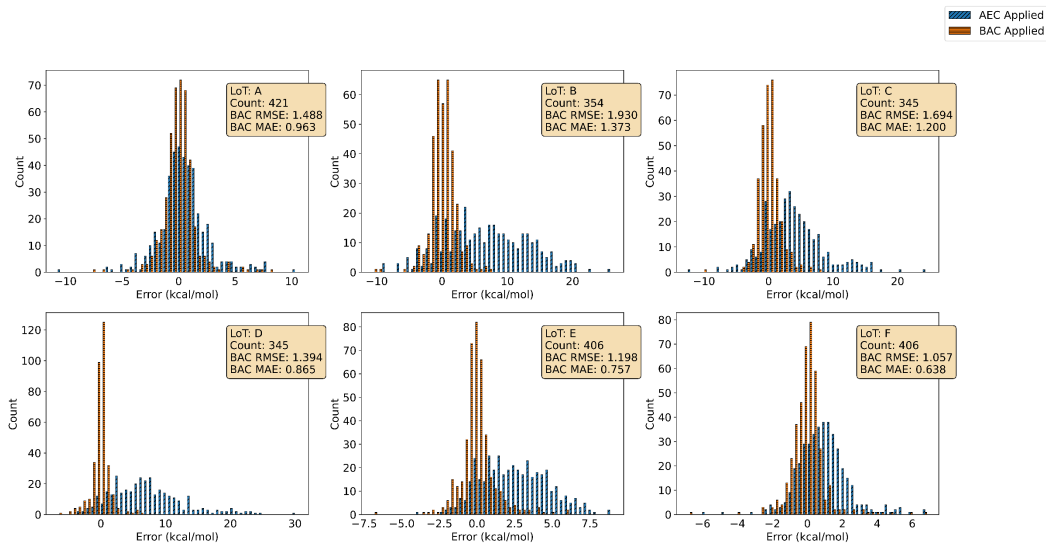
Figure 3.4. Distribution of errors in heats of formation before and after applying Melius-type BACs. The RMSE and MAE are leave-one-out cross-validation errors.

difference between these two errors.

As seen in Table 3.3, Petersson-type BACs almost always had lower training errors (both MAE and RMSE). However, Melius BACs more often than not had slightly lower cross-validation errors, and always had a smaller change in going from training errors to cross-validation errors. This suggests that Melius BACs might generalize better to types of molecules that are not well represented in the reference set, and that it might be easier to over-fit Petersson-type BACs. This makes intuitive sense, as Petersson-type BACs fit a single parameter to an fairly broad notion of a bond type (for example, treating C-C bonds in benzene and ethane as the same). It is therefore possible that Petersson-type BACs could be fit on a reference set containing (for example) C-C bonds from a reference set that does not contain many examples of the targets C-C bond. Melius-type BACs on the other-hand are inherently more flexible in distinguishing between bond types, though a diverse reference set is still crucial.

Overall, though, the difference in cross-validation errors between the types of BACs is not substantial, especially for higher levels of theory. The possibility of being able to generalize better and requiring slightly fewer parameters does give Melius-type BACs a slight advantage.

Anlyzing the performance of isodesmic reaction methods is a bit more involved, though. This can be done by calculating $\Delta_f^{\mathrm{iso}} H_{298}^\circ$ for every species in the reference set (temporarily

Table 3.3. Training and cross-validation errors for Petersson (P) and Melius (M) type BACs in kcal/mol. Δ is the increase in going from training errors to cross-validation errors. Values are bolded to indicated which type of BAC performed better.

LoT	BAC Type	Training Errors		Cross-Validation Errors		Δ	
		RMSE	MAE	RMSE	MAE	RMSE	MAE
A	P	**1.19**	**0.77**	**1.49**	**0.93**	0.30	0.16
	M	1.37	0.90	**1.49**	0.96	**0.12**	**0.06**
B	P	**1.86**	**1.21**	**2.18**	**1.43**	0.32	0.22
	M	2.11	1.48	2.37	1.62	**0.26**	**0.14**
C	P	**2.11**	**1.26**	2.47	1.49	0.36	0.23
	M	2.15	1.37	**2.29**	**1.47**	**0.14**	**0.10**
D	P	2.41	**1.20**	2.78	**1.45**	0.37	0.25
	M	**2.38**	1.38	**2.49**	1.46	**0.11**	**0.08**
E	P	**1.00**	**0.66**	1.21	0.78	0.21	0.12
	M	1.07	0.68	**1.20**	**0.76**	0.13	0.08
F	P	**0.89**	**0.57**	1.14	0.69	0.25	0.12
	M	0.97	0.59	**1.05**	**0.63**	**0.08**	**0.04**

excluding this species from the reference set while it is the target species) and then calculating the error as $\Delta_f^{\mathrm{iso}} H_{298}^{\circ}$ - $\Delta_f^{\mathrm{ref}} H_{298}^{\circ}$. This can be done not only for every level of theory but also reaction class as defined previously.

There are two complexities worth considering, though. First, not all species can be calculated using isodesmic reaction. For example, if a certain type of bond or constraint is only found in a small number of species in the reference set, it may not be possible to construct an isodesmic reaction using only the species in the reference set (excluding the target). This becomes more and more problematic for higher order reaction classes. Furthermore, sometimes the MILP solver did not converge onto a solution quickly enough, leaving some species without data. The second complexity is that while a certain reaction class might be specified, it is always possible that the median isodesmic reaction used for the final value for $\Delta_f^{\mathrm{iso}} H_{298}^{\circ}$ just so happens to be a member of a higher order reaction class. Figures 3.5, 3.6, and 3.7 show errors for each reference species that could be calculated with isodesmic reactions for reaction classes at least as strict as RC2, RC3, and RC4, respectively. The actual reaction class found are given different colors in stacked in the distribution.

Looking at Figure 3.5, we see that although a reaction class as low as RC2 was requested (and thus only RC2 type constraints were considered when using Equation (3.3), the median reaction found was always a higher order reaction class. In fact, the median reaction was always at least of class RC3 or higher. This is not entirely surprising, as the distinction
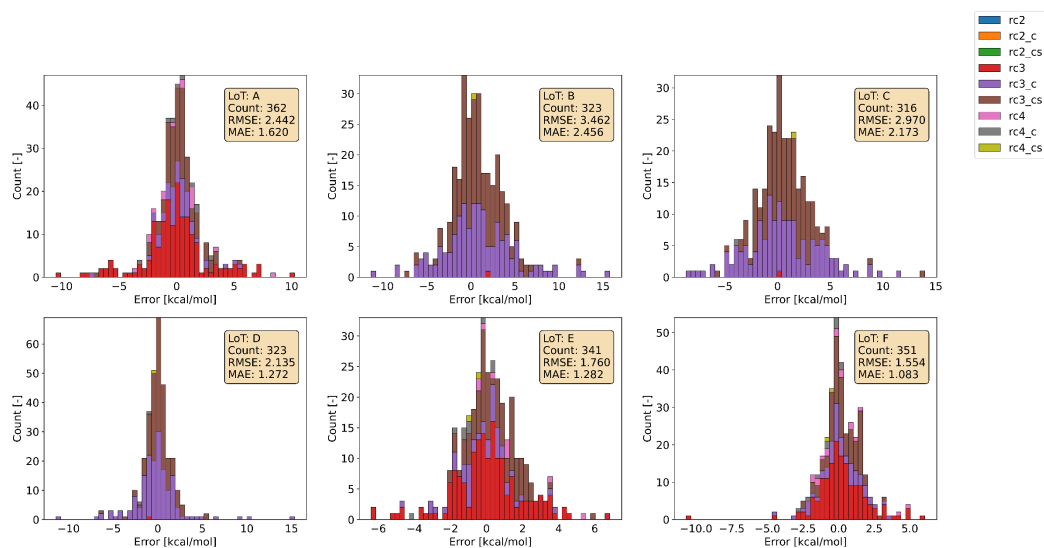
Figure 3.5. Errors between $\Delta_f^{\mathrm{iso}} H_{298}^\circ$ and known values of $\Delta_f^{\mathrm{ref}} H_{298}^\circ$ using species from the reference set and isodesmic reactions of class RC2 or stricter. The actual reaction class of found reaction are given different colors as seen in the legend
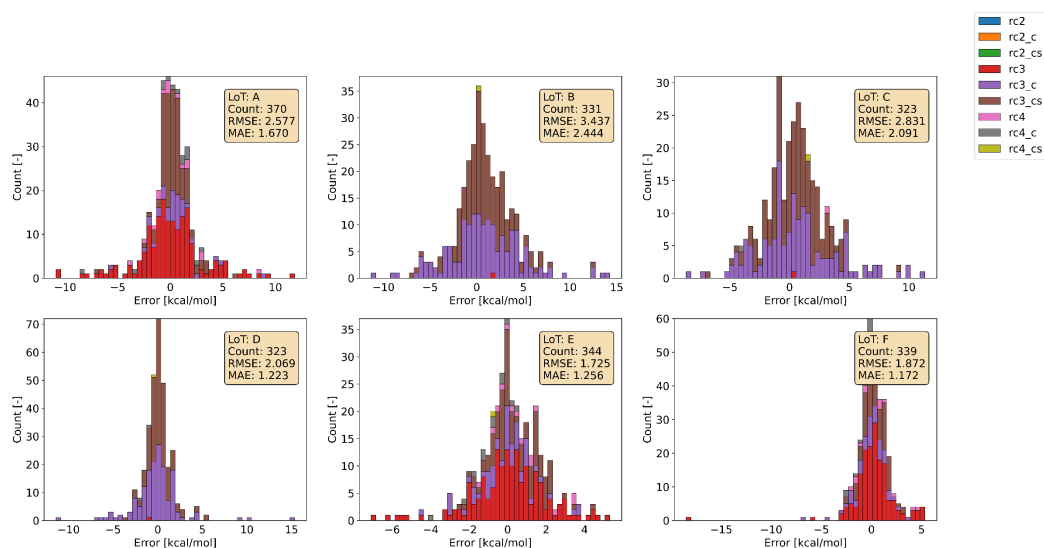
Figure 3.6. Errors between $\Delta_f^{\mathrm{iso}} H_{298}^\circ$ and known values of $\Delta_f^{\mathrm{ref}} H_{298}^\circ$ using species from the reference set and isodesmic reactions of class RC3 or stricter. The actual reaction class of found reaction are given different colors as seen in the legend

Figure 3.7. Errors between $\Delta_f^{\mathrm{iso}} H_{298}^\circ$ and known values of $\Delta_f^{\mathrm{ref}} H_{298}^\circ$ using species from the reference set and isodesmic reactions of class RC4 or stricter. The actual reaction class of found reaction are given different colors as seen in the legend
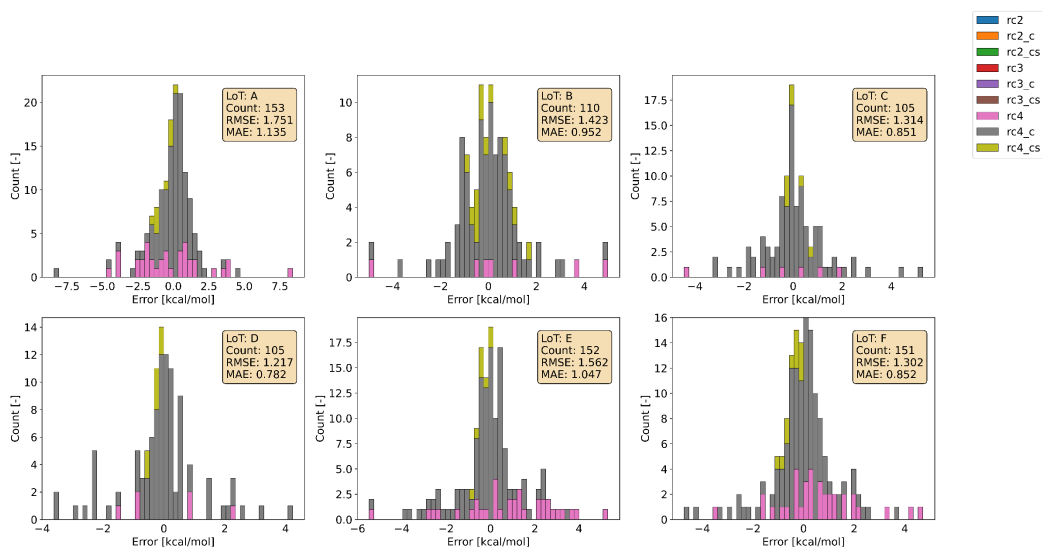
between RC2 and RC3 is small, with RC3 only considering the bond orders as additional information. Because of this, Figures 3.5 and 3.6 look very similar. While the more strict sub-classes were often satisfied as well in both cases, The number of RC4 class reaction found by chance was extremely small, as RC4 reactions are a comparatively bigger step up. In fact, Figure 3.7 shows that a significantly small portion of the reference species can be calculated using RC4 class reactions or higher. It appears at first glance that the errors associated with RC4 reactions are smaller than either RC2 or RC3, but making this direct comparison is tenuous due to the different populations of species being represented. Species that can be calculated by RC4 reactions are likely well represented in the reference set, which likely includes a disproportionate number of easier to calculate species such as simple alkanes.

Either way, even if the smaller populations were affording an advantage, the errors shown so far by using isodesmic reactions appears significantly larger than that of either BAC approach. The distribution of errors appears to be wider, and there appear to be more outlier values for isodesmic reactions than there were for either BAC type. Given this and the fact that many species cannot be calculated using isodesmic reactions, it does not appear to be worthwhile to utilize isodesmic reactions using solely the approach described so far. However, there are a few remedies worth exploring.

We note that the optimization procedure in Equation (3.3), while capable of finding isodesmic reactions automatically, is setup to prioritize isodesmic reactions that are simple. Other than ensuring that the constraints of the reaction class are satisfied, the optimization itself does not attempt to differentiate between isodesmic reactions using some metric of quality, leaving this task entirely up to the reaction classes. This does not allow for any further optimization within a reaction class. For example, one aspect that is worth considering is the quality of the species included in the isodesmic reaction. For example, some species in the reference set might have a poor geometry at the given level of theory. Buerger et al. were aware of this and considered this problem in their work. Their solution to this problem was to perform a careful prunning of the reference set to remove species that seemed to be inconsistent with the rest of the reference set in yielding higher errors.

Instead of using the prunning approach of Buerger et al., we considered trying to include some metric for this into the optimization procedure itself. With this, we considered an optimization that would minimize the expected errors of the involved species in the isodesmic reaction. One way to estimate this is to use the difference in the known heat of formation of the reference species with the heat of formation calculated with just AECs at the given level of theory. Equation (3.4) shows this approach, which tries to include species that likely are well represented by the given level of theory while still enforcing the constraints of the reaction class. Note that because including more species in the isodesmic reaction would only increase the objective function, this objective function behaves similarly to that of Buerger et al. in Equation (3.3) in trying to yield simplistic reactions all other factors being equal.

$$
\begin{aligned}
\min_{\nu} \quad & \sum_{i}^{N_{\text{ref}}} \left[|\nu_i| \cdot |\Delta_f^{\text{ref}} H_{298}^{\circ}(i) - \Delta_f^{\text{calc}} H_{298}^{\circ}(i)| \right] \\
\text{s.t.} \quad & \sum_{i}^{N_{\text{ref}}} \nu_i c_{ij} = 0 \quad \forall j \in \{1, \ldots, N_{\text{constraints}}\}, \\
& \nu_i > 0 \quad \forall i \in \{1, \ldots, N_{\text{ref}}\}
\end{aligned}
\tag{3.4}
$$

Using Equation (3.4) as the optimization algorithm, Figure 3.8 shows the resulting errors for using isodesmic reactions of class RC2 and higher. Looking at the results though, it appears that the errors are not significantly different from those achieved in Figure 3.5 where Equation (3.3) was used. There is some evidence to suggest that this is simply because
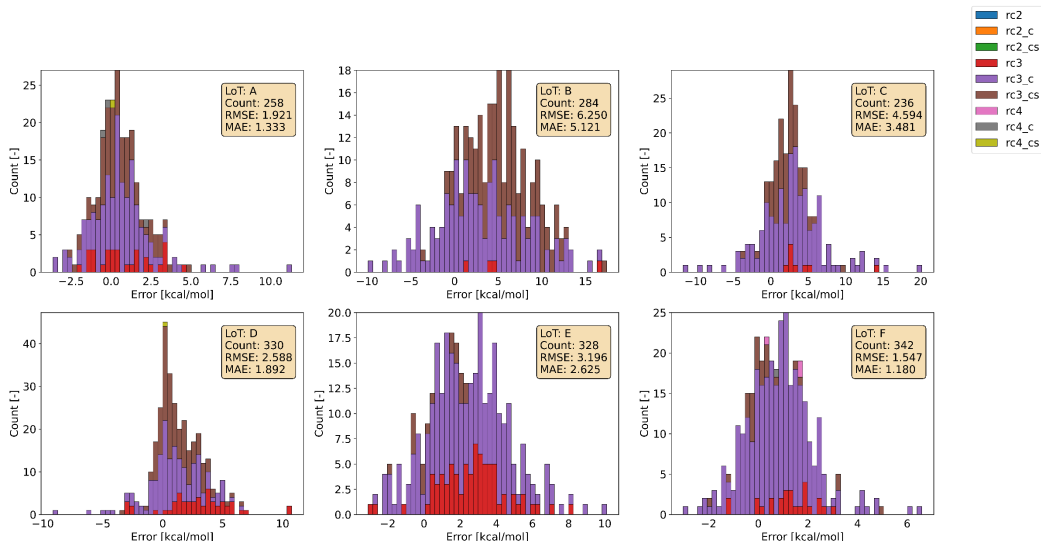
Figure 3.8. Errors between $\Delta_f^{\mathrm{iso}} H_{298}^\circ$ and known values of $\Delta_f^{\mathrm{ref}} H_{298}^\circ$ using the optimization algorithm from Equation (3.4). Isodesmic reactions of class RC2 or stricter were considered. The actual reaction class of found reaction are given different colors as seen in the legend

$|\Delta_f^{\mathrm{ref}} H_{298}^\circ(i) - \Delta_f^{\mathrm{calc}} H_{298}^\circ(i)|$ is a poor estimator for the quality of the included reference species (i.e. its expected error at this given level of theory). We note that $\Delta_f^{\mathrm{calc}} H_{298}^\circ$ are simply the heats of formation with AECs applied to them (see the methodology section where we explain this implementation detail), which as we see in Figures 3.3 and 3.4 have a wide distribution, usually not centered around zero. This means that this optimization function will actually favor species at the tail end of this distribution, so it is quite possible that these species are actually poor choices.

Given this, it is worth considering a different estimator for the quality of the reference species at the given level of theory. The next obvious choice here is to replace $\Delta_f^{\mathrm{calc}} H_{298}^\circ$ with $\Delta_f^{\mathrm{BAC}} H_{298}^\circ$, as the errors in heats of formation calculated with BACs are much narrower in distribution, and crucially centered around zero. Equation (3.5) updates Equation (3.4) with this approach in mind. Note that BAC data is only being used to help us algorithmically favor reference species that are likely well captured at this level of theory. Any BACs applied to the reference species and target would cancel out anyways in an isodesmic reaction approach. The BACs used for this implementation were Melius-type BACs.
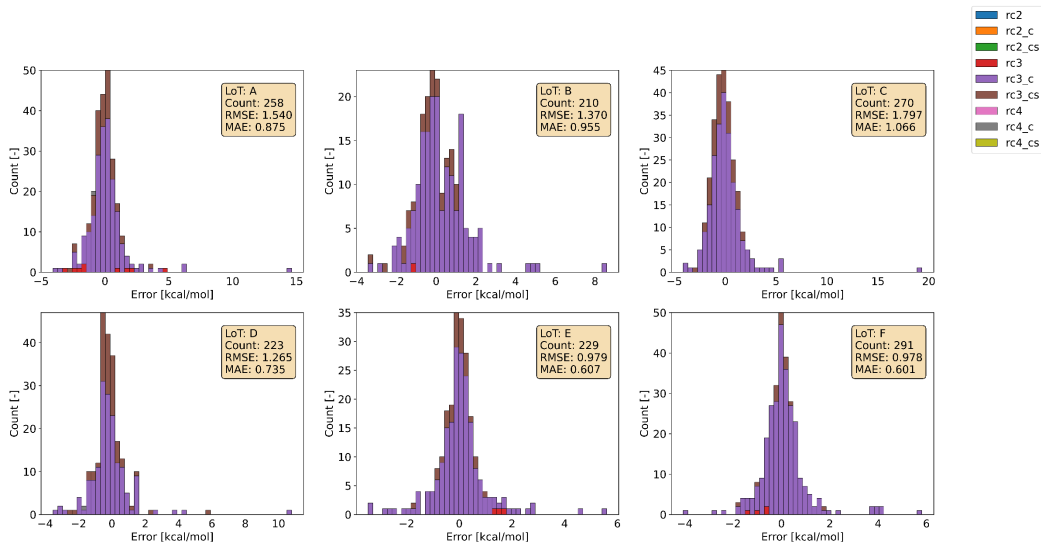
Figure 3.9. Errors between $\Delta_f^{\mathrm{iso}} H_{298}^\circ$ and known values of $\Delta_f^{\mathrm{ref}} H_{298}^\circ$ using the optimization algorithm from Equation (3.5). Isodesmic reactions of class RC2 or stricter were considered. The actual reaction class of found reaction are given different colors as seen in the legend

$$
\begin{aligned}
\min_{\nu} \quad & \sum_{i}^{N_{\mathrm{ref}}} \left[|\nu_i| \cdot |\Delta_f^{\mathrm{ref}} H_{298}^\circ(i) - \Delta_f^{\mathrm{BAC}} H_{298}^\circ(i)| \right] \\
\text{s.t.} \quad & \sum_{i}^{N_{\mathrm{ref}}} \nu_i c_{ij} = 0 \quad \forall j \in \{1, \ldots, N_{\mathrm{constraints}}\}, \\
& \nu_i > 0 \quad \forall i \in \{1, \ldots, N_{\mathrm{ref}}\}
\end{aligned}
\tag{3.5}
$$

Figure 3.9 shows the resulting errors from using the optimization scheme given by Equation (3.5) for reaction class RC2 or stricter. The distribution of errors given by this approach is much narrower. Furthermore, at a quick glance the errors using this approach appear to be on-par or even slightly better than those obtained through BAC approaches. However, once again we note that this direct comparison is tenuous, as the population of species is not the same, as once again this approach could not find isodesmic reactions for all of the reference species.

To better compare the performance of BAC approaches to this new isodesmic reaction algorithm, we constructed parity plots, where the x-axis is the error from isodesmic reactions while the y-axis is the error from applying Melius-type BACs. Each data point represents a single reference species that was able to be calculated both by BACs and isodesmic reactions. This gets around the issue of separate populations, and allows for a fair comparison between
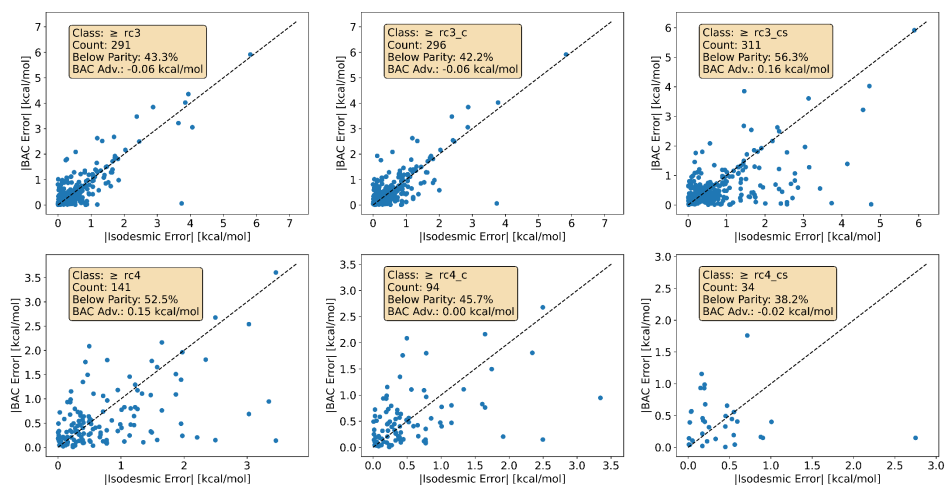
Figure 3.10. Parity plots showcasing the performance of isodesmic reactions found from Equation (3.5) and Melius-type BACs for a given species in the reference set at LoT F. Data points below the parity line indicate that using isodesmic reactions resulted in a lower error for that species than using Melius-type BACs. The BAC advantage (abbreviated "BAC Adv.") is the average difference between using Melius-type BACs versus isodesmic reactions, with a negative value indicating that isodesmic reactions performed better.

the method. Furthermore, an abundance of points above or below the parity line tells us which method is performing better most of the time.

The parity plot comparing cross-validation errors of Melius-type BACs and isodesmic reactions found using Equation (3.5) can be seen in Figure 3.10. All of the data correspond to LoT F, and are separated out by the various reactions classes (strictly held, so any RC3 reactions are not also RC4). Comparing these two methods, it appears that the species appear above the parity line (favoring isodesmic reactions) slightly more often than they appear below it (favoring BACs), though this is not consistent across all reaction classes. While this may be the case, averaging the errors for BACs and isodesmic reactions for these species yields an advantage of essentially zero, suggesting that both methods are roughly on-par performance wise. That being said, we are still left with the fact that Melius-type BACs work for all species, which is not true of the isodesmic reaction approach, given Melius-type BACs a significant advantage.

As a final observation, we note that in Figure 3.10 that the error from isodesmic reactions goes down as the reaction class is increased. The species that can be predicted using these higher order reactions classes are also well predicted by using Melius-type BACs though. It

is unclear if this is because being able to use higher order reaction classes means that the target species is well represented in the reference database, which likely also helps Melius-type BACs, or if these well represented species also happen to be species that are better captured by the given level of theory. Either way, it may be possible to predict which species have especially low errors at a given level of theory based on if they can be calculated using a higher order isodesmic reaction scheme.

3.4   Conclusions

In this work, we implemented several approached for correcting heats of formation obtained through quantum chemistry calculations. Part of this work included compiling a database of reference species and data from the literature that contained a diverse set of molecules so that these approaches could be applied to a broad range of molecules. This implementation is available to use with the open-source ARKANE software,[11] which allows other people to use these methods and even contribute new parameters for additional levels of theory using the developed API for fitting parameters and adding new information to the database. Because all of the information needed to fit parameters is stored in the reference species files in the database, it is also possible to replicate these fittings.

As for the performance of the various methods, Petersson-type and Melius-type BACs show very similar performance, though Melius-type BACs appear to generalize better. Both approaches are a substantial improvement over just using AECs, though, as previously noticed in the literature. By comparison isodesmic reactions do not appear to perform as well as using BACs, unless care is taken to optimize for including reference species that likely have low errors. In this work we built on the optimization method of Beurger et al. to achieve this end, and showed that the performance was on-par or slightly better than using BACs. Although we did not do so here, it would be interesting to compare this approach with the prunning approach used by Buerger et al. to achieve a similar end of prefering certain reference species. While the accuracy of these isodesmic reactions were comparable to that of BACs, isodesmic reaction approaches still have the downside that not all species can be calculated using them, as the reference set needs to include enough species with relevant features.

With this, in general we recommend using Melius-type BACs, as it has some slight

advantages. That being said, calculations using Petersson-type BACs should be viewed as equivalent, especially if the target species is well represented in the reference set. Similarly, calculations from isodesmic reactions could also be viewed as equivalent if care was taken to choose appropriate reference species, and higher reaction classes are used.

3.5   References

(1)   Reuther, A.; Kepner, J.; Byun, C.; Samsi, S.; Arcand, W.; Bestor, D.; Bergeron, B.; Gadepally, V.; Houle, M.; Hubbell, M.; Jones, M.; Klein, A.; Milechin, L.; Mullen, J.; Prout, A.; Rosa, A.; Yee, C.; Michaleas, P. In *2018 IEEE High Performance extreme Computing Conference (HPEC)*, IEEE: Waltham, MA, 2018, pp 1–6.

(2)   Petersson, G. A.; Malick, D. K.; Wilson, W. G.; Ochterski, J. W.; Montgomery, J. A.; Frisch, M. J. Calibration and comparison of the Gaussian-2, complete basis set, and density functional methods for computational thermochemistry. *The Journal of Chemical Physics* **1998**, *109*, 10570–10579.

(3)   Anantharaman, B.; Melius, C. F. Bond Additivity Corrections for G3B3 and G3MP2B3 Quantum Chemistry Methods. *J. Phys. Chem. A* **2005**, *109*, 1734–1747.

(4)   Buerger, P.; Akroyd, J.; Mosbach, S.; Kraft, M., *A systematic method to estimate and validate enthalpies of formation using error-cancelling balanced reactions*, Publication Title: Combustion and Flame Issue: 179 ISSN: 15562921, 2018; Vol. 187.

(5)   Ruscic, B.; Pinzon, R. E.; Morton, M. L.; von Laszevski, G.; Bittner, S. J.; Nijsure, S. G.; Amin, K. A.; Minkoff, M.; Wagner, A. F. Introduction to Active Thermochemical Tables: Several "Key" Enthalpies of Formation Revisited. *J. Phys. Chem. A* **2004**, *108*, 9979–9997.

(6)   Ruscic, B.; Pinzon, R. E.; Laszewski, G. v.; Kodeboyina, D.; Burcat, A.; Leahy, D.; Montoy, D.; Wagner, A. F. Active Thermochemical Tables: thermochemistry for the 21st century. *J. Phys.: Conf. Ser.* **2005**, *16*, 561–570.

(7)   Riplinger, C.; Neese, F. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *Journal of Chemical Physics* **2013**, *138*, ISBN: 1089-7690 (Electronic) 0021-9606 (Linking), DOI: 10.1063/1.4773581.

(8)   Pavošević, F.; Peng, C.; Pinski, P.; Riplinger, C.; Neese, F.; Valeev, E. F. SparseMaps—A systematic infrastructure for reduced scaling electronic structure methods. V. Linear scaling explicitly correlated coupled-cluster method with pair natural orbitals. *The Journal of Chemical Physics* **2017**, *146*, 174108.

(9)   Paulechka, E.; Kazakov, A. Efficient DLPNO-CCSD(T)-Based Estimation of Formation Enthalpies for C-, H-, O-, and N-Containing Closed-Shell Compounds Validated Against Critically Evaluated Experimental Data. *Journal of Physical Chemistry A* **2017**, *121*, 4379–4387.

(10) Dobek, F. J.; Ranasinghe, D. S.; Throssell, K.; Petersson, G. A. Evaluation of the heats of formation of corannulene and C60 by means of inexpensive theoretical procedures. *Journal of Physical Chemistry A* **2013**, *117*, ISBN: 1520-5215 (Electronic)\r1089-5639 (Linking), 4726–4730.

(11) Grinberg Dana, A.; Johnson, M.; Allen, J.; Sharma, S.; Raman, S.; Liu, M.; Gao, C.; Grambow, C.; Goldman, M.; Ranasinghe, D.; Gillis, R.; Payne, A. M.; Li, Y.-P.; Dames, E.; Buras, Z.; Vandewiele, N.; Yee, N.; Merchant, S.; Buesser, B.; Class, C.; Goldsmith, F.; West, R.; Green, W. *Automated Reaction Kinetics and Network Exploration (Arkane): A Statistical Mechanics, Thermodynamics, Transition State Theory, and Master Equation Software*; preprint; Chemistry, 2022.

(12) Ruscic, B.; Bross, D. H. Active Thermochemical Tables (ATcT) values based on ver. 1.122e of the Thermochemical Network, Argonne National Laboratory (2019); available at ATcT.anl.gov.

(13) Cioslowski, J.; Schimeczek, M.; Liu, G.; Stoyanov, V. A set of standard enthalpies of formation for benchmarking, calibration, and parametrization of electronic structure methods. *The Journal of Chemical Physics* **2000**, *113*, 9377–9389.

(14) Benson, S. W. Thermochemistry and kinetics of sulfur-containing molecules and radicals. *Chem. Rev.* **1978**, *78*, 23–35.

(15) NIST Chemistry WebBook, NIST Standard Reference Database Number 69, 10.18434/T4D303.

(16) Pedley, J. B.; Naylor, R. D.; Kirby, S. P., *Thermochemical Data of Organic Compounds*; Springer Netherlands: Dordrecht, 1986.

(17) Burcat, A.; Ruscic, B.; Chemistry; Tech, T.-I. I. o. Third millenium ideal gas and condensed phase thermochemical database for combustion (with update from active thermochemical tables). **2005**, DOI: 10.2172/925269.

(18) Johnson, R. D. J. I. NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101, Release 20, August 2019, 10.18434/T47C7Z.

(19) Mardirossian, N.; Head-Gordon, M. ω B97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *Journal of Chemical Physics* **2016**, *144*, ISBN: 0741110741, 0–23.

(20) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615.

(21) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **2010**, *132*, 154104.

(22) Shao, Y. et al. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Molecular Physics* **2015**, *113*, 184–215.

(23) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.;

Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian~16 Revision C.01, 2016.

(24) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.

(25) Neese, F. The ORCA program system. *WIREs Comput Mol Sci* **2012**, *2*, 73–78.

(26) Neese, F. Software update: the ORCA program system, version 4.0. *WIREs Comput Mol Sci* **2018**, *8*, DOI: 10.1002/wcms.1327.

(27) Yu, H. S.; Fiedler, L. J.; Alecu, I.; Truhlar, D. G. Computational thermochemistry: Automated generation of scale factors for vibrational frequencies calculated by electronic structure model chemistries. *Computer Physics Communications* **2017**, *210*, 132–138.

(28) Ochterski, J. W. Thermochemistry in Gaussian, 2000.

(29) Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **2020**, *17*, 261–272.

(30) Branch, M. A.; Coleman, T. F.; Li, Y. A Subspace, Interior, and Conjugate Gradient Method for Large-Scale Bound-Constrained Minimization Problems. *SIAM J. Sci. Comput.* **1999**, *21*, 1–23.

(31) Berkelaar, M.; Eikland, K.; Notebaert, P. lp_solve.

(32) Hart, W. E.; Watson, J.-P.; Woodruff, D. L. Pyomo: modeling and solving mathematical programs in Python. *Math. Prog. Comp.* **2011**, *3*, 219–260.

(33) Hart, W. E.; Laird, C. D.; Watson, J.-P.; Woodruff, D. L.; Hackebeil, G. A.; Nicholson, B. L.; Siirola, J. D., *Pyomo — Optimization Modeling in Python*; Springer Optimization and Its Applications, Vol. 67; Springer International Publishing: Cham, 2017.

Chapter 4

GCxGC-FID/MS and NMR analysis of Low-Temperature Closed-System Pyrolysis of Type I and II Kerogens for Validating Detailed Kinetic Models

Chapter Abstract

With the growth of automatic mechanism generation software and the availability of faster and faster computers and computational methods, larger and larger chemical systems can be studied using detailed kinetic models. This can include the use of surrogate models or fragment chemistry models, which use elementary reactions with *ab initio* rates to model larger system while abstracting away some of the computational complexity or the difficulty in precisely defining the starting material. One such large and complex chemical system that is within reach of a more detailed kinetic approach is the pyrolysis of kerogen, which is the non-soluble organic fraction of sedimentary rocks that can form oil and gas under geological conditions. While numerous kerogen pyrolysis studies exist in the literature, not all of these studies produce data that is ideal for validating these kinetic models. In this work, we add to the existing data for kerogen pyrolysis by performing closed-system, isothermal pyrolysis of two different kerogen shales at relatively low temperatures of 250°C and 270°C and 5-10 days. Under these conditions, we show that a significant portion of the kerogen has converted to soluble bitumen and oil, although the thermal maturity of these products is low. We are also able to identify key oil products after pyrolysis under these conditions, which could

give mechanistic insights when validating detailed kinetic models. Finally, we show that quantitative ssNMR results can be obtained even when using high-field instruments, and that newer methods like multi-CP/MAS ssNMR can substantially reduce the time needed for these experiments. This makes experiments on feedstocks such as kerogen likely more accessible for further study.

4.1 Introduction

Over at least the past 50 years, kerogen, the organic matter found in sedimentary rocks, has been the focus of countless studies both from a pyrolysis perspective and a kinetic modeling perspective. Given the oil and gas potential of kerogen found in oil shales, there is tremendous amount of interest in understanding the underlying kinetic mechanisms forming these products. Although great progress towards this end has been made, this is a challenging task given the complex nature of kerogen. First, it is important to note that the definition of kerogen is a practical one rather than a precise chemical one. By definition, kerogen is the fraction of organic matter that is not soluble in common organic solvents.[1] The soluble fraction is then usually split into a high molecular weight portion called bitumen, and a lower molecular weight oil portion. The precise chemical makeup of a given kerogen sample is dependent on the origin depositing matter, though kerogens are usually classified into type I, II, II-S, and III based on atom H/C and C/O ratio along with sulfur content. While we know of many biomarkers and their corresponding precursors that make up kerogen,[2] it is impossible to determine the exact molecular composition of a kerogen sample (though great work has been done to generate representative kerogen structures and study their properties[3–7]), and even obtaining bulk information about the initial kerogen structure and composition has always been challenging.

Given this, a significant advance towards studying kerogen was the use of solid-state NMR (ssNMR) with magic-angle spinning (MAS) and cross-polarization (CP) pulse sequences to measure different types of carbon in kerogen. CP/MAS ssNMR goes at least as far back as 1977, with work by Resing et al., who were able to measure the amount of aromatic carbon in a Green River kerogen sample.[8] An explosion in use of this technique can be seen in the late 1970s and 1980s, and included further advances to get more detailed information from ssNMR spectra. For example, Opella and Frey developed a tech-

nique for distinguishing between protonated and non-protonated carbon by inserting a delay (with no decoupling pulse on ^{1}H) between the cross polarization pulse and the acquisition (pulse sequences like this are commonly referred to as dipolar-dephasing experiments or non-quaternary suppression) .[9] Using pulse sequences like these, Solum et al. was able to build on this work by using data from both variable contact time experiments and dipolar de-phasing experiments to determine various structural parameters including average aromatic cluster size and fraction of bridgehead carbons.[10] Wilson[11] provides an excellent overview of the details and considerations behind these and other ssNMR experiments of interest to kerogen samples.

In addition to being used to determine the compositional and structural nature of kerogen feedstocks, ssNMR was used to gain valuable insights into the behavior of kerogen during pyrolysis. For example, important early works by Miknis et al. showed that the oil/gas generation potential of kerogen correlates well with the amount of aliphatic carbon present regardless of kerogen type,[12] and then later showed that the mass of aromatic carbon in pyrolyzed shales is roughly equal to the mass in the starting material.[13] Shortly thereafter, Hershkowtiz then performed pyrolysis experiments with care to closing the mass balance across products to note that the total amount of aromatic carbon between the solid residue and bitumen/oil increased over the course of the reaction.[14]

After this, it was common to try and perform pyrolysis experiments and fit a kinetic model to the data to describe the underlying mechanism. For example, Behar et al. per-formed isothermal, closed system pyrolysis on gold tubes of type II and III kerogen and fit a first-order kinetic model to the observed products.[15] To try and account for the difference in reactivity among similar types of molecules, another common modeling strategy used was to use a distribution of activation energies and fit to non-isothermal data. Burnham gives a good review of different ways to model kerogen pyrolysis using various distribution models, including various discrete, continuous, and isoconversional models.[16] These models are generally able to describe the bulk transformations of kerogen into various bitumen, oil, and gaseous components with a relatively small number of reactions. However, some caution should be used with these models. Landais et al. performed gold tube pyrolysis on type II and III kerogen across a broad range of temperature, and noted the difference in behavior in pyrolysis at lower temperatures for longer times and pyrolysis at higher temperatures for correspondingly shorter times.[17] Given that these activation energy distribution models are

usually fit to specific temperature ramps and ranges, usually above 300°C, it is therefore possible that these models do not accurately capture the behavior of kerogen pyrolysis at lower temperatures.

A different modeling approach that may not have this disadvantage to the same extent are detailed kinetic models using elementary reactions with *ab initio* rates. This modeling approach has been gaining in popularity with the use of automatic mechanism generation software.[18–22] With models built using software like the Reaction Mechanism Generator (RMG), the rate parameters are never fit to bulk experimental data like that obtained from kerogen pyrolysis studies but are instead based on data, calculations, or estimates for individual elementary reactions. Furthermore, the software aims to add only reactions that seem to be important at the user-specified conditions. Once the model has been generated, the model predictions and then usually compared to bulk experimental data to see how well the model is able to predict the observed experimental data. If the model and the data reasonably match, this gives confidence that the model is capturing the underlying chemistry properly. Because of this, the model likely is able to extrapolate to other conditions better than models that might have been overfit to bulk experimental data.

Of course, one downside of this modeling strategy is that generating fully-detailed models for large and complex chemical systems like kerogen are too difficult and expensive, and thus not currently possible. However, it is conceivable to generate fully-detailed models for surrogate compounds that might behave similarly to kerogen. Additionally, it might be possible to generate a detailed model that only keeps tract of the relevant functional groups or fragments of kerogen, abstracting away some of the complexity while still capturing the relevant underlying chemistry.

To see if these modeling approaches are feasible, though, high quality experimental data for kerogen pyrolysis is need for validating these models. Of course, such data is prevalent in the literature as mentioned above. A recent example that does an excellent and thorough job of characterizing both the initial starting material and the pyrolysis products is the work of Solum et al.[23] and Fletcher et al.,[24] who performed non-isothermal open-system pyrolysis of a Green river kerogen. For validating detailed models, though, isothermal closed-system pyrolysis data is usually preferred. This is because it is easier to simulate detailed models without having to worry about mass transport into and out of the system. Isothermal data across a broad range of temperatures is also preferred, as it is easier to evaluate the

model performance isothermally across the range, whereas ramped simulations might not be too sensitive to the exact reactivity at the lowest temperature in the ramp. High quality isothermal closed-system kerogen pyrolysis data exists in the literature, including work by Behar et al. who performed isothermal closed-system (as well as open-system) pyrolysis on type I, II, II-S, and III kerogen for temperatures as low as 300°C.[25] In this work, to obtain pyrolysis data for even lower temperatures of 250°C and 270°C, including obtaining GCxGC-FID/MS data that could identify key species that can give insights into the underlying mechanisms at play.

4.2 Methods

4.2.1 Preparation of Kerogen Samples

In this work we studied a type I Green River kerogen, along with a type II kerogen sample. The kerogen samples were demineralized using the procedure outlined by Solum et al.,[23] which closely follows the procedure of Vandegrift et al.,[26] though an additional wash with zinc chloride solution was used to further demineralize the type I kerogen sample. The zinc chloride solution was prepared using a ratio of 300 grams of zinc chloride per 100 mL of 3 M HCl.

A 20 gram portion of the type I Green River kerogen that had only completed the first step of the demineralization procedure (washing with methanol and DCM under reflux) was set aside for further use. We will refer to this material as "solvent-washed" for the remainder of this work. Since this material was abundant, we used this material to help us choose the final pyrolysis conditions. We also pyrolyzed solvent-washed samples alongside the fully demineralized samples in order to get more data, as well as determine what effect if any the additional inorganic minerals had on the pyrolysis. Because of the solvent washing, this sample should not contain any soluble bitumen or oil, but does retain the initial minerals in the shale.

A third fully demineralized sample of type III kerogen was also obtained, though there was only enough sample to characterize by ssNMR and FTIR, so this sample was not pyrolyzed.

4.2.2   Inorganic Analysis

The total inorganic content of the solvent-washed type I kerogen and the fully demineralized type I and type II samples were determined using thermogravimetric analysis (TGA) on a TA Instruments Q500-1579 instrument. This analysis was performed under air flux at a flow rate of 45 mL/min and with an initial heating rate of 10°C/min to 550°C. After holding at this temperature for 60 minutes, a heating rate of 10°C/min was once again used to reach a temperature of 750°C, which was held for 120 minutes. The weight of the remaining residue after 60 minutes at 550°C was taken as the total weight of the inorganic residue, while the weight lost after bringing the sample to 750°C was taken as the amount of carbonates in the sample.

4.2.3   ssNMR

To obtain quantitative spectra, a multiple cross-polarization (multi-CP) pulse sequence was used similar to that found in the literature.[27] We used a sequence of six 90-100% ramped CP pulses with a contact time of 1.1 ms and a delay of 0.2 seconds in between pulses. The recycle delay of 1 second was chosen to be well above 5 times the slowest T_{1H} as measured using the standard inversion recovery experiment. The number of CP pulses, delay time between pulses, and ramp were optimized against a known sample of poly(α-methylstyrene) (Millipore Sigma 81520-250MG, CAS No. 25014-31-7, Mn 106000 Da), chosen because it has a known amount of various protonated/non-protonated aliphatic and aromatic carbon atoms that are resolvable. A discussion on this method validation can be found in the results and discussion section.

4.2.4   Preparation of Samples in Gold Tubes

Closed-system pyrolysis was carried out inside of gold tubes, which allow for thermal and mechanical equilibrium with the rest of the tube reactor without exchanging material and minimizing possible wall reactions that might be catalyzed by stainless steel.[28] Gold tubing was obtained from XRF Scientific (99.99% Au, 5 mm OD, 0.2 mm thick) and initially cut to a length of 45-55 mm. One end of the gold tube was then arc welded shut before being annealed using a propane torch.

The gold tubes were then filled with approximately 300 mg of kerogen for solvent-washed

samples, while fully demineralized samples were filled with about 100-200 mg, as these samples are less dense. To provide an inert atmosphere, the gold tubes were then carried into a glove box with an argon atmosphere (max 5 ppm oxygen, usually less than 1 ppm), where they were then crimped shut (an effort was made during this crimping step to remove some of the "empty" space in the gold tube that would otherwise be occupied by argon, both to allow room for expansion and to minimize the amount of contaminating oxygen). The crimped gold tubes were then sealed inside of headspace vials with a 20 mm aluminum caps and a 3.2 mm PTFE/white silicone septa. This allowed for the gold tubes to be transported to the arc welder while maintaining their argon atmosphere. The gold tubes were then quickly removed from headspace vials and the remaining unsealed (but crimped) end of the gold tube was then arc welded shut. The final length of the gold tubes were usually around 35-50 mm.

4.2.5   Pyrolysis Reactor

Gold-tube pyrolysis experiments were performed inside of a stainless steel tube reactor (HiP part number TOC7-20G) sitting inside of a Lindberg Blue tube furnace. Figure 4.1 shows an schematic representation of this set up, with a cutaway in Figure 4.2 of the interior of the tube reactor. One end of the tube reactor was capped with a K-type thermocouple (HiP part number 209440L), while the other end of the reactor was connected to both a pressure relief valve and an ISCO 100DM syringe pump filled with deionized water. Custom stainless-steel metal inserts were created and placed inside of the reactor, both to reduce the volume of the reactor and to hold in place (up to) 4 gold tubes near the tip of the thermocouple. An experimental run consisted of loading the reactor with the metal inserts along with the prepared gold tubes containing the samples for pyrolysis. The reactor was then pressurized to 300 bar with deionized water at room temperature with the syringe pump operating in constant pressure mode to control the reactor pressure. After this pressure testing of the reactor, the tube furnace was then closed and turned on. A thermocouple built into the tube furnace was used to control the temperature of the tube furnace, while the internal K-type thermocouple near the samples was logged with a data recorder every second. In a typical run, the internal temperature as measured by the K-type thermocouple reached its equilibrium value within roughly two hours.

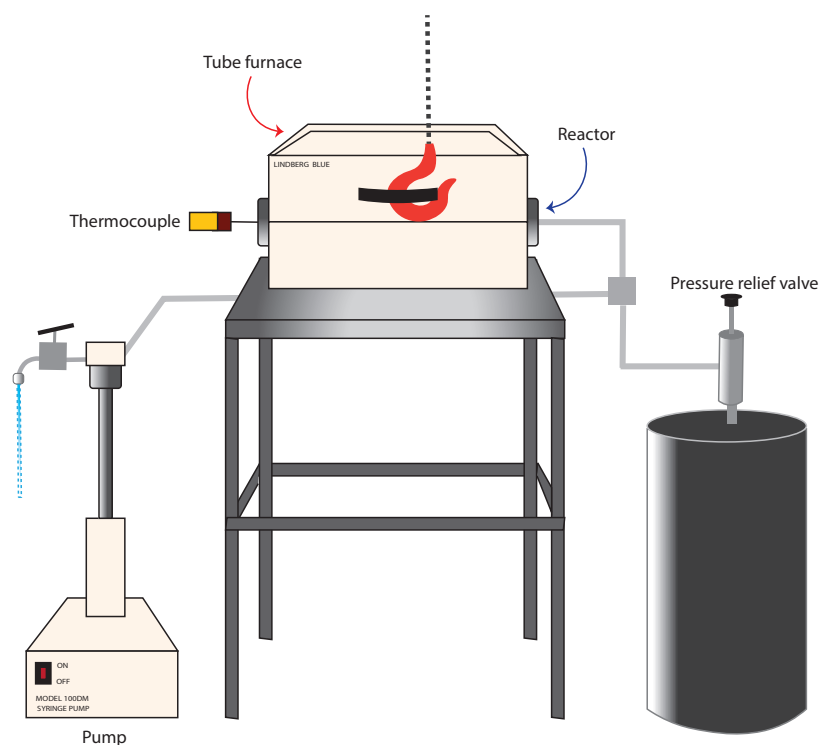At the completion of the run time, the tube furnace heater was turned off, and the lid

Figure 4.1. Experimental setup for gold-tube pyrolysis experiments, including a tube reactor enclosed by a tube furnace. A syringe pump (left) is used to control the pressure of the reactor to 300 bar, while a pressure relief valve on the right can empty the contents of reactor into a large barrel in the event of overpressurization. A gantry crane with hook is used to open the lid of the tube furnace for cooling down the reactor.
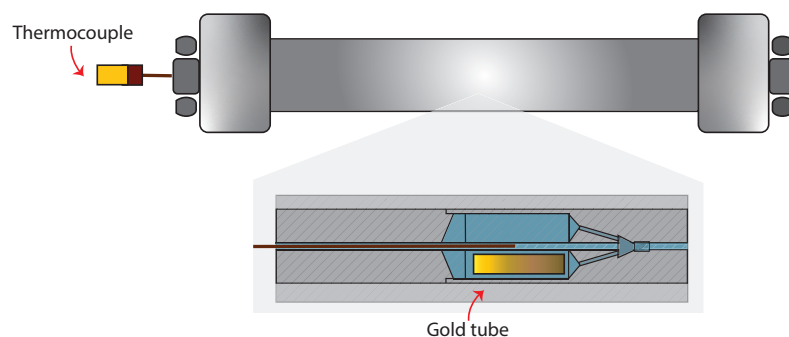

Figure 4.2. Cutaway of the inside of the tube reactor, showcasing the small chambers to hold the gold tubes in place right by the thermocouple down the center line. Steel inserts take up most of the volume of the reactor, but small channels allow for water from the external syringe pump to reach the gold tube to pressurize its contents to 300 bar.

to the tube furnace was lifted open using a gantry crane hook to allow the reactor to cool off faster. Typically the reactor would take about 5 hours to cool to below 50°C. The run time was taken to be the time between when the tube furnace heater was first switched on and when it was turned off (i.e. not including the cooldown time). Since the internal temperature was logged, though, it is possible to account for the precise nature of the heat up and cooldown periods, which are also much shorter than the run times in this study.

After the reactor had sufficiently cooled down, the pressure inside the reactor was reduced to atmospheric pressure, and then the tube reactor was opened up to retrieve the gold tubes. The tubes were then dried and re-weighed. Any gold tubes with a mass difference of more than 1 mg after pyrolysis were considered to be busted and discarded.

4.2.6 Reaction Conditions

In this study we performed pyrolysis experiments at two different temperatures, 270°C and 250°C, each with two different time points of 120 hours and 240 hours. The solvent-washed samples were pyrolyzed at all four conditions, but due to a lack of material the demineralized samples were only pyrolyzed at the 120 hour time point for both temperatures (two conditions in total).

These conditions were chosen because preliminary experiments using the solvent-washed kerogen showed noticeable changes in the solids as seen via ssNMR at conditions near these. For example, the composition of aliphatic carbon in the residual solids could change by at least ten percentage points even for the shorter time point and lower temperature. Although the changes to the residual solids are noticeable, these conditions are very much in the low conversion regime, so only small amounts of liquid and gaseous products are generated.

4.2.7 Analysis of Gaseous Products

Gaseous pyrolysis products were analyzed with GC-FID/TCD using an SRI MG #5 Model 8610C GC equipped with a thermal conductivity detector (TCD) and a flame ionization detector with an attached methanizer (FID). A HayeSep D packed column (6 ft × 1/8 in) was used with nitrogen as the carrier gas. The column was held isothermally at 110°C, and a runtime of 180 minutes was used to see molecules as big as butane. All samples were manually injected into the GC using a 1 mL syringe. In each case an air blank was run for 20 minutes to ensure there were no background hydrocarbon signals.

To manually extract the sample, the gold tube (after pyrolysis) was placed inside of 20 mL headspace vial. After the vial was sealed, a 21 gauge needle was used to pierce the gold tube on both sides through the septum of the headspace vial to release the gases from the gold tube into the headspace vial. The gases were allowed to diffuse throughout the headspace vial for at least 1 minute before 1 mL of gasses from the headspace vial were drawn up and directly injected into the GC. The gold tube was then placed in a new headspace vial with a new cap and septum to prevent further loss of liquid products for subsequent steps.

4.2.8 Chloroform Extraction

To separate the liquid products (more commonly referred to as tar, which we take to be any non-gaseous products that are soluble in chloroform) from the residual solids (char), we performed and extraction step using deuterated chloroform (to allow for further analysis of the liquids by NMR). In this extraction step, a known amount of internal standard was added to a 10 mL headspace vial so that the total amount of liquid products could be quantified. The gold tube (post gas analysis) was then cut into small sections and its contents were scraped into the headspace vial. The sections gold tube itself were also added to headspace vial, as it is difficult to scrape off all of the solids from the gold. The vial was then filled with deuterated chloroform, and capped. The sealed vial was then sonicated at room temperature for 1 hour to extract the liquid products into the chloroform before being chilled in a refrigerator for 30 minutes. To separate the extracted liquids from the residual solids the contents of the vial (except for the gold tube pieces) were poured into a centrifuge tube and centrifuged at 6000 rpm for 5 minutes. Since chloroform is more dense than the residual solid kerogen, the solids float to the surface, so a small needle and syringe we used to extract the liquids from beneath the solids. The extracted liquid was placed in 2 mL GC vials for further analysis.

To remove the residual solids from the remaining liquid, additional chloroform was added to the centrifuge tube and the tube was sealed and shaken to mix the contents. The tube was then further centrifuged at 6000 rpm for 5 minutes and as much liquid was removed from the tube as possible via needle and syringe without removing any of the solids. This process was repeated several times until the liquid in the tube was clear. The solids were then allowed to air dry to remove the last of the residual chloroform.

We primarily used iodoform (Millipore Sigma 109452-5G, CAS No. 75-47-8, 99%) as an internal standard for this extraction step. Iodoform has a few interesting properties that make it ideal for use as an internal standard. It is a non-volatile solid, and has exactly 1 hydrogen and 1 carbon atom that each have unique NMR chemical shifts (4.92 ppm in ^{1}H and -161.5 ppm in ^{13}C) that do not overlap with any aliphatic or aromatic carbon peaks. Furthermore, because the mass of carbon and hydrogen per molecule is a small percentage, iodoform can be added neat and yield ^{13}C and ^{1}H signals of similar size to the pyrolysis products in NMR. In this work, 100-200 mg of iodoform was usually used.

While iodoform was primarily used, some of the earlier runs (type I and type II demineralized at 270°C and 120 hours) in this work used hexamethyldisilane (Millipore Sigma, CAS No. 217069-5G, 98%). This internal standard was abandoned because it is slightly volatile and it is difficult to distinguish with tetramethylsilane in NMR.

4.2.9 Solution-Phase NMR

Solution-phase NMR was performed on a Bruker Avance Neo 600MHz spectrometer, with the exception of two samples (type I and type II demineralized at 270°C and 120 hours) which were run on a Bruker Avance Neo 500MHz spectrometer. The 600MHz was equipped with a 5mm helium-cooled QCI-F cryoprobe, while the 500MHz was equipped with 5mm liquid-nitrogen cooled Prodigy broad band observe cryoprobe.

To obtain quantitative spectra, we used chromium(III) acetylacetonate (Cr(acac)$_3$, Millipore Sigma, 574082-25G, 99.99% trace metals basis) at a concentration of 5 mg/mL, which has been shown to significantly reduce the relaxation delay needed without sacrificing quantitative accuracy for pyrolysis oils.[29] At this concentration of Cr(acac)$_3$, we measured (using the standard population inversion experiment) the longest T1 relaxation time among all of the ^{13}C peaks and determined that a relaxation delay of 15 seconds was appropriate, as this was more than 5 times the longest T1 relaxation time. For proton spectra a similar procedure was used and a relaxation delay of 4 seconds was deemed appropriate.

^{13}C spectra were obtained using the Bruker "zgdepg_noNOE" pulse sequence, which uses a double-echo sequence to suppress the probe background, along with power-gated decoupling of ^{1}H with no NOE enhancement. A 0.5 second acquisition time was used along with a 14.5 second recycle delay (for a total time of roughly 15 seconds for relaxation) for 2048 scans. For ^{1}H spectra a relaxation delay of 4 seconds was used for 256 scans.

4.2.10 GC×GC-FID/MS

Liquid pyrolysis products were also analyzed via two-dimensional gas chromatography with quadrupole mass spectrometry (MS) and a flame ionization detector (FID). A modified Agilent 7890 instrument (with FID) equipped with a Zoex Corporation ZX2 thermal modulator was used, with an attached Agilent 5975C quadrupole mass spectrometer. RXi-5HT (30 m × 0.25 mm ID × 0.25 μm film thickness) was used as the primary column, which to a good approximation separates components based on their boiling point. The end of the primary column was attached to a deactivated fused silica column (1.5 m × 0.25 mm ID) to yield a 1 m long modulation loop, which was exposed at a single point to separate hot and cold jets for thermal modulation. The end of the modulation loop was attached to the secondary column of BPX-50 (2 m × 0.15 mm ID × 0.25 μm film thickness), which to a good approximation separates components based on polarity, with oxygenated and multi-ring species being retained in the column for longer. The secondary column was also encased in a secondary oven, which was held 25°C hotter than the primary oven. The end of the secondary column was attached to more deactivated fused silica, which connected to a splitter to split the flow to both the FID and the MS.

For sample runs, a primary oven ramp of 3°C from 45°C to 300°C was used, with the secondary oven also being ramped to maintain a temperature of 25°C above the primary oven. A modulation time of 16 seconds was used with a hot jet duration of 0.75 seconds. A split inlet held at 350°C was used, with a split ratio of 1:1, a column flow of 2 mL/min, and a septum purge flow of 3 mL/min. An auto-injector was used to deliver an injection volume of 5 μL (though the internal standard method used accounts for variation in injection volumes).

Before and after each run, an acetone blank was injected through the instrument to purge the columns of any material from previous runs. In these purge runs, all columns were kept at 250°C for 90 minutes, and the cold jet of the modulation loop was turned off (while the hot jet was left on to pulse every 16 seconds).

4.3 Results and Discussion

4.3.1 Solid-State NMR Method Validation

While solid-state NMR (ssNMR) of kerogen shale samples is pretty common, in general great care must be taken when trying to achieve quantitative results, as cross-polarization (CP) is inherently not a quantitative technique, and direct polarization (DP) experiments are usually too time intensive to perform due to the long relaxation delays needed. Furthermore, most ssNMR studies of kerogen shale samples utilize low-field instruments (e.g. a proton frequency of 100 MHz and a carbon frequency of 25 MHz), as opposed to the high-field instruments available for this work (note that low-field instruments benefit from needing lower spinning speed to eliminate spinning sidebands, whereas the higher resolution of high-field instruments is an advantage that goes unused given the broad peaks in kerogen spectra). Therefore, we decided it would be prudent to validate our ssNMR methodology to make sure our results are quantitative. While we ultimately settled on using a multi-CP pulse sequence and believe it to be reasonably quantitative, we tried a few other techniques as part of this method validation that are worth mentioning.

One common solution to obtaining quantitative spectra while still using CP is to perform variable contact time experiments to account for the different spin dynamics between various types of carbon in the sample. Solum et al.[10] outline the details for this experiment, where a series of spectra are taken at various contact times and the peak area as a function of contact time is used to fit a spin kinetic model to determine what the fully magnetized peak area would be for all peaks in the spectra. The exact model is given in Equation 4.1, where τ_{cp} is the contact time and M_0 is the peak area to use for quantitation. $T_{1\rho}^H$ is the time constant for T1 relaxation of protons in the rotating frame, and T_{CH} is the time constant for the transfer of magnetization from proton to carbon during the cross polarization.

$$M(\tau_{cp}) = M_0(exp(-\tau_{cp}/T_{1\rho}^H) - exp(-\tau_{cp}/T_{CH})) \tag{4.1}$$

Finding values of M_0 for each peak in the spectra then is a three parameter non-linear fitting to the variable contact time data obtained. It is also possible to measure $T_{1\rho}^H$ directly as the exponential decay time constant for variable spinlock time experiments (Bruker "cpht1rho" pulse sequence), reducing this to a two parameter non-linear fitting problem.
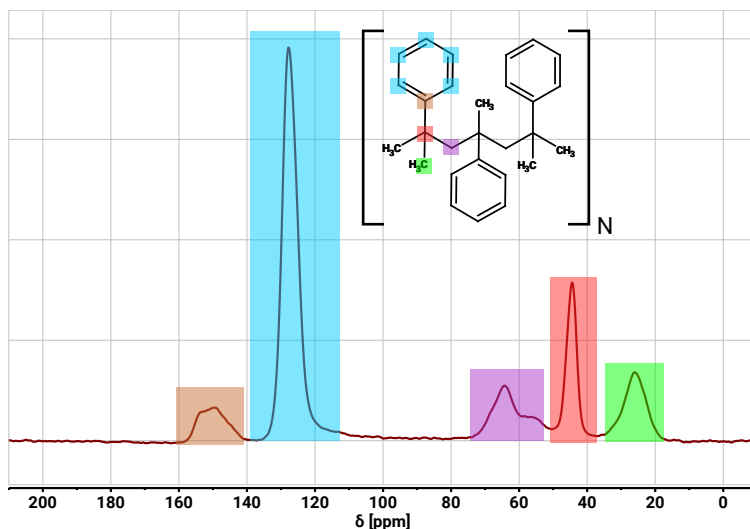
Figure 4.3. CP spectra of PAMS using a contact time of 2 ms for 2500 scans. A 2D representation of three monomer units of PAMS are shown, and are color matched to their corresponding peaks in the NMR spectrum.

We first tried these variable contact time experiments to obtain quantitative kerogen spectra. To validate this methodology, though, we decided to test a sample of known composition to verify that we could obtain quantitative results. For a known standard sample we used solid poly(α-methylstyrene) or PAMS with a number average molecular weight of roughly 106,000. Figure 4.3 shows the ssNMR spectra obtained for PAMS, along with its 2D structure to showcase the various peaks in the sample.

PAMS was chosen as our known standard for validation because it contains both protonated and non-protonated varieties of both aliphatic and aromatic carbons. This helps test that the ssNMR method can properly quantify each type of carbon, given that non-protonated carbons usually do not show up as strongly as protonated carbons in CP experiments. This is also useful for validating methods for distinguishing between protonation states such as dipolar dephasing experiments. While these are ideal properties for a ssNMR standard in this use case, it should be noted that morphology and the mobility of atoms in the solid can have a significant effect on the spin dynamics of the material. We have not shown that PAMS has similar mobility to that of kerogen, and in fact there are good reasons to believe that these two materials are quite different. Therefore, it is possible that the spin dynamics of PAMS are such that it is harder to obtain quantitative spectra for PAMS than it is for kerogen.

	150 ppm	128 ppm	64 ppm	44 ppm	26 ppm
Expected M_0	1.0	5.0	1.0	1.0	1
Fitted M_0	1.14	3.60	0.94	1.3	1

Table 4.1. Variable contact time method validation for PAMS, showing the expected values for M_0 in comparison to the fitted values obtained from the model in Equation 4.1. All value of M_0 were normalized such that M_0 for the methyl peak at 26 ppm is exactly 1.

After measuring T_1^H to choose a suitable recycle delay of 3 seconds, $T_{1_\rho}^H$ was measured (for all 5 peaks in the spectra) using variable spinlock experiments, which was used in the fits of subsequent variable contact time experiments. For the variable contact time step, 16 experiments in total were run with contact times ranging from 0.5 ms to 10 ms. Values for M_0 were then obtained for each of the 5 peaks in the spectra by fitting the model given in Equation 4.1. The results of these fits are shown in Table 4.1

As shown by the comparison, some of the fits are reasonable close (methylene carbons at 64 ppm and surprisingly the non-protonated carbon at 150 ppm). However, larger errors of 39% and 30% are noticeable for the protonated aromatic carbon and quaternary aliphatic carbons, respectively, relative to the methyl peak. Because the quaternary carbons are overestimated while the protonated aromatic carbons are underestimated, though, the worst case difference between these peaks is a multiplicative factor of $1.3/(3.60/5) = 1.80$.

One possible source for these errors is due to the simplifications made to arrive at Equation 4.1. Kolodziejski and Klinowski[30] derived a set of differential equations describing the kinetics of cross polarization, and showed the simplifications that can be made to arrive at Equation 4.1. In particular, there are three assumptions that are made to arrive at Equation 4.1: that ^{13}C spins are dilute compared to ^{1}H, that $T_{CH}/T_{1_\rho}^C \approx 0$, and $T_{CH}/T_{1_\rho}^H \approx 0$. The first assumption that ^{13}C is dilute compared to ^{1}H is a good assumption, as the natural abundance of ^{13}C is only 1%, and there are more protons in the sample than carbon. Since we directly measure $T_{1_\rho}^H$ and fit values for T_{CH}, we can test the validity of the third assumption. $T_{1_\rho}^H$ was usually around 10 ms, while T_{CH} was usually a few tenths of a millisecond, with the exception of the non-protonated aromatic and aliphatic carbon peaks, which had a T_{CH} of 3.33 ms and 1.71 ms, so it appears that the assumption that $T_{CH}/T_{1_\rho}^H \approx 0$ does not always hold. If we still assume that $T_{CH}/T_{1_\rho}^C \approx 0$ still holds, though, Kolodziejski and Klinowski also derived and expression given by Equation 4.2 that we can use instead.

	150 ppm	128 ppm	64 ppm	44 ppm	26 ppm
Expected M_0	1.0	5.0	1.0	1.0	1
Fitted M_0	0.81	3.76	0.94	1.18	1

Table 4.2. Corrected values of M_0 for PAMS using Equation 4.2 assuming that prior fit values for T_{CH} were valid.

$$M(\tau_{cp}) = \frac{M_0}{1 - (T_{CH}/T_{1\rho}^H)}(exp(-\tau_{cp}/T_{1\rho}^H) - exp(-\tau_{cp}/T_{CH})) \qquad (4.2)$$

If we assume that the fitted values for T_{CH} are correct and use Equation 4.2, we can correct-in-post the values for M_0. These corrected values for M_0 are given in Table 4.2.

Applying this correction yield slightly better results, with the only error significantly above 20% being the protonated aromatic carbon peak at 33%. It is unclear why the error for this peak is still relatively high. We have not considered the assumption that $T_{CH}/T_{1\rho}^C \approx 0$, though the fitted T_{CH} for this peak was 0.48 ms, which is small, and relaxation times for carbon are typically longer than that for hydrogen. Still, if it turned out that this assumption did not hold for this particular peak (but did for others), the models of Kolodziejski and Klinowski would predict that this peak is being under predicted.

While variable contact time experiments are the most common method for getting quantitative results, one particularly promising method that can also be used is to perform a multi-CP experiment as posed by Johnson and Schmidt-Rohr,[27] where multiple cross-polarization pulses are used prior to acquisition in order to overcome the loss of magnetization from $T_{1\rho}^H$ relaxation. This method has the advantage of being able to obtain quantitative data from a single experiment, as opposed to on the order of 10 experiments needed for variable contact time. To use the method properly, though, there are several parameters that need to be optimized, including the number of cross-polarization pulses, the contact time, and the delay time in between pulses. Using PAMS again as a standard, we ran a series of experiments and determined the optimal parameters to be 6 pulses with a contact time of 1.1 ms and a delay of 0.2 s in between pulses. The resulting peak integrals from this optimal multi-CP experiment are shown in Table 4.3.

As shown in Table 4.3, the multi-CP experiment yielded reasonably quantitative results in just a single run, with no peak having an error greater than 20% relative to the methyl carbon, though the worst case difference (in this case between the methylene carbon and the non-protonated carbon) is still a factor of $1.2/0.84 = 1.43$.

	150 ppm	128 ppm	64 ppm	44 ppm	26 ppm
Expected Peak Area	1.0	5.0	1.0	1.0	1
Measured Peak Area	0.84	5.94	1.19	1.20	1

Table 4.3. Multi-CP method validation for PAMS, showing the expected peak areas in comparison to the measured peak areas directly from the multi-CP spectrum. All peak areas were normalized such that the peak area for the methyl peak at 26 ppm is exactly 1.

Given that the multi-CP method seems to perform slightly better than the variable contact time experiments, and the advantage of using just a single experiment to get quantitative data directly, we decided to use the multi-CP method for this work. It is important to note that the errors in quantitation are still significant for PAMS, though again it is possible that it is much harder to achieve quantitative results for PAMS than it is for kerogen. Indeed, later on we discuss the measured composition of a Green River type I kerogen using this method, which agrees almost perfectly with values from the literature for a Green River kerogen[23] where they validated their variable contact time experiments with direct polarization experiments, which is an inherently quantitative technique. Therefore, this multi-CP methods likely yields quantitative results that are even more accurate than our validation with PAMS would suggest, though it is worth noting then that this may not be true for other types of samples, or maybe even other types of kerogen.

4.3.2 Dipolar Dephasing Experiments for High-Field Instruments

In addition to obtaining quantitative spectra, it is desirable to obtain further structural information about kerogen from ssNMR. For example, the broad peaks in the aromatic and aliphatic portion of the spectra can be broken up further into further features based on chemical shift. Another strategy, though, is to elucidate additional structural features by differentiating between protonated and non-protonated carbons. For example, Solum et al.[23] used dipolar dephasing experiments to separately quantify protonated and non-protonated carbons in a type I Green River kerogen. With this information, they were able to determine additional structural information of the kerogen, including average aromatic cluster size, using a procedure developed in prior work.[10] To do this, a series of experiments were performed using progressively longer dipolar dephasing delays, and fitting the peak areas to a model to figure out how much of the decay follows a fast Gaussian decay (protonated carbons) versus a slower Lorentzian decay (non-protonated carbons).
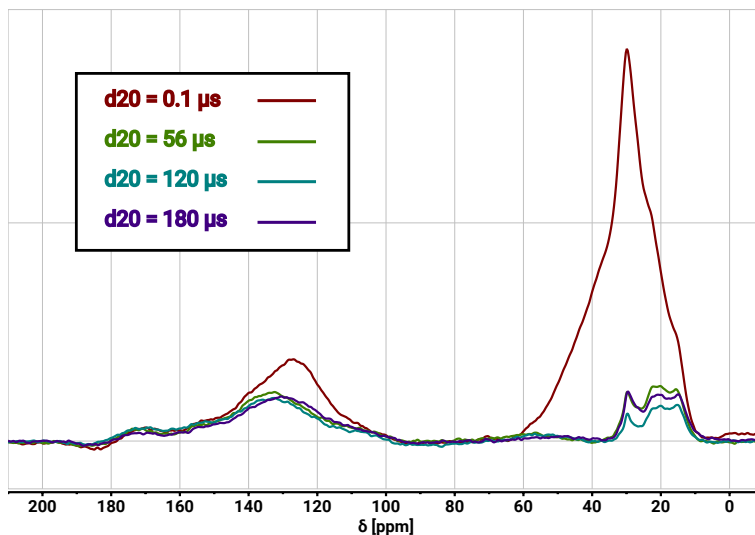
Figure 4.4. TOSS dipolar dephasing experiments on demineralized type II kerogen at various dipolar dephasing delays (d20). Due to oscillation caused by the rotor spinning rate, the longest delay (purple) does not result in the smallest peak areas

There are some challenges to performing these experiments using a high-field instrument, though. High-field instruments require much faster spinning speeds to remove spinning sidebands from the spectra; however, high spinning speeds in dipolar dephasing experiments can cause problems. Newman created a model for how the spinning rate effects the signal in dipolar dephasing experiments and validated the results experimentally.[31] Large spinning speeds results in oscillations in signal strength over the dephasing time, making it very difficult to fit a model for the Lorentzian decay of non-protonated carbons. In their work, large oscillations were observed for ammonium tartrate spun at just 8.5 kHz on an instrument with a carbon frequency of 50 MHz. We were also able to observe large oscillations when spinning our samples at 20 kHz using the standard Bruker "cpnqs" pulse sequence, rendering the data useless.

One possible solution to this problem is to spin the sample at lower spinning speeds and use other techniques to eliminate spinning sidebands. To perform dipolar dephasing experiments on a high-field instrument, we used the Bruker "cptoss_nqs" pulse sequence based on the work of Raleigh et al.,[32] which uses the total suppression of sidebands (TOSS)[33] technique to eliminate sidebands. Figure 4.4 shows as series of spectra taken using this technique at various dephasing times for the demineralized type II kerogen sample studied in this work.

While it appears that the protonated carbon signal has been mostly eliminated, and large oscillation are not observed, there are still some difficulties apparent from Figure 4.4. Notably, the signal strength of the non-protonated aliphatic carbons at the longest dephasing delay is actually a decent bit larger than than of the second longest dephasing delay. This indicates that minor oscillations are still present. In theory, it should be possible to determine the proper Lorentzian decay despite these oscillations by using longer dephasing delays, though there is an upper limit on the allowed dephasing delay in order to prevent TOSS pulses from overlapping (in our case this was just under 200 μs). Rayleigh et al. mention a few solutions to this problem though, such as inserting an integer multiple rotor periods in between the 90° pulses, or switching to the four pulse TOSS variant. A large number of scans would likely be needed to overcome signal-to-noise issues. This technique was not pursued further in this work.

4.3.3 Kerogen Initial Composition

The initial composition for the starting materials used in this work, along with that of a demineralized type III kerogen (which was not available in sufficient quantities for pyrolysis), were primarily characterized by multi-CP ssNMR. Figure 4.5 shows the multi-CP spectra acquired for these materials.

In this spectra, two broad peaks are visible, one for the aromatic carbon centered around 125 ppm, and one for the aliphatic carbon centered around 30 ppm. As expected, the Green River type I samples are rich in aliphatic carbon, while the type II and type III kerogens are progressively richer in aromatic carbon, with the type III kerogen sample being majority aromatic carbon. We can break up the spectra into further regions by considering carbonyl/oxyl carbons that reside downfield of the aromatic carbons in the 220-165 ppm range, and aliphatic carbons that are attached to oxygen (C-O) in the 90-50 ppm range. Small shoulders for both carbonyl/oxyl carbon and aliphatic C-O are visible on the type I and type II kerogens, whereas a clear C-O peak centered at 75 ppm is visible in the type III kerogen, and its carbonyl/oxyl shoulder is much more pronounced. Table 4.4 shows the quantification of these peak areas for each of these samples as a percentage of the total amount of carbon in the sample.

Since our type I kerogen is a Green River kerogen sample, which is well studied in the literature, we can compare this quantification with those in the literature. Depending on
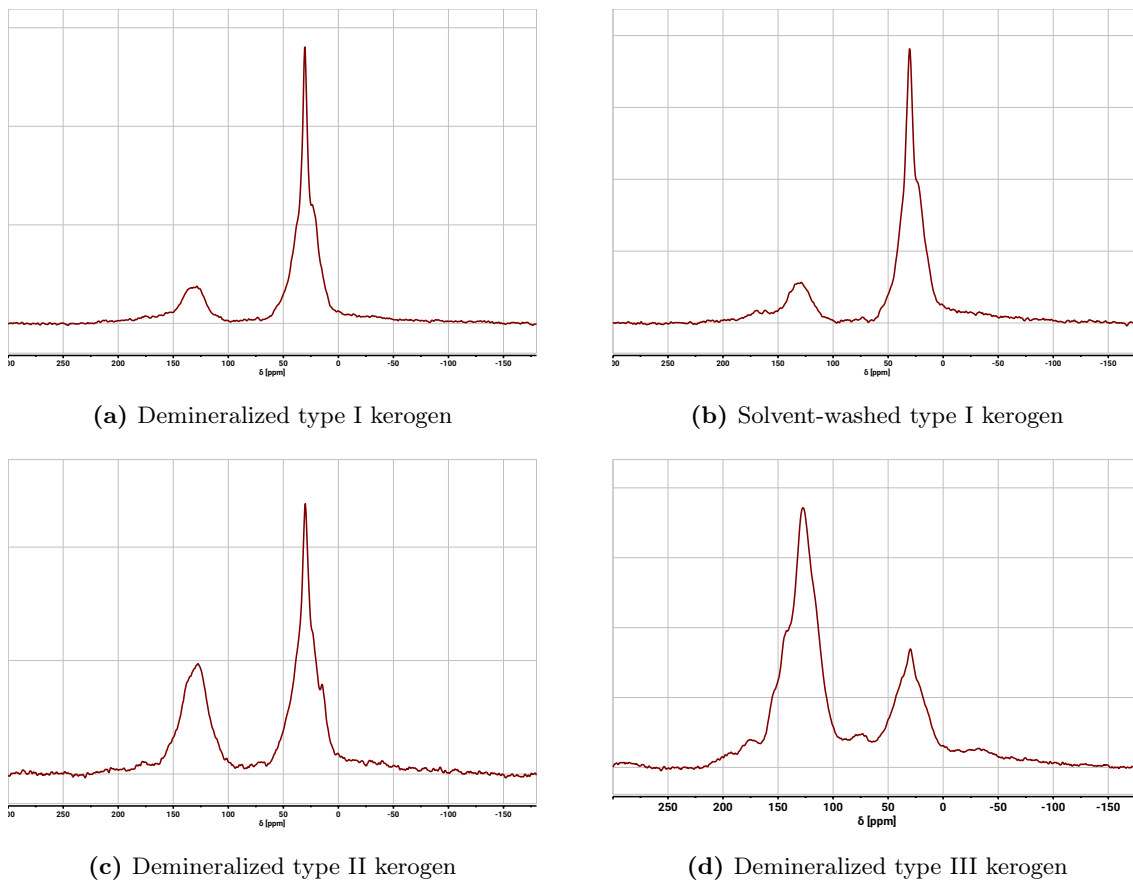
(a) Demineralized type I kerogen

(b) Solvent-washed type I kerogen

(c) Demineralized type II kerogen

(d) Demineralized type III kerogen

Figure 4.5. Multi-CP spectra of kerogen samples used as starting materials in this work. Each spectra has been individually scaled up to show detail.

Sample	Carbonyl/oxyl 220-165 ppm	Aromatic 165-90 ppm	Aliphatic 90-0 ppm	Aliphatic C-O 90-50 ppm
Demineralized type I	3.6	19.9	76.5	5.4
Solvent-washed type I	3.8	19.0	77.2	4.0
Demineralized type II	3.4	37.5	59.1	5.9
Demineralized type III	5.5	61.5	33.0	8.7

Table 4.4. Composition of kerogen starting materials as quantified from multi-CP ssNMR. All values are percentages of the total amount of carbon. The first three columns sum to 100%, while the third and last column are not mutually exclusive.

the depth of shale sample taken, Solum et al.[23] measured a total aliphatic carbon content of 75%, and an aromatic carbon content (excluding carbonyl/oxyl carbons) of 21%, both within 1% of the values obtained in this work. As mentioned previously, this gives us confidence in our multi-CP methodology, as these number were verified in the work of Solum et al. using the inherently quantitative direct polarization technique.

It is also worth comparing the results for the demineralized and solvent-washed varieties of the type I kerogen. Because the demineralization procedure involves the use of nitric and hydrofluoric acids, it is possible that the organic material in the kerogen has been altered. Additionally, the excess of minerals present in the solvent-washed sample could distort the ssNMR spectra, as ferromagnetic components can exacerbate the issue of spinning sidebands. Looking at Table 4.4, though, the composition of solvent-washed and demineralized type I kerogen differ by no more than 1%. Figure 4.6 shows these small differences more clearly by superimposing these spectra and normalizing their areas. The spectra appear to overlap almost perfectly, with the exception of a few slight deviations in the carbonyl/oxyl region and the methylene region around 30 ppm.

It is also important to quantify the amount of inorganic residue present in the sample. Table 4.5 shows the quantification of the inorganic matter as determined by TGA. Solvent-washed type I kerogen, which only went through the first step of the demineralization procedure, is mostly inorganic residue at 65% as expected. On the other hand, the demineralized type I kerogen sample only has 7.1% inorganic residue, a small amount of which is carbonates. These results are consistent with the results obtained from ssNMR. In fact, if we assume that the demineralized type I kerogen sample is 92.9% organic matter visible on ssNMR and the solvent-washed type I sample is 35% organic matter visible on ssNMR and account for the mass of sample packed into the rotor, we almost replicate the area normalized plot of Figure 4.6 within 5% (the solvent-washed spectra has an area of 1.05 times that of the demineralized sample). The demineralization procedure for the type II sample appears to not have been as successful, with an inorganic reside of 22.9%.

Also, note that the mineral carbonate like the 14.1% fraction for solvent-washed type I kerogen shown in Table 4.5 are not visible in CP (or multi-CP) ssNMR due to the lack of protons near these minerals. This is consistent with findings of Solum et al.[23] who compared spectra of shale samples using both CP and single-pulse (also known as direct polarization) methods, and showed that the mineral carbonates were only visible in the
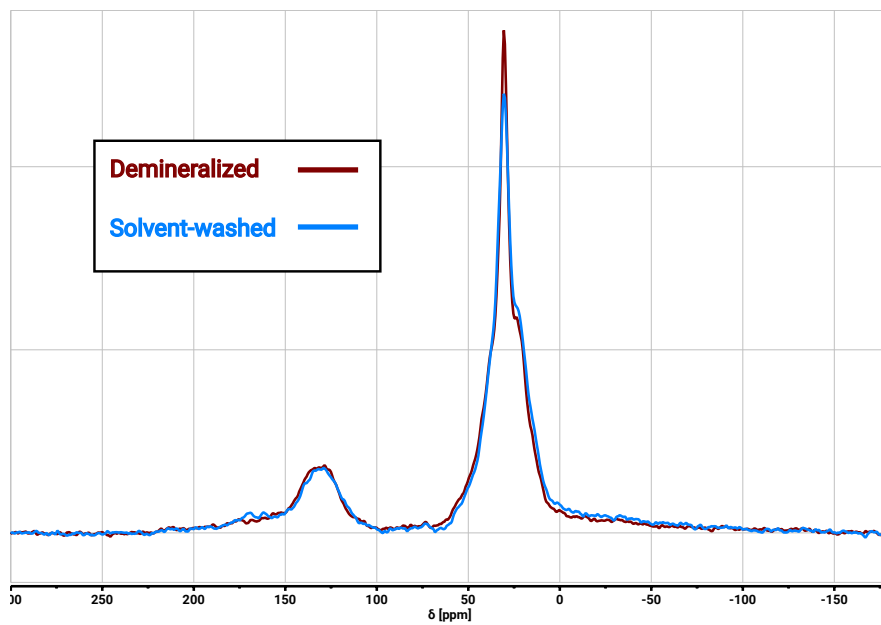
(a) Area normalized
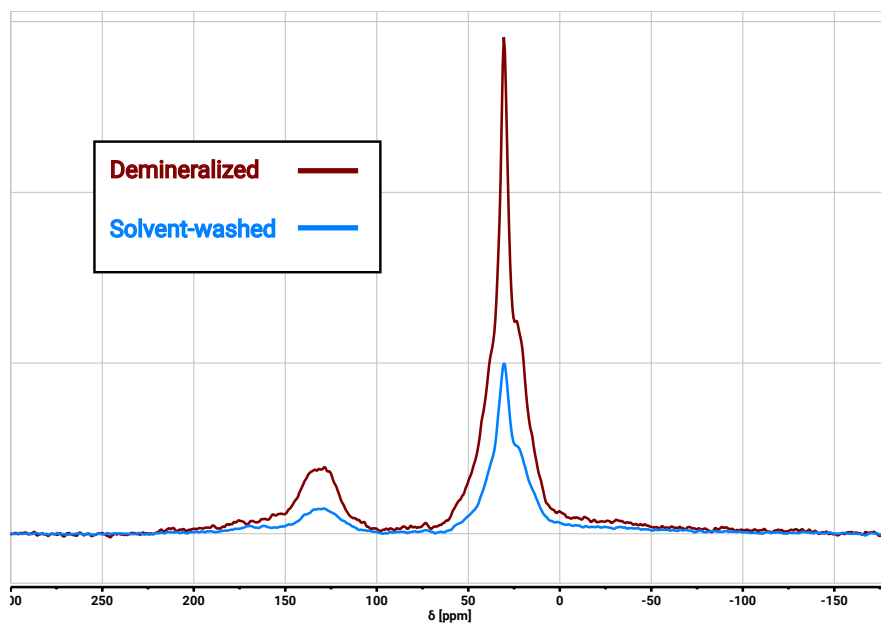
(b) Mass normalized

Figure 4.6. Comparison of multi-CP spectra for type I demineralized kerogen (red) versus type I kerogen that was only solvent-washed. Normalizing by area (top) makes it easy to see small compositional differences while normalizing by the mass of sample packed into the rotor (bottom) makes it clear that there is less organic matter in the solvent-washed sample

Sample	Inorganic residue	Carbonates
Demineralized type I	7.1	2.5
Solvent-washed type I	65.0	14.1
Demineralized type II	22.9	1.2

Table 4.5. Quantification of the amount of inorganic residue present in the starting feedstock kerogens as a percentage. The inorganic residue is taken as the amount of mass remaining after being held at 550°C for 60 minutes, and includes the mass of the carbonates. The amount of carbonates was taken as the further mass lost when heated to and held at 750∘C for 120 minutes.

Sample	Carbonyl/oxyl 220-165 ppm	Aromatic 165-90 ppm	Aliphatic 90-0 ppm	Aliphatic C-O 90-50 ppm
270°C for 120 hours				
Solvent-washed type I (1)	4.3	27.3	68.3	3.3
Solvent-washed type I (2)	3.6	25.7	70.7	2.5
Demineralized type I	0.6	34.6	64.8	1.7
Demineralized type II	1.6	50.0	48.4	3.8
250°C for 120 hours				
Solvent-washed type I	6.5	26.2	67.3	7.0
Demineralized type I	2.0	32.1	65.8	3.4
250°C for 240 hours				
Solvent-washed type I	3.9	26.9	69.2	4.6

Table 4.6. Composition of organic matter in the residual solids after pyrolysis of various kerogen samples as determined by multi-CP ssNMR. Duplicate runs are indicated in the sample name in parenthesis (e.g. (1))

single-pulse spectra.

4.3.4   ssNMR of pyrolyzed samples

The analysis for pyrolyzed samples were separated by solids, liquids (soluble in chloroform), and gases. The solids were primarily analyzed multi-CP ssNMR to quantify the composition of the organic compounds in the solid residue. Table 4.6 details the compositional break down for all pyrolyzed samples (individually) in this work.

First, from the replicate solvent-washed data at 270°C and 120 h, we can see that the results are fairly repeatable, with both replicate runs having mostly similar compositions likely within the true precision of our measurements. Secondly, it is clear that the kerogens have been noticeably pyrolyzed, even at these low temperature conditions. In addition to being able to detect liquid and gaseous products as we discuss later on, the solid residue has a noticeably different composition than the starting material for both types of kerogen. Figure 4.7 makes it easier to see these compositional differences between the demineralized type I
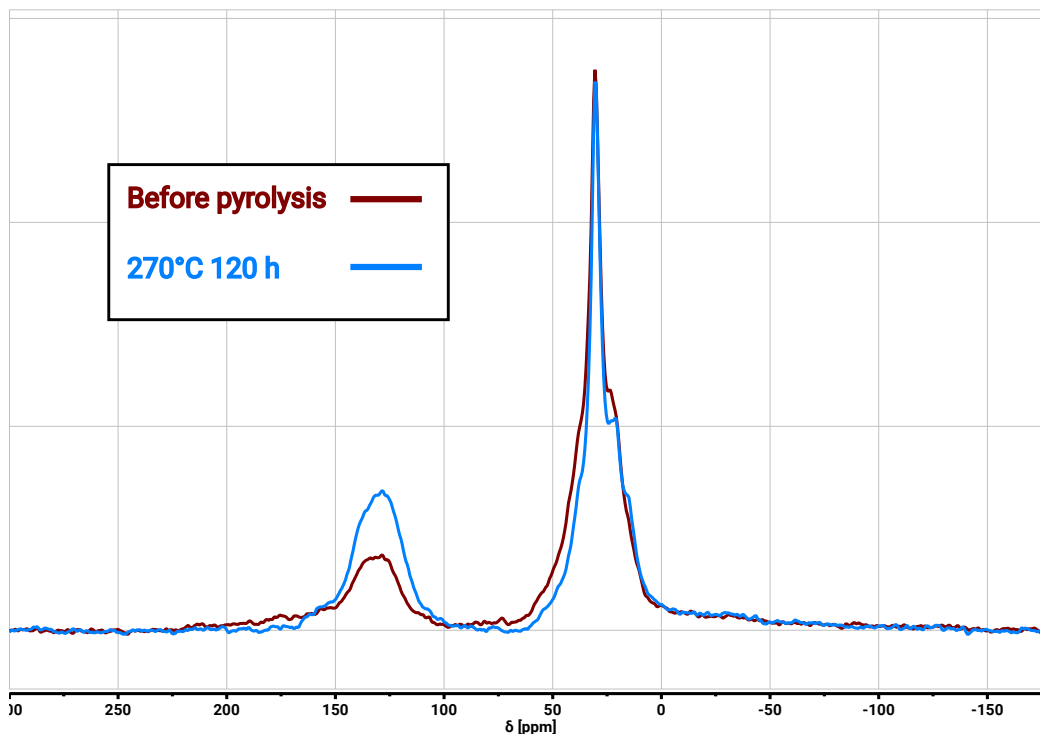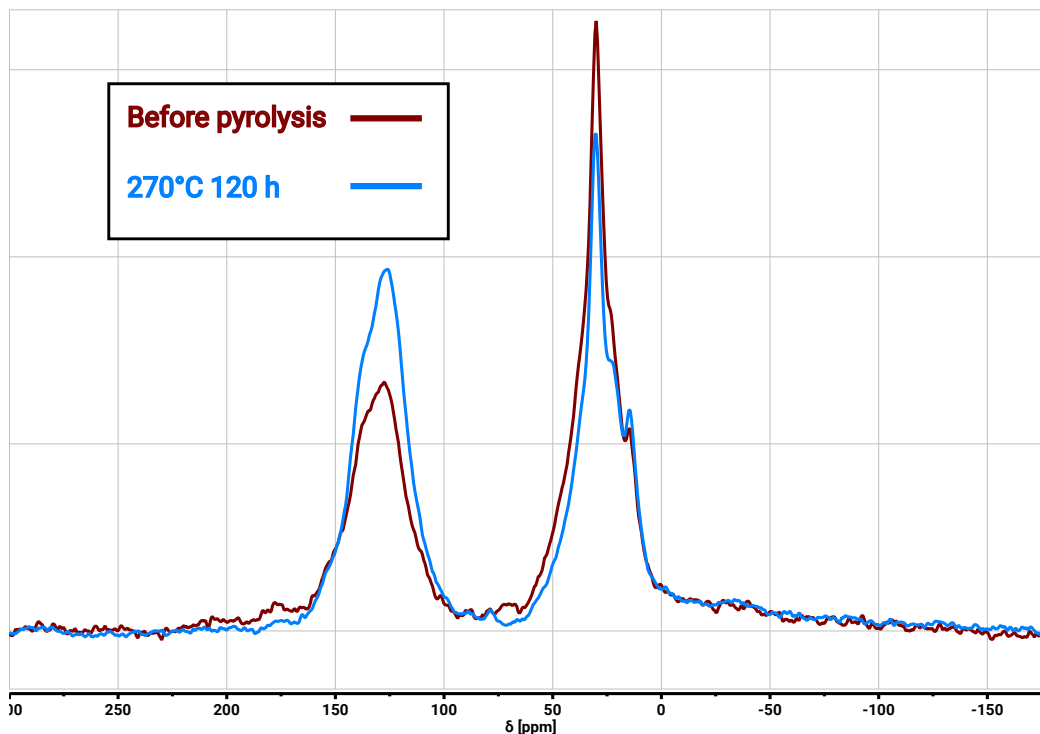
Figure 4.7. Comparison of demineralized type I kerogen before and after pyrolysis at 270°C for 120 h. The spectra have been normalized by area to showcase differences in composition

kerogen both before and after pyrolysis. The fraction of aromatic carbon in the pyrolyzed sample is noticeably higher after pyrolysis, with the pyrolyzed sample having less aliphatic carbon downfield of 30 ppm (usually where methine and quaternary carbons appear along with some methylene carbons). Additionally, the pyrolyzed sample seems to have released most of the carbonyl/oxl carbons and aliphatic C-O carbons.

To rule out the possibility that these compositional differences are not a result of pyrolysis but rather due to the solid-liquid extraction (SLE) procedure performed after pyrolysis, a small amount of demineralized type I kerogen (unpyrolyzed) was sent through the SLE procedure. As seen in Figure 4.8 shows that the composition of the kerogen before and after SLE is essentially the same. Therefore, it is clear that the compositional changes in the residual solids after pyrolysis are from the pyrolysis itself.

Figure 4.9 show a comparison of type II demineralized kerogen before and after pyrolysis at 270°C and 120 h. Much like the type I sample at these conditions, the pyrolyzed demineralized type II kerogen has noticeably more aromatic carbon than its starting material. The loses in aliphatic carbon are seen primarily downfield of 20 ppm, including a noticeable
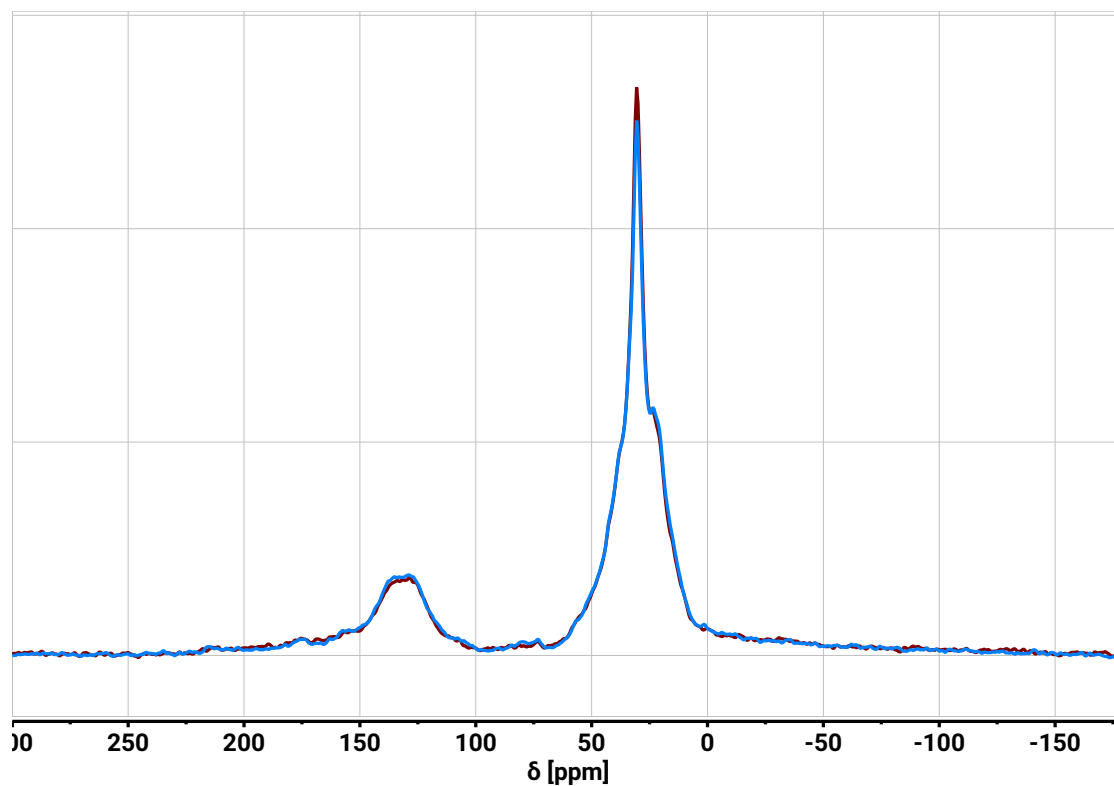
Figure 4.8. Comparison of multi-CP spectra of the demineralized type I kerogen starting material (red) and a small amount of this starting material that underwent the solid-liquid extraction procedure (SLE). Both samples are unpyrolyzed. This confirms that difference in multi-CP spectra between pyrolyzed and unpyrolyzed samples are not simply due to the SLE procedure.
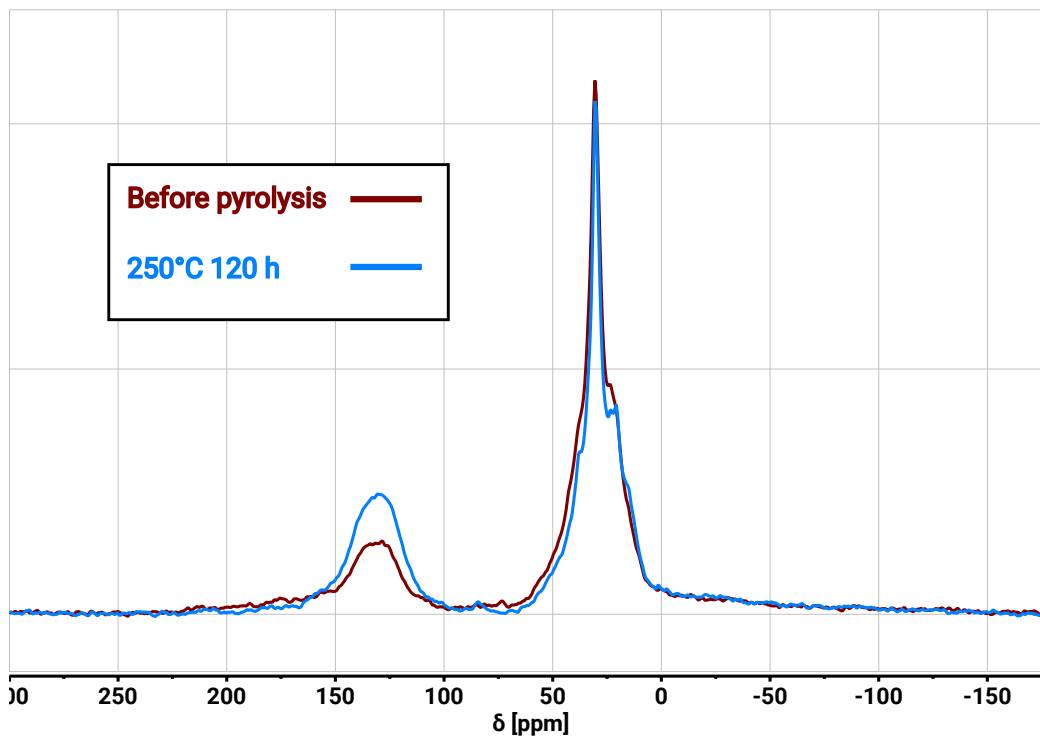
Figure 4.9. Comparison of demineralized type II kerogen before and after pyrolysis at 270°C for 120 h. The spectra have been normalized by area to showcase differences in composition

drop in methylene carbons around 30 ppm. Like the demineralized type I kerogen, the demineralized type II sample appears to lose sizeable portion of its carbonyl/oxyl and aliphatic C-O carbons.

It is also interesting to compare the pyrolyzed type I kerogen samples that were solvent-washed as opposed to those that went through the full demineralization procedure. Figure 4.10 compares the pyrolyzed solvent-washed type I kerogen both to its starting material and to the demineralized type I kerogen sample that was pyrolyzed at the same conditions. It is clear from the comparison with its starting material that the residual solids after pyrolysis for solvent-washed sample are compositional different from the starting material, with a shift towards more aromatic carbon at the expense of aliphatic carbon. However, this sample also shows distinct compositional differences with the pyrolyzed demineralized sample. It appears that the demineralized sample is more reactive, showing an even greater shift from aliphatic carbon to aromatic carbon. Furthermore, it appears that the shale sample retained more of their oxygenated carbons, particularly their carbonyl/oxyl carbons.

We see similar results at the lower temperature point as well. Figure 4.11 shows the
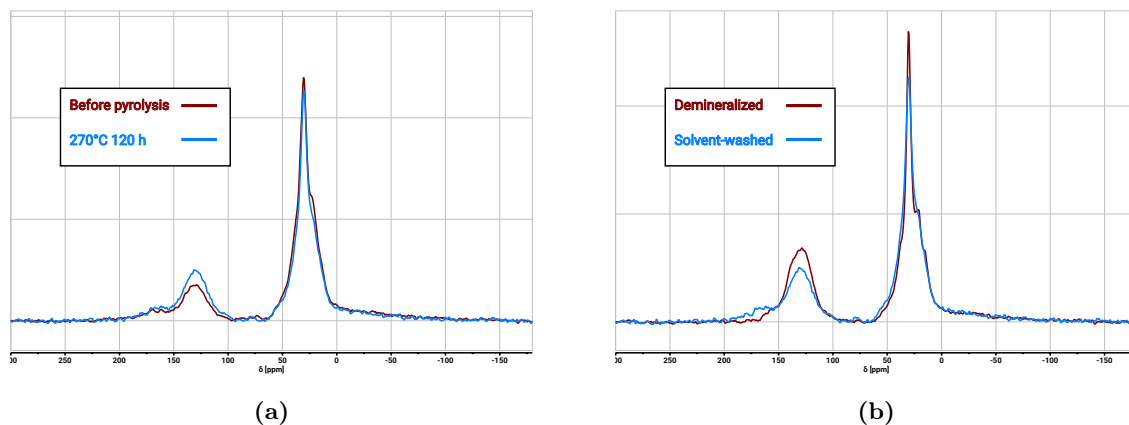
Figure 4.10. Comparison of multi-CP spectra of (a) solvent-washed kerogen before and after pyrolysis at 270°C for 120 h and (b) demineralized versus solvent-washed kerogen both pyrolyzed at 270°C at 120 h. In both cases the spectra are area normalized and enlarged to show compositional differences

multi-CP spectra of type I demineralized kerogen after pyrolysis at 250°C for 120 h. In comparison with the starting material, once again the pyrolyzed sample has notably more aromatic carbon at the expense of aliphatic carbon. Additionally, the aliphatic C-O and carbonyl/oxyl carbons appear to have mostly been released. Figure 4.12 adds back in the demineralized type I sample pyrolyzed at 270°C and we see that the two samples appears fairly similar despite the difference in temperature, though the 270°C sample does have a higher fraction of aromatic carbon, so it appears to have achieved a higher conversion as expected.

Finally, Figure 4.13 shows the difference between the demineralized and solvent-washed type I kerogens at this lower temperatures. Once again, the solvent-washed sample appears to be less reactive than the demineralized kerogen. Furthermore, the solvent-washed sample seems to have retained more of its aliphatic C-O and carbonyl/oxyl carbons.

From the ssNMR data for the pyrolyzed solids, it seems pretty conclusive that the samples are undergoing noticeable changes in composition during pyrolysis despite the relative low temperatures. Furthermore, it is interesting that the demineralized samples appear to be more reactive and retain less aliphatic C-O and carbonyl/oxyl carbons compared the solvent-washed samples at the same conditions. There are many reasons why this could be the case. For one, it is possible that demineralization procedure altered the kerogen in subtle ways. While Figure 4.6 makes it clear that there are no significant compositional changes between these starting materials, it is still possible that small chemical changes or

Figure 4.11. Comparison of demineralized type I kerogen before and after pyrolysis at 250°C for 120 h. The spectra have been normalized by area to showcase differences in composition

changes to the morphology of the kerogen occurred that are significant from a reactivity stand point even if they aren't from a compositional stand point. It is also not clear what effect if any the large fraction of inorganic matter in the solvent-washed sample is having on the underlying chemistry. It should be noted that this behavior matches with some previous observations in the literature. Baruah and Tiwari performed TG-FTIR pyrolysis on a type II Indian oil shale, and observed that demineralized samples began generating lighter hydrocarbons much earlier than oil shale samples with minerals still present.[34] They also suggested that the minerals may be having a catalytic effect on coking and cracking reactions in the generated oil, so these two phenomenon are not mutually exclusive.

4.3.5   GC-FID/TCD for gaseous products

After the gold tubes containing the pyrolyzed samples were retrieved from the reactor, gaseous products were quantified using using GC-FID/TCD by puncturing the gold tube inside of headspace vials to extract the gaseous product with a syringe. Figure 4.14 shows a typical FID chromatogram of the products observed. Note that since the instrument for
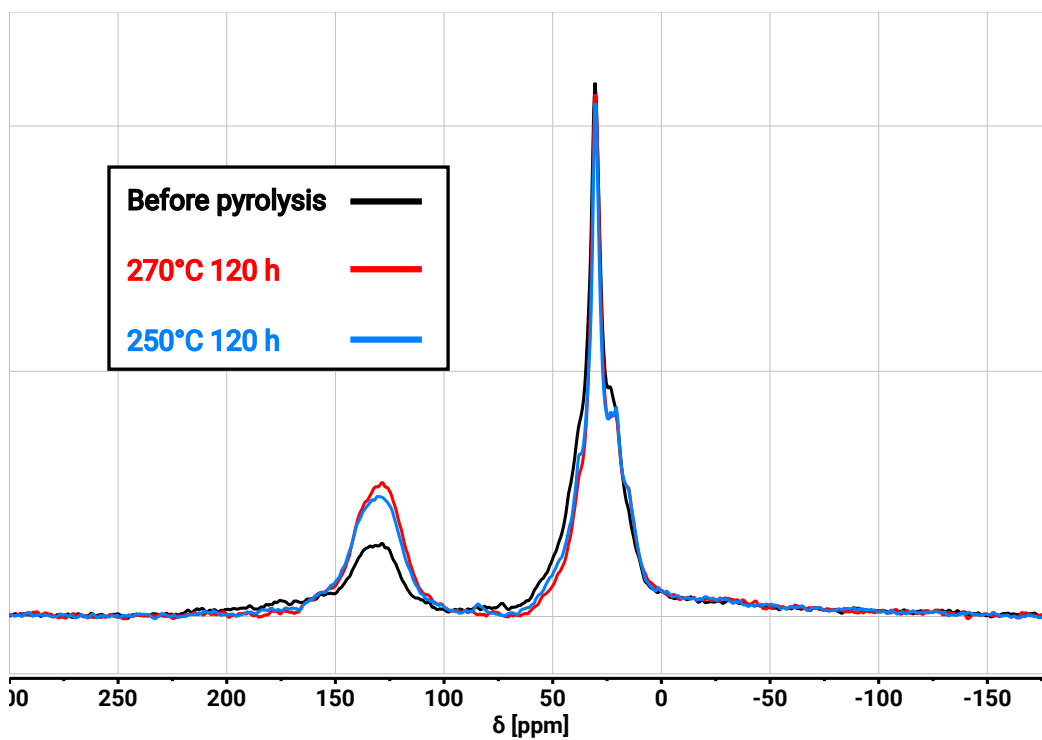
Figure 4.12. Comparison of demineralized type I kerogen before pyrolysis, after pyrolysis at 250°C for 120 h, and after pyrolysis for 270∘C for 120 h. The spectra have been normalized by area to showcase differences in composition
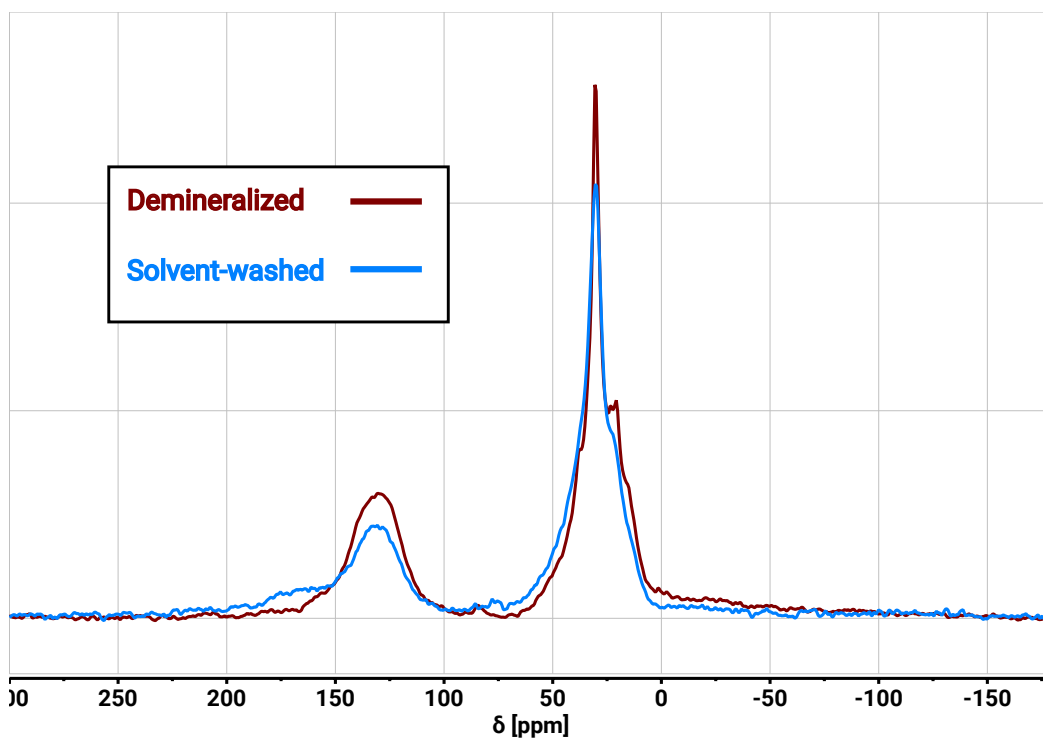
Figure 4.13. Comparison of multi-CP spectra of demineralized and solvent-washed type I kerogen both pyrolyzed at 250°C for 120 h. The spectra have been normalized by area to showcase differences in composition.
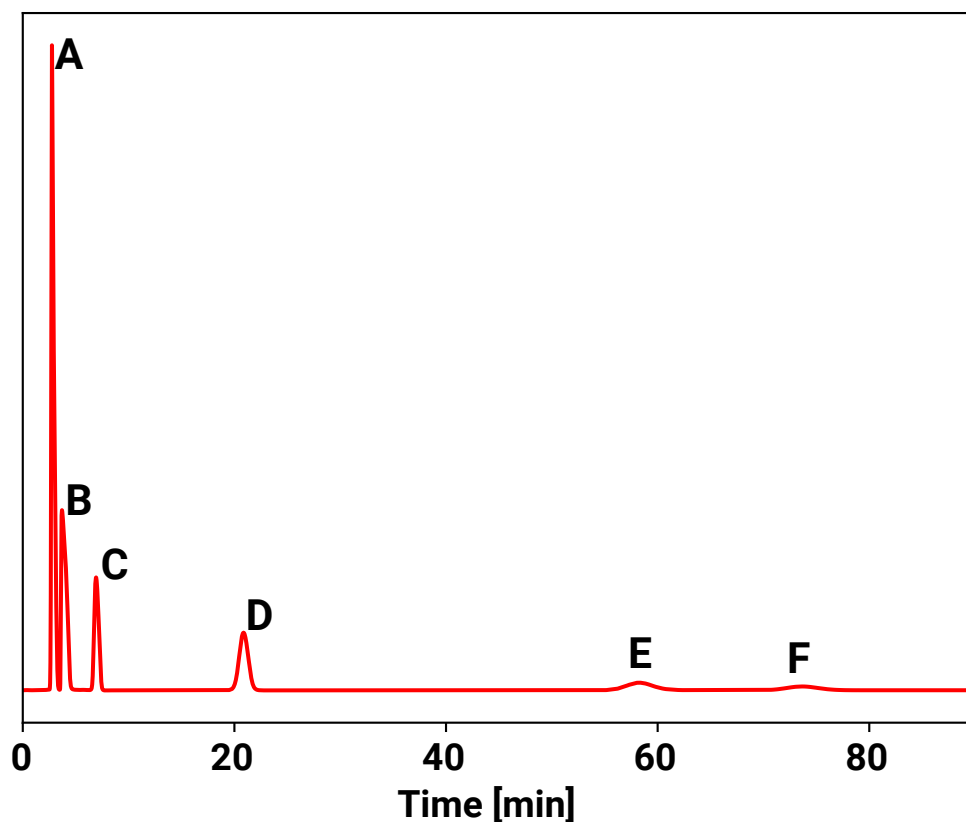
Figure 4.14. GC-FID chromatogram of gaseous products formed from the pyrolysis of demineralized type I kerogen at 270°C and 120 h. Peaks include: (A) methane, (B) CO_2, (C) ethane, (D) propane, (E) isobutane, (F) n-butane.

this analysis was equipped with a methanizer, CO_2 can be seen in the FID as well, though this signal includes background CO_2. Hershkowitz et al. noted in their pyrolysis study of Green River kerogen that mineral carbonates are the source of CO_2.[14]

As seen in Figure 4.14, due to the long GC runtime we are able to see gases as large as C4, with both isobutane and n-butane having signals large enough to be quantified. The gas observed are mostly the saturated alkane gases; however, on some spectra we were able to see small peaks above the baseline for gases such as ethylene and propylene. These peaks were not quantified due to the peaks barely being above the baseline. In all samples methane was the dominant gas observed, with a significant drop-off for ethane and propane, and a further drop off for the butane isomers. To quantify the relative composition of the gaseous products, the peak areas were assumed to be proportional to the mass of the species (this is equivalent to assuming that the molar composition is proportion to the peak area times

Sample	Methane	Ethane	Propane	Isobutane	N-butane
270°C for 120 hours					
Solvent-washed type I kerogen (1)	1	0.206	0.149	0.066	0.031
Solvent-washed type I kerogen (2)	1	0.188	0.126	0.057	0.035
Demineralized type I kerogen	1	0.301	0.319	0.125	0.075
Demineralized type II kerogen	1	0.480	0.238	0.035	0.062
250°C for 120 hours					
Solvent-washed type I kerogen	1	0.101	0.047	nq	nq
Demineralized type I kerogen	1	0.162	0.106	0.045	0.022
250°C for 240 hours					
Solvent-washed type I kerogen	1	0.126	0.065	0.027	0.010

Table 4.7. Ratio of mass fractions of gaseous products relative to methane for pyrolyzed kerogens at various conditions. Due to abnormalities with the baseline, some peaks are labeled "nq" for not quantified, although these peaks were observed.

the number of carbons as is typical) and normalized to give methane a value of 1 for easy comparison. Table 4.7 shows the relative composition of the gaseous products observed for the experimental runs in this work.

The replicate runs for the solvent-washed kerogen samples pyrolyzed at 270°C and 120 hours shows that the results are reproducible, within a few percent relative to methane. Given this, we can safely say that the demineralized kerogen samples give off a higher fraction ethane and propane relative to methane than the corresponding solvent-washed samples across the various conditions studied. Given that the amount of ethane and propane (relative to methane) increased for the solvent-washed sample pyrolyzed at 250°C when the time was increased from 120 h to 240 h, we take the position that the increase in ethane/propane relative to methane is indicative of higher conversions. If true this would be consistent with the ssNMR data for the residual solids after pyrolysis, where we noted that the demineralized samples showed higher conversion relative to their corresponding solvent-washed samples. Furthermore, this would suggest that a much lower conversion was achieved for both sample types at 250°C as opposed to 270°C as expected. Other interesting observations include the fact that among the demineralized samples at 270°C and 120 h the type II sample favors more ethane to propane while ethane and propane are about equal in the type I sample. Finally, isobutane was more abundant than n-butane in the type I kerogen samples, whereas the opposite is true in the type II samples. Work by Connan and Cassou suggests that ratios of isobutane/butane greater than even 0.8 is usually a sign of thermal immaturity, which we would expect given the low temperatures of this work.[35]

4.3.6 GCxGC-FID/MS Analysis of Liquid Pyrolysis Extracts

After extracting pyrolysis products into deuterated chloroform, these liquid products were analyzed by GCxGC-FID/MS. It should be noted, though, that the inlet to the GC is only 350°C and to prevent damage to the column the primary column only reaches 300°C, so heavier products may not be observed in GCxGC-FID/MS. Figure 4.15 shows the GCxGC-MS chromatogram (though all quantifications were performed using the GCxGC-FID chromatogram) of pyrolysis products for demineralized type I kerogen pyrolyzed at 270°C and 120 h.

As seen in Figure 4.15, there are hundreds of products that are visible of a wide variety of types. Most of the prominent peaks were successfully identified using the available MS data, though we were conservative with this identification, choosing only to ID a peak only when both the MS data was convincing and when the compound had an expected boiling point near the other peaks in the area. In total we were able to identify more than 60 peaks (circled on the chromatogram), though it is likely possible to identify many more peaks if we take a less conservative approach.

The sub-figures of Figure 4.15 also distinguishes between the identifiable peaks based on the types of compounds observed. The largest type of compound observed were the linear alkanes, with most of the mass coming from C20 and above (up to C31 can be observed). Below C20, in addition to linear alkanes we also see peaks for isoprenoids, starting with 2,6,10,14-tetramethylheptadecane, along with phytane and pristine, and going all the way down to 2,6-dimethylundecane. We also see many aromatic compounds, including toluene, and tri/tetra-methylated naphthalenes. The mass of these compound is much smaller than the total mass of all aliphatics, but larger than what is suggested by NMR (discussed later on), which is further indication that not all compounds are visible in GCxGC. We also see peaks for oxygenated compounds, which are mostly ketones, aldehydes, and ethers. Of the peaks we identified, all of the oxygenated compounds are aliphatic with the exception of diphenyl ether. Finally, there were some additional aliphatic species that we identified, most notably hopane species in the upper right-hand corner of the chromatogram. A full listing of all identifiable species and their relative composition can be found in Appendix C.
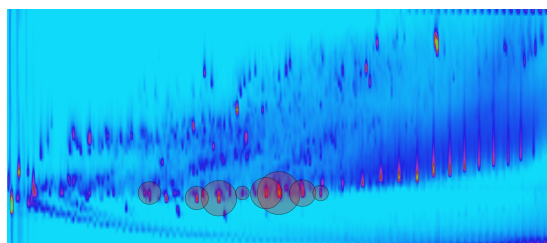
In addition to the internal standard, we also observed some contaminant peaks that were deemed to not have been part of the pyrolysis products. For example, both tetra-
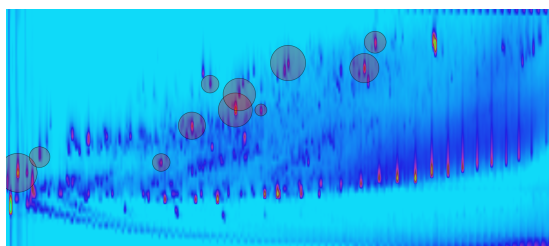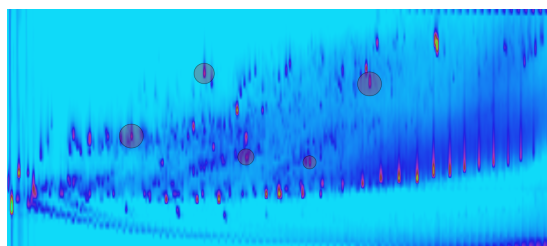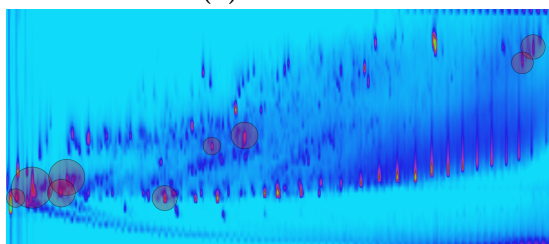
103
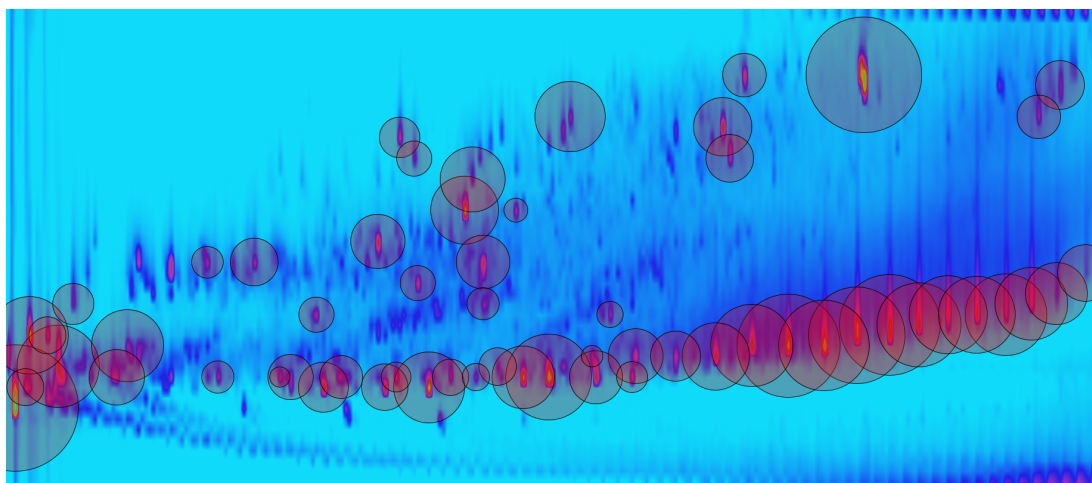
(a) Identifiable peaks
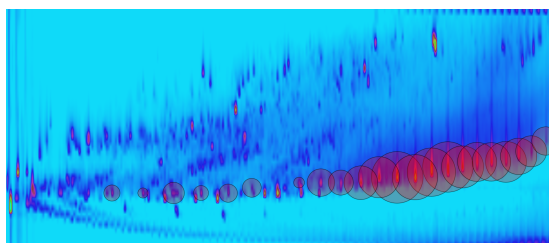
(b) Linear alkanes

(c) Isoprenoids
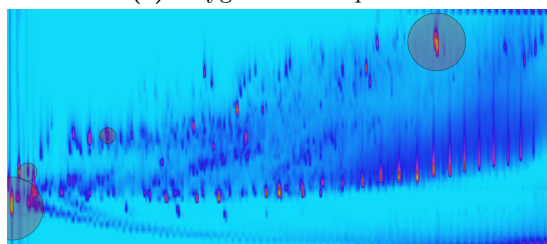

(d) Aromatics

(e) Oxygenated compounds

(f) Miscellaneous aliphatics

(g) Contaminants and internal standard

Figure 4.15. GCxGC-MS chromatogram of pyrolysis products produced from pyrolyzing demineralized type I kerogen at 270°C and 120 h. Those peaks that could be easily identified using the MS data are highlighted, and the sub-figures distinguish between the various types of compounds observed.

chloroethylene and hexachloroethane were observed in the chromatogram. These products likely formed from the solvent, deuterated chloroform, during sonication for the extraction. At high power, sonication is a known technique for eliminating tri-halomethane species in water, with chloroform being the most reactive and iodoform being the least reactive.[36] We also identified a large amount of 2,2'-methylenebis[6-(1,1-dimethylethyl)-4-ethyl-phenol as a contaminant, which elutes in the upper right-hand corner of the chromatogram. Given its prevalence, we determined this to be a contaminant, though it is unclear how this compound entered the sample (it is not visible in GCxGC spectra from washing the starting materials with chloroform prior to pyrolysis).

One type of compounds not observed in this work that has been seen in other works are carotenoids (tetraterpenoids). For example, Wang et al. performed pyrolysis-gas chromatography (py-GCxGC-FID) of Green river kerogen and observed a significant amount of carotenoids that appear at times after C33 and slightly higher up on the secondary column.[37] While Wang et al. used a lower inlet temperature of 300°C, they ramped to a much higher temperature of 390°C, which we were unable to do in this work without damaging the column. Given that we can observe up to C31, we likely just missed seeing the tail end of these carotenoids.

Given the low temperature pyrolysis of this work, we expect the results that are consistent with thermal immaturity. There are a couple of observations that point to this being the case. For example, the ratio of C18/phytane usually increases with thermal maturity. We observe a C18/phytane ratio of 0.94 for demineralized type I Green River kerogen after 270°C and 120 h. For comparison, Chong et al. performed isothermal pyrolysis on Green River kerogen at much higher temperatures of 400°C, 425°C, and 440°C.[38] Even after just a few minutes, they observed a C18/phytane ratio of at least 2-3, and found that this ratio would grow to as high as 5-7 with longer reaction times. Given that both studies use the same starting material, this shows that as expected the thermal maturity in this work is low.

It should be noted that Chong et al. separated their products effectively into two products: an oil product that was collected in a dry-ice-cooled trap, and a pyrobitumen, which was extracted from the pyrolyzed shale with benzene. The oil products had an average molecular weight of around 250 g/mol, while the pyrobitumen had an average molecular weight closer to 1000 g/mol. At small reaction times the oil product had an atomic H/C

ratio around 1.82, while the pyrobitumen had an atomic H/C ratio closer to 1.7. We mention this because at low reaction times, Chong et al. observed much more of their pyrobitumen product compared to their oil product. While we did not fractionate our liquid extracts, the products that can be seen on GCxGC-FID/MS are likely similar to the oil products of Chong et al.. Indeed, from the products we identified we see a mass-averaged atomic H/C ratio of 1.80, which is very similar to that of the oil product. Given this, it is likely that the majority of the pyrolysis products are not visible in the GCxGC chromatograms of this work.

Other supporting observations include the fact that the pristane/phytane (Pr/Ph) ratio for demineralized type I Green River kerogen after 270°C and 120 h is 1.8-2.3. Koopmans et al. performed isothermal hydrous pyrolysis experiments on Green River kerogen for 72 h for temperatures ranging from 240°C to 350°C.[39] At 240°C, they observe a Pr/Ph ratio of about 0.6. This ratio starts to increase dramatically around 280°C and levels off around 2 at higher temperatures. The Pr/Ph of our demineralized sample is more consistent with that of the higher temperatures. This can perhaps be explained by a combination of the longer reaction times in this work, as well as the fact that the demineralized samples in this work appear to be more reactive than shale samples like those used by Koopmans et al., and even differences between hydrous and non-hydrous pyrolysis.

If we take all of the information so far together, including the significant compositional changes of solids, the markers indicating low thermal maturity as expected, and the higher Pr/Ph ratio, the evidence so far points to the likely formation of many precursors to lighter oil and gas products in the liquid phase (the formation of an initial pyrobitumen and a small amount of oil) without achieving a higher amount of conversion among these precursors to lighter products.

4.3.7 Solution-phase NMR of Liquid Pyrolysis Extracts

Given that it is likely that there is a large fraction of liquid products that are not observable on GCxGC-FID/MS, we decided to perform solution-phase NMR on the liquid pyrolysis products. Figure 4.16 shows a typical ^{1}H spectrum that was obtained.

Most of the observed ^{1}H appears in the aliphatic window of 2 - 0.5 ppm, with peaks for the internal standard (iodoform) and tetramethyldisilane (TMS) also visible. The peak at 7.26 ppm is the residual peak for chloroform, though the tiny shoulder up-field might be
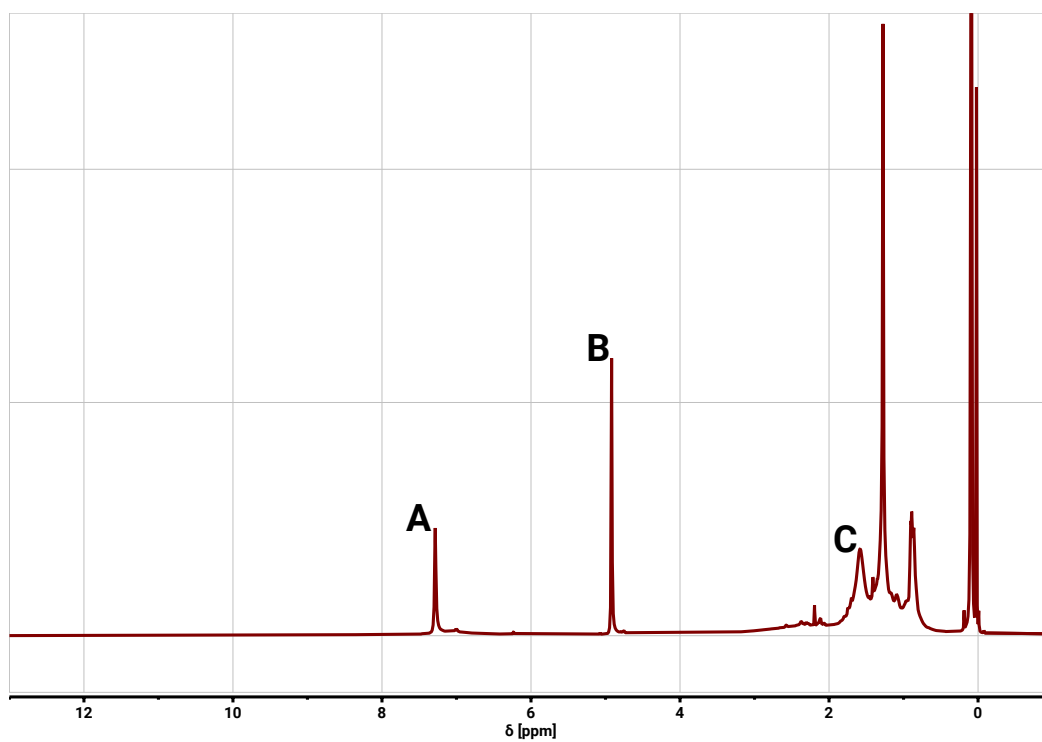
Figure 4.16. Solution-phase ^{1}H NMR of liquid extracts of demineralized kerogen after pyrolysis at 250°C and 120 h. In addition to the broad peaks for aliphatic ^{1}H from 0.5 ppm - 2 ppm, there are peaks for (A) residual H-CCl$_3$ from the solvent, (B) H-CI$_3$ peak for the internal standard, and (C) trace water peak around 1.6 ppm within the broad aliphatic peak.
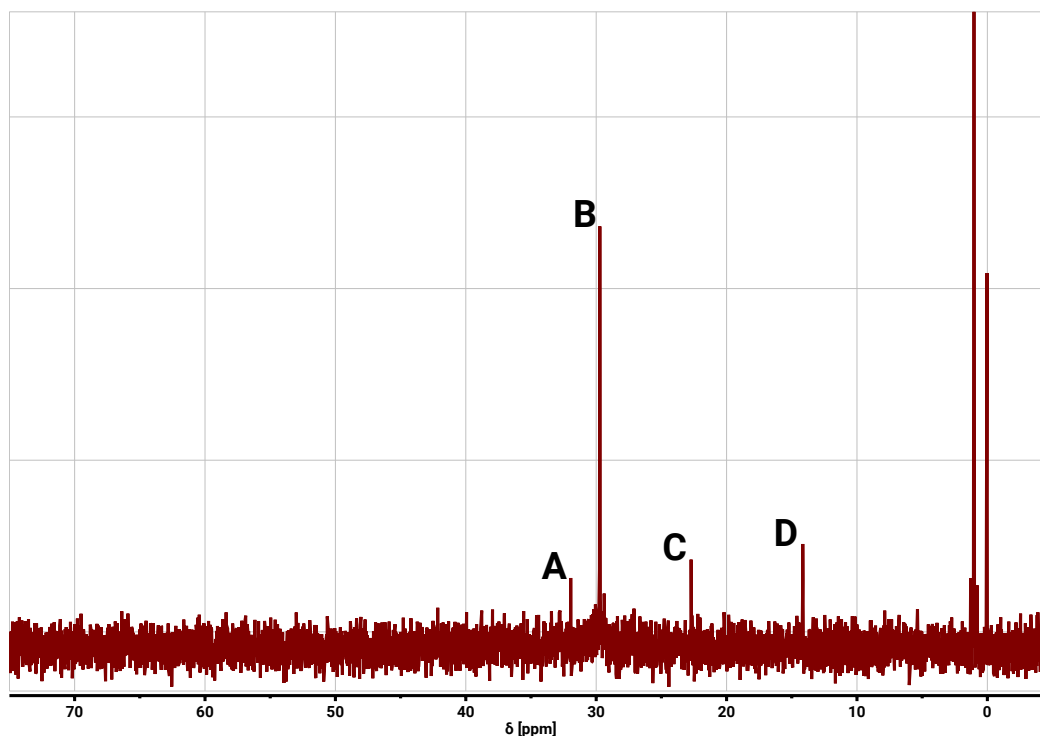
Figure 4.17. Solution-phase ^{13}C NMR of liquid extracts of demineralized kerogen after pyrolysis at 250°C and 120 h. Despite many scans, the aliphatic carbon signals are barely above the noise of the baseline, but 4 peaks are still visible: (A) tertiary carbons, (B) methylene carbons, (C) methylene carbons adjacent to a primary carbon, and (D) methyl carbons

due to aromatic compounds. Although it is difficult to be certain due to the residual solvent peak, it appears that the amount of aromatic ^{1}H is tiny in comparison to the amount of aliphatics. A peak for residual water in the solvent is also observed around 1.6 ppm. Before quantification, this water peak was removed by matching to the peak height at 1.6 ppm and using the peak width observed in a blank run of solvent.

Figure 4.17 shows the aliphatic portion of the ^{13}C NMR spectra for the same sample. We did not observe any peaks above noise in the aromatic portion of the spectra, and the iodoform peak of the internal standard is far upfield and not shown. Despite using 2048 scans, the peaks are barely above the noise, making quantification difficult. It all but certain that due to the lack of signal/noise that there are several peaks that we cannot observe. For example, Fletcher et al.[24] was able to observe many more peaks after pyrolysis at higher temperatures, using assignments from Dalling et al. to identify various peaks.[40] This should lead to an underestimation of the amount of carbon in the liquid extracts. Using the peak assignments of Dalling et al., though, we see that the most prominent peak is that

of methylene carbons as expected. a small amount of tertiary carbons can be observed as well, along with methyl carbons.

Given the low signal-to-noise of the ^{13}C spectra, we decided to first quantify the amount of ^{1}H. By using the internal standard method, we are able to determine the total mass of ^{1}H that was extracted into the liquid phase. Table 4.8 shows this mass of ^{1}H for the pyrolyzed samples were iodoform was used as the internal standard. Since the initial amount sample pyrolyzed, we can convert this into a mass fraction, X_H, which is the fraction of mass of the entire sample that ended up as ^{1}H in the liquid extracts. This quantity is not as useful as X'_H, which is the ratio of the measured mass of ^{1}H to the mass of organic matter in the original sample using the data from Table 4.5.

With this information, we can estimate what fraction of the original organic mass made it into the liquid extracts, X'. Even without quantification from the available ^{13}C NMR data, we know that a typical heavy oil has an atomic H/C ratio between 1.5 and 2 (in fact in this suspected thermally immature regime data from Chong et al.[38] suggests that the products should have an atomic H/C ratio closer to 1.7 to 1.8. We will consider the broader range of 1.5 - 2 to be conservative). Using this ratio (and only considering the mass of carbon and hydrogen) Table 4.8 should a likely range for X', with lower fractions coming from an H/C ratio of 2 (less carbon to add) and higher fractions coming from an H/C ratio of 1.5. We expect the true X' to fall somewhere in this range, and that this estimate is more accurate than incorporating the ^{13}C NMR data directly. For example, using the ^{13}C data yielded atomic H/C ratios usually between 2 and just over 3, which is obviously too high, and also consistent with our assumption that we are unable to quantify all of the peaks (i.e. carbon mass appears to be missing yielding a high H/C ratio).

As seen in Table 4.8, the results appear fairly repeatable, with both solvent-washed samples at 270°C and 120 h agreeing within a few percent for X'. The values for X' across all samples indicate that somewhere between 1/4 and 1/3 of all of the starting organic mass went into the liquid extracts, which is a significant amount. Furthermore, comparing the demineralized and solvent-washed samples at 250°C and 120 h shows that a higher fraction of the starting organic mass ends up in the liquid extracts, consistent with our previous observations that the demineralized samples appear to be more reactive.

Finally, we note that Miknis et al.[13] showed that the total mass of aromatic carbon in the residual solids is roughly equal to that of the mass of aromatic carbon in the starting

Sample	Starting Mass [mg]	Measured H [mg]	X_H [%]	X'_H [%]	Likely X' [%]	H/C Needed [-]
270°C for 120 hours						
Solvent-washed type I kerogen (1)	305.6	4.33	1.42	4.04	28.3-36.4	1.95
Solvent-washed type I kerogen (2)	310.8	4.13	1.33	3.80	26.6-34.2	1.82
250°C for 120 hours						
Solvent-washed type I kerogen	298.3	3.66	1.23	3.50	24.5-31.5	1.83
Demineralized type I kerogen	87.7	3.95	4.50	4.85	33.9-43.6	1.85

Table 4.8. Quantification of pyrolysis liquid extracts from solution-phase NMR.

material (although of course the composition of aromatic carbon goes up due to the loss of aliphatic carbon) regardless of pyrolysis temperature, including those covered in this work. If we make this assumption that the mass of aromatic carbon before pyrolysis is exactly equal to the mass of aromatic carbon in the residual solids, we using the data from Table 4.5 and Table 4.6 we can work out what X′ should be based on the ssNMR observations, and use this to work out the atomic H/C for the liquid products given the known mass of ^{1}H from solution-phase NMR of the liquid extracts. Table 4.8 shows the H/C ratio needed to make this assumption hold exactly (e.g. if the hydrocarbons in the liquid extracts had an atomic H/C ratio of 1.85 for the demineralized sample pyrolyzed at 250°C and 120 h, the measured amount of hydrogen plus the predicted amount of carbon in the liquid extracts would have the same mass as what was lost from the solids as measured via ssNMR). These needed H/C ratios are in the 1.8 - 1.9 range, which is certainly in the correct ballpark. This matches up almost perfectly with the measured H/C ratio of products visible in GCxGC. However, it is likely that the true H/C ratio for the liquid extracts is less than 1.8 given that the GCxGC is not able to see higher molecular weight species that in the literature have lower values for the H/C ratio. It should be noted that there are a couple sources of error that explain any discrepancies here, including the fact that we have ignored the mass of heteroatoms in the analysis, and that it is possible that we have underestimated the amount of water present.

4.4 Conclusions

In this work, low-temperature closed-system pyrolysis was performed on demineralized type I (Green River) and type II kerogen shales along with "solvent-washed" type I shale with minerals in order to obtain valuable data for validating detailed kinetic models of kerogen maturation. Although this work typically involves the use of low-field NMR instruments, we showed that this analysis can be done using high-field while still obtaining quantitative results. This coupled with the fact that the multi-CP method is able to achieve quantiative results with a shorter experimental time than more typically used variable contact time experiments means that these kinds of experiments are more accessible given the prevalence of high-field instruments. Despite the relatively low temperatures of this work, data collected on the residual solids and liquid extracts after pyrolysis show significant conversion of kerogen into soluble bitumen and oil products (as much as 1/4 or 1/3 of the original organic mass). Many markers and observations from both the gaseous and liquid products are consistent with achieving low thermal maturity as might be expected at these low temperatures, indicated that most of the mass formed is a higher molecular weight bitumen as opposed to oil that is not visible on GCxGC. That being said, many oil compounds are visible in GCxGC, including linear alkanes, isoprenoids, oxygenated compounds, and a sizeable portion of aromatics like toluene, xylenes, and naphthalenes. Future work will entail identifying even more of the minor products present in the GCxGC data, which could provide valuable mechanistic insight or information about the starting precursors or both. While a sizeable portion of aromatics can be seen in the oil products, the oil plus bitumen (i.e. the full liquid extracts) appear to be sparse in aromatics based on solution-phase NMR data for these conditions, suggesting that lack of aromatic compounds in the larger bitumen fraction at this stage in the pyrolysis.

Additional observations include the fact that the demineralized kerogen samples appear to be more reactive than shales that were only solvent-washed to remove the initial bitumen (but still contain mostly inorganic matter), even though other than the inorganic content the two samples have nearly identical macro-composition among organic compounds as can be measured by ssNMR. If the inorganic matter present in the solvent-washed shale is playing a catalytic role, it does not appear to speed up the formation of the initial soluble bitumen and oil products, though we can't rule out that the inorganic matter is playing a catalytic

effect to selectively produce certain products in the oil or bitumen. Otherwise, it is possible that undetectable differences in chemical structure between the solvent-washed sample and demineralized sample caused by the demineralization procedure could exist that are playing an out-sized role in reactivity. As a final explanation perhaps this is simply just a matter of the organic matter being more concentrated in the demineralized samples in the solid-phase, though the constant pressure nature of these experiments should keep the density of the oil, gas, and likely even the bitumen products roughly constant.

Taken altogether, this work shows that the initial formation of a soluble bitumen and oil phase occurs at relatively low temperatures and times, even if the thermal maturation of this products into oil takes longer reaction times or higher temperatures. This is valuable information for validating detailed kinetic models, especially when taken with the more abundant data available at higher temperatures.

4.5   References

(1)   *Kerogen: insoluble organic matter from sedimentary rocks*; Durand, B., Ed.; Editions technip: Paris, 1980.

(2)   Peters, K. E.; Walters, C. C.; Moldowan, J. M., *The Biomarker Guide: Volume 2, Biomarkers and Isotopes in Petroleum Systems and Earth History*, 2nd edition, OCLC: 1277067567; Cambridge University Press: Cambridge, 2007.

(3)   Siskin, M.; Scouten, C. G.; Rose, K. D.; Aczel, T.; Pabst, R. E. Detailed structural characterization of the organic material in rundle Ramsay Crossing and Green River oil shales. *Composition, Geochemistry and Conversion of oil shales* **1995**, ISBN: 978-94-010-4140-9, 978-94-011-0317-6, 143–158.

(4)   Lille, U.; Heinmaa, I.; Pehk, T. Molecular model of Estonian kukersite kerogen evaluated by 13C MAS NMR spectraq. **2003**, 6.

(5)   Ungerer, P.; Collell, J.; Yiannourakou, M. Molecular Modeling of the Volumetric and Thermodynamic Properties of Kerogen: Influence of Organic Type and Maturity. *Energy Fuels* **2015**, *29*, 91–105.

(6)   Tong, J.; Jiang, X.; Han, X.; Wang, X. Evaluation of the macromolecular structure of Huadian oil shale kerogen using molecular modeling. *Fuel* **2016**, *181*, 330–339.

(7)   Pan, S.; Wang, Q.; Bai, J.; Chi, M.; Cui, D.; Wang, Z.; Liu, Q.; Xu, F. Molecular Structure and Electronic Properties of Oil Shale Kerogen: An Experimental and Molecular Modeling Study. *Energy Fuels* **2018**, *32*, 12394–12404.

(8)   Resing, H. A.; Garroway, A. N.; Hazlett, R. N. Determination of aromatic hydrocarbon fraction in oil shale by 13C n.m.r. with magic-angle spinning. *Fuel* **1978**, *57*, 450–454.

(9) Opella, S. J.; Frey, M. H. Selection of nonprotonated carbon resonances in solid-state nuclear magnetic resonance. *J. Am. Chem. Soc.* **1979**, *101*, Publisher: American Chemical Society, 5854–5856.

(10) Solum, M. S.; Grant, D. M.; Pugmire, R. J. 13C Solid-State NMR of Argonne Premium Coals. *Energy and Fuels* **1989**, *3*, ISBN: 0887-0624, 187–193.

(11) Wilson, M. A., *N.M.R. techniques and applications in geochemistry and soil chemistry*, 1st ed; Pergamon Press: Oxford [Oxfordshire] ; New York, 1987.

(12) Miknis, F. P.; Netzel, D. A.; Smith, J.; Mast, M.; Maciel, G. E. 13C NMR measurements of the genetic potentials of oil shales. *Geochimica et Cosmochimica Acta* **1982**, *46*, 977–984.

(13) Miknis, F. P.; Szeverenyi, N. M.; Maciel, G. E. Characterization of the residual carbon in retorted oil shale by solid-state 13C n.m.r. *Fuel* **1982**, *61*, 341–345.

(14) HERSHKOWITZ, F.; OLMSTEAD, W. N.; RHODES, R. P.; ROSE, K. D., Molecular Mechanism of Oil Shale Pyrolysis in Nitrogen and Hydrogen Atmospheres In *Geochemistry and Chemistry of Oil Shales*; ACS Symposium Series 230, Vol. 230, Section: 15; AMERICAN CHEMICAL SOCIETY: 1983, pp 301–316.

(15) Behar, F.; Kressmann, S.; Rudkiewicz, J.; Vandenbroucke, M. Experimental simulation in a confined system and kinetic modelling of kerogen and oil cracking. *Organic Geochemistry* **1992**, *19*, 173–189.

(16) Burnham, A. K. Historical Perspective on the Maturation of Modeling Coal and Kerogen Pyrolysis. *Energy Fuels* **2021**, *35*, 10451–10460.

(17) Landais, P.; Michels, R.; Elie, M. Are time and temperature the only constraints to the simulation of organic matter maturation? *Organic Geochemistry* **1994**, *22*, ISBN: 0146-6380, 617–630.

(18) Gao, C. W.; Allen, J. W.; Green, W. H.; West, R. H. Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms. *Computer Physics Communications* **2016**, *203*, Publisher: Elsevier B.V., 212–225.

(19) Liu, M.; Grinberg Dana, A.; Johnson, M. S.; Goldman, M. J.; Jocher, A.; Payne, A. M.; Grambow, C. A.; Han, K.; Yee, N. W.; Mazeau, E. J.; Blondal, K.; West, R. H.; Goldsmith, C. F.; Green, W. H. Reaction Mechanism Generator v3.0: Advances in Automatic Mechanism Generation. *J. Chem. Inf. Model.* **2021**, *61*, 2686–2696.

(20) Broadbelt, L. J.; Stark, S. M.; Klein, M. T. Computer Generated Pyrolysis Modeling: On-the-Fly Generation of Species, Reactions, and Rates. *Industrial and Engineering Chemistry Research* **1994**, *33*, ISBN: 0888-5885, 790–799.

(21) Warth, V.; Battin-Leclerc, F.; Fournet, R.; Glaude, P. A.; Côme, G. M.; Scacchi, G. Computer based generation of reaction mechanisms for gas-phase oxidation. *Computers and Chemistry* **2000**, *24*, ISBN: 0097-8485, 541–560.

(22) Vandewiele, N. M.; Van Geem, K. M.; Reyniers, M.-F.; Marin, G. B. Genesys: Kinetic model construction using chemo-informatics. *Chemical Engineering Journal* **2012**, *207-208*, Publisher: Elsevier B.V. ISBN: 1385-8947, 526–538.

(23) Solum, M. S.; Mayne, C. L.; Orendt, A. M.; Pugmire, R. J.; Adams, J.; Fletcher, T. H. Characterization of Macromolecular Structure Elements from a Green River Oil Shale, I. Extracts. *Energy & Fuels* **2014**, *28*, 453–465.

(24) Fletcher, T. H.; Gillis, R.; Adams, J.; Hall, T.; Mayne, C. L.; Solum, M. S.; Pugmire, R. J. Characterization of Macromolecular Structure Elements from a Green River Oil Shale, II. Characterization of Pyrolysis Products by 13 C NMR, GC/MS, and FTIR. **2014**, DOI: 10.1021/ef500095j.

(25) Behar, F.; Vandenbroucke, M.; Tang, Y.; Marquis, F.; Espitalie, J. Thermal cracking of kerogen in open and closed systems: Determination of kinetic parameters and sto-ichiometric coefficients for oil and gas generation. *Organic Geochemistry* **1997**, *26*, ISBN: 0146-6380, 321–339.

(26) Vandegrift, G. F.; Winans, R. E.; Scott, R. G.; Horwitz, E. Quantitative study of the carboxylic acids in Green River oil shale bitumen. *Fuel* **1980**, *59*, 627–633.

(27) Johnson, R. L.; Schmidt-Rohr, K. Quantitative solid-state 13C NMR with signal enhancement by multiple cross polarization. *Journal of Magnetic Resonance* **2014**, *239*, Publisher: Elsevier Inc., 44–49.

(28) Lazar, C.; Cody, G. D.; Davis, J. M. A kinetic pressure effect on the experimental abiotic reduction of aqueous CO_2 to methane from 1 to 3.5kbar at 300°C. *Geochimica et Cosmochimica Acta* **2015**, *151*, Publisher: Elsevier Ltd ISBN: 0016-7037, 34–48.

(29) Wang, R.; Luo, Y.; Jia, H.; Ferrell, J. R.; Ben, H. Development of quantitative ^{13} C NMR characterization and simulation of C, H, and O content for pyrolysis oils based on ^{13} C NMR analysis. *RSC Adv.* **2020**, *10*, 25918–25928.

(30) Kolodziejski, W.; Klinowski, J. Kinetics of cross-polarization in solid-state NMR: A guide for chemists. *Chemical Reviews* **2002**, *102*, ISBN: 0009-2665, 613–628.

(31) Newman, R. Analysis of Results from Interrupted-Decoupling NMR Pulse Sequences Combined with High-Speed Magic-Angle Spinning R. *Journal of Magnetic Resonance* **1990**, *86*, 176–179.

(32) Raleigh, D. P.; Olejniczak, E. T.; Vega, S.; Griffin, R. G. An analysis of sideband suppression techniques in magic-angle sample spinning NMR. *Journal of Magnetic Resonance (1969)* **1987**, *72*, 238–250.

(33) Dixon, W. T.; Schaefer, J.; Sefcik, M. D.; Stejskal, E. O.; McKay, R. A. Total suppression of sidebands in CPMAS C-13 NMR. *Journal of Magnetic Resonance (1969)* **1982**, *49*, 341–345.

(34) Baruah, B.; Tiwari, P. Compositional and kinetic study of thermal degradation of kerogen using TG-FTIR , NMR , and microscopic study. *AIChE Journal* **2022**, *68*, DOI: 10.1002/aic.17396.

(35) Connan, J.; Cassou, A. Properties of gases and petroleum liquids derived from terrestrial kerogen at various maturation levels. *Geochimica et Cosmochimica Acta* **1980**, *44*, 1–23.

(36) Shemer, H.; Narkis, N. Sonolytic Destruction of Trihalomethanes by Ultrasonic Irradiation. *Isr. J. Chem.* **2006**, *46*, 27–32.

(37) Wang, F. C.-Y.; Walters, C. C. Pyrolysis Comprehensive Two-Dimensional Gas Chromatography Study of Petroleum Source Rock. *Anal. Chem.* **2007**, *79*, 5642–5650.

(38) Chong, S.-L.; Wu, R.-Y.; Miknis, F. P.; Turner, T. F. CHARACTERISTICS OF PYROBITUMEN AND OIL OBTAINED FROM GREEN RIVER OIL SHALE PYROLYSIS. *Fuel Science and Technology International* **1989**, *7*, 347–376.

(39) Koopmans, M. P.; Rijpstra, W. C.; Klapwijk, M. M.; de Leeuw, J. W.; Lewan, M. D.; Sinninghe Damsté, J. S. A thermal and chemical degradation approach to decipher pristane and phytane precursors in sedimentary organic matter. *Organic Geochemistry* **1999**, *30*, 1089–1104.

(40) Dalling, D. K.; Pugmire, R. J.; Grant, D. M.; Hull, W. E. The use of high-field carbon-13 NMR spectroscopy to characterize chiral centers in isopranes. *Magn. Reson. Chem.* **1986**, *24*, 191–198.

Chapter 5

Conclusions and recommendations for future work

Generating detailed kinetic models for large systems can be an arduous task that comes with many challenges due solely do the size of the system, which are on top of numerous challenges that comes with developing *ab initio* kinetic models in general. That being said, advances in automatic mechanism generation software, quantum chemistry methods, and experimental techniques have brought this task closer to being obtainable in practice. In just this work we were able to construct a detailed pyrolysis model for not just one but a mixture of three intermediate sized molecules. While molecules on the size of 20 heavy atoms pale in comparison to the size and complexity of something like biomass or kerogen, generating a model with thousands of species and hundreds of thousands of reactions that accurately predicts the underlying chemistry is impressive. Furthermore, with these models and existing technology we are able to gain considerable insight into the underlying chemistry of large and complex systems. That being said, it is clear from this work that many challenges and unanswered questions remain, those these hint at great next steps for future work.

In Chapter 2, we showed that RMG can be used to generate a detailed kinetic model to predict the pyrolysis chemistry of a mixture of dodecylbenzene, undecane, and toluene. Although this task came with challenges due to the size of the molecules and relatively low temperatures studied, it should not be too surprising that the final generated model was able to predict the underlying chemistry reasonably well. After all, many systems that involve the pyrolysis or combustion of hydrocarbons have been successfully studied with RMG. That being said, this work shows that RMG can be used to generate relatively large

kinetic models, especially when the underlying chemistry has been well studied.

That is not to say that the underlying chemistry here was fully understood prior to this work, nor that this work did not provide some insights here. In fact, the absence of isomerization products from initial models that were observed experimentally ultimately led to the determination of rate parameters for phenyl migration reactions by Khanniche et al.[1] (work by others). Given this success, and the fact that there are hundreds of unidentified peaks present in the experimental data, future work should entail identifying as many of these peaks as possible, as this might lead to new insights and calculations to capture even more of the underlying chemistry. One particular area of focus should be on larger alkane products, as it appears that the formation of these species might be responsible for the necessary hydrogen atoms to form the observed lighter products. This is especially important because if this is correct, currently the RMG model does not capture this behavior accurately. Although the model predicts the formation of a particular large alkane product with sufficient mass to explain the hydrogen atom balance, it does not predict the diversity of alkane species we see experimentally. Perhaps the structure of these species, once identified, will give us some insight into the relevant formation pathways and why they appear to be missing from the models so far.

In addition to gleaning more insights from the experimental data, there is one more task of future work related to this model that appears especially ripe for studying in the short term. In Chapter 2 we ultimately concluded that our assumption of ideal gas behavior when simulating the model, despite the near supercritical conditions of the actual experiments, was not the major source of error between the model and the experimental data. With recent advances, though, it should be possible to test this conclusion further, and perhaps it is even possible to not have to rely on this assumption going forward. For example, many mechanism simulation codes have recently incorporated more accurate equations of state (for example, the Peng-Robinson equation of state has been incorporated into Cantera). More importantly, though, the proliferation of machine learning to predicting chemical properties means that it is likely possible to accurately predict the necessary critical property data for all species in the model so that an equation of state like Peng-Robinson can be used when simulating the mechanism. It would be interesting to re-simulate the three-component surrogate model of this work using this approach to confirm our analysis about the ideal gas assumption, and perhaps to use this as the simulation strategy going forward.

While the modeling work in Chapter 2 was able to take advantage of already known thermochemistry and kinetic parameters for hydrocarbon pyrolysis, future work studying other large chemical systems will likely require model refinement through many quantum chemistry calculations. With regards to large systems, Chapter 3 showed that many modern quantum chemistry methods that can be used on large molecules are able to achieve great accuracy, such as the DLPNO-CCSD(T)-F12 methods. The work in Chapter 3 showed that while corrections are needed for each level of theory in order to realize accurate results, Melius-type BACs, Petersson-type BACs, and even isodesmic reactions (when used with care) are sufficient for this task. Going into this work, we expected isodesmic reactions to be of particular use for larger molecules, as there was some literature that suggested that this was the ideal approach for larger molecules. It was therefore at first a bit surprising that the isodesmic reaction approach was only ever on-par with the BAC approach even in the best case scenario with higher order reaction classes. In hindsight, perhaps this result should have been expected, as it was already known that BACs and isodesmic reaction are mathematically closely related, differing only in the number of reference species that directly affect the final correction. While using higher order reaction classes yielded more accurate results, it turned out that BACs were equally accurate for these species that could be calculated this way.

Given this, it is our view that with a proper implementation and a large and diverse reference set, current BAC approach are more than sufficiently accurate. It is tempting to try to apply more advanced methods (for example machine learning methods, though the small amount of available training data could be an issue) to yield better accuracy, and while there are probably gains to be had here, these gains are likely small, and may not be worth the effort. Instead, future work should focus on continuing to improve the user experience of the existing implementation, making it as easy to use and update as possible.

As for the experimental characterization of kerogen pyrolysis in Chapter 4, while this work highlights many experimental techniques that can be used to understand larger systems going forward, there are many remaining questions about this particular work. In this work, we showed that even at relatively low temperatures, a substantial amount of liquid products are formed. The liquid products appear to be mostly higher molecular weight bitumen as opposed to lower molecular weight oil that is observable with gas chromatography. Given this plus the suspected small amount of gaseous products formed, there are some serious

questions about the nature of this evolution of liquid products at these low temperatures. For example, one possible explanation is that these bitumen products are the result of chemical reactions that form these bitumen products from even larger precursors in the solid kerogen that are not soluble. An alternative explanation, though, is that some sort of physical change is occurring to the kerogen solid structure that is allowing these bitumen products to be leached out from the kerogen. In this explanation, the bitumen products could have been present prior to pyrolysis, and this heating only allowed for their removal. If this second explanation is correct, though, it is worth noting that the structural change appears to occur to a lesser extent at lower temperatures and shorter times, much like we would expect if the bitumen is instead being formed from chemically cracking larger precursors.

If physical/structural changes are the primary explanation, perhaps microscopy could yield some interesting insights. For example, recent work by Baruah and Tiwari[2] included microscopy analysis of kerogen heated at 10°C/min from room temperature up to 600°C. As early as 200°C they were able to observe changes to the porous structure of kerogen, including seeing visible fissures, which they attributed to being a result of forces from the products formed. It would be interesting to perform similar analysis on the solid residues of Chapter 4 (which have been preserved) to see what changes are present between samples at different temperatures and times.

Of course, these changes could be do to the formation of bitumen from cracking larger precursors as Baruah and Tiwari suggest, so this alone may not provide the necessary clarity. Because of this, it might be useful to follow up this work with model compound studies where compounds containing certain functional groups are pyrolyzed at these low temperatures. For example, it is often suggested that heteroatom chemistry, notably sulfur and oxygenated functional groups, are behind the observed reactivity under these conditions. Molecules that mimic these functional groups or suspect biological precursors could be pyrolyzed to see how reactive they are at these conditions.

It is our view that while forming the bitumen products (through either of the proposed mechanisms) happens at these relatively low temperatures, it is likely that the bitumen itself will not crack to form substantial amounts of oil and gas until much higher temperatures or longer reaction times. Understanding when this cracking will occur is essential, though, so future work should also entail pyrolyzing kerogen at slight higher temperatures on both sides of 300°C.

While this should yield useful experimental data to validate future kinetic models, it is likely important to understand the makeup of the bitumen itself. Unfortunately, these products are not visible in gas chromatography unlike lighter oil products, so not as much is known about these products. To obtain additional data, it may be worth exploring methods for determining the molecular weight distribution of the bitumen using various mass spectrometry techniques. The ideal technique would be to use matrix-assisted laser desorption ionization (MALDI), as this is a soft ionization technique that would yield the molecular ion peaks. In fact Kim et al.[3] were able to use this exact method to determine the molecular weight of a bitumen sample.

Another aspect not well understood from the experiments of Chapter 4 is what effect the minerals are having on the pyrolysis. It is often suggested that the minerals provide a catalytic effect; however, in these experiments the samples with a substantially higher mineral content produced less bitumen and liquid products, not more. The samples containing minerals seem to retain more of their aliphatic C-O and carbonyl/oxyl carbons. It is possible that this is being compounded with other effects. For example, if the formation of bitumen is really just due to structural or physical changes, perhaps the minerals are providing extra structure that is slowing this process down (but leaving the possibility open that the minerals catalyze cracking reactions of the bitumen itself). Either way, it would be best to isolate specific minerals from the kerogen matrix, and perform model compound studies with an without the presence of these minerals to confirm if a catalytic effect exists.

Taking a step back from all of this and looking at all of these aspects taken together, there are some final conclusions we can draw. While the large kinetic model generated in Chapter 2 is impressive, trying to model a system on the size of something like kerogen in full detail is simply not possible, or at least conceivable for the time being. Instead, we suspect that the experimental data obtained in Chapter 4 will be more useful for validating other modeling strategies such as fragment chemistry approaches, which while still detailed do not try to keep track of every species but rather keep track of the relevant functional groups.

This does not mean that fully detailed kinetic models of intermediate-sized model compounds will no longer be useful going forward. In fact, the results of this work seems to suggest quite the opposite. The low temperature data of Chapter 4 seems to say more about the starting material rather than giving us hints about the underlying chemistry at

play. Many of the species present in the data are well known bio-markers that seem to be direct products of expected precursors. Furthermore, when a new species is identified, there are always question about whether this species was present in the starting material, simply cleaved from the starting material, or formed through an intricate set of steps from the many intermediate products. This is much easier to sort out in model compound studies where the starting material is well known, and it is easier to speculate how certain products might have been formed. Generating large fully-detailed kinetic models of these model compounds then helps us understand if our current understanding of the underlying chemistry is sufficient to explain the observed data, or if there are new reaction pathways to be discovered.

Finally, given the results of this work that show that it is possible (though at times grueling) to generate detailed kinetic models to describe the pyrolysis of large hydrocarbon systems, the next obvious area to focus on would be to generate detailed kinetic models for biomass, where heteroatom chemistry plays an even more important role. Many of the lessons learned in this work should be directly applicable to the study of biomass. For example, model compound studies are very common (for example, lignin is often studied by studying the behavior of its monomeric constituents), and the experimental methods of this work should be applicable to biomass as well. If we had to speculate, the greatest challenge that such work will face is the refinement of kinetic and thermodynamic parameters through quantum chemistry calculations. This will likely involve facing many obstacles from conformers to rotor scans (which are especially important when hydrogen bonding becomes available with the presence of heteroatoms). That being said, while many challenges remain here, these are both items that are actively being research by many individuals who work on automatic mechanism generation, so perhaps pursuing such work will be frutiful on both fronts.

5.1   References

(1)   Khanniche, S.; Lai, L.; Green, W. H. Kinetics of Intramolecular Phenyl Migration and Fused Ring Formation in Hexylbenzene Radicals. *The Journal of Physical Chemistry A* **2018**, *122*, 9778–9791.

(2)   Baruah, B.; Tiwari, P. Compositional and kinetic study of thermal degradation of kerogen using TG-FTIR , NMR , and microscopic study. *AIChE Journal* **2022**, *68*, DOI: 10.1002/aic.17396.

(3) Kim, J. G.; Kim, J. H.; Song, B.-J.; Lee, C. W.; Lee, Y.-S.; Im, J. S. Empirical approach to determine molecular weight distribution using MALDI-TOF analysis of petroleum-based heavy oil. *Fuel* **2016**, *186*, 20–23.

Appendix A

Supporting Information for:

Detailed Reaction Mechanism for 350-400 °C Pyrolysis of an Alkane, Aromatic, and Long-Chain Alkylaromatic Mixture

Section S1: Final Kinetic Mechanism

The final kinetic mechanism is given by `chem_annotated.inp`, which can be simulated with Chemkin-Pro[1] or converted to a Cantera file to be simulated with Cantera.[2] If one wants to simulate the mechanism with RMG, `species_dictionary.txt` will be helpful. All files are publicly available at `https://github.com/kspieks/3_Surrogate_Oil_2021`.

Section S2: GitHub Commit Strings

RMG is an open source software project that is actively updated. This section reports the GitHub version commit hashes that were used to generate the kinetic mechanism for this work. To reproduce the initial model generation, one should use the `input.py` file provided at `https://github.com/kspieks/3_Surrogate_Oil_2021` and check out the respective GitHub commits for RMG-Py and RMG-database. Note that mechanism generation did not converge for the mechanism shown in this work to the tolerance specified in the input file. Instead, RMG was terminated after achieving a sufficient number of core species.

RMG-Py: `062afc533ae6e03c7edae7d5dfa139ef90edb265`

RMG-database: `753e53c80f3c9af7b47770500c1b5e26b87cb600`

To update the thermodynamic and kinetic parameters using those from the most recent version of RMG-database, use the provided `update_thermo_kinetics.py`. The following GitHub version string was used when updating the parameters.

RMG-database: `82c0462b1154cdc7668efccb78576f69cc7a89d2`

Many of the thermodynamic and kinetic parameters for the model in this work were updated to match a newer version of RMG-database. Here, we emphasize the species and reactions with the most significant changes. The biggest improvement to the model's predictions came from updating the thermodynamics of the radical(Benzyl_S_dihydronaphthalene) group in RMG-database which greatly impacted the enthalpy of three 1-octylnaphthalene radicals. These values were originally added by Lai et al.[3] and updated by Liu et al.[4] in RMG-database PR 333. Previously, these radicals had a $\Delta_f H^o$ of about -34 kcal mol^{-1} i.e. it was thermodynamically favorable to create this radical. This had several consequences, such as causing these radical concentrations to be too high when simulating our initial kinetic mechanism. The initial species trajectories were far from the experimentally measured values. Further, the large number of radicals also made the model more stiff and thus slower to integrate. After updating this group, the radicals had a $\Delta_f H^o$ of about 18 kcal mol^{-1}; this more reasonable value fixed the previous issues. $\Delta_f S^o$ was unchanged after updating the parameters. The next biggest change came from updating the thermodynamics of the polycyclic(s2_4_4_ene_1) group in RMG-database. The $\Delta_f H^o$ for this group doubled from about 26 kcal mol^{-1} to 58 kcal mol^{-1}, which similarly made 15 alkyl-bicyclo[2.2.0]hex-2-ene radical species more unstable. We acknowledge Hao-Wei Pang for updating the value of this group in RMG-database PR 443 using the result from a CBS-QB3 calculation. There were an additional 50 species whose $\Delta_f G^o$ only increased by fewer than 3 kcal mol^{-1}. The ΔG_f^o of the remaining species were unchanged.

Updating the kinetic parameters had a much smaller impact on the model's predicted trajectories. For completeness, we identify reactions whose rate constant at 400°C changed by over a factor of 100. If we look at reactions whose species thermodynamics was not estimated using the radical(Benzyl_S_dihydronaphthalene) or polycyclic(s2_4_4_ene_1) group, we identify 53 reactions, all of which are unimolecular reactions in which an alkyl radical from a straight-chain alkylaromatic attacks its own benzene ring to form a fused ring. Breaking aromaticity should not be favorable. Indeed, all reactions had a larger activation

energy after updating the parameters. Some examples include reactions that were calculated by Khanniche et al.[5]

Section S3: Rate-of-Production Analysis

Here, we show representative rate-of-production (ROP) plots that were used to understand the reaction pathways responsible for producing the experimentally measured observables. The relative ordering of reactions from ROP analysis remains essentially constant during the first few hours of the reaction simulation so only one representative time point is shown here. Since the general reaction template is the same for many species entering the respective pathways, only one representative example is shown. For instance, the main pathway for producing straight-chain alkanes is shown in Figure 1.1. The first step is for benzyl to abstract a hydrogen from undecane, creating a pool of various undecyl radicals. These radicals decay to an alkene and a primary alkane radical, which reacts with undecane to form the final alkane product. Note that dodecylbenzene is abbreviated as DDB.

ROP for C7 at 180 min

ROP for 1-HEPTYL at 180 min
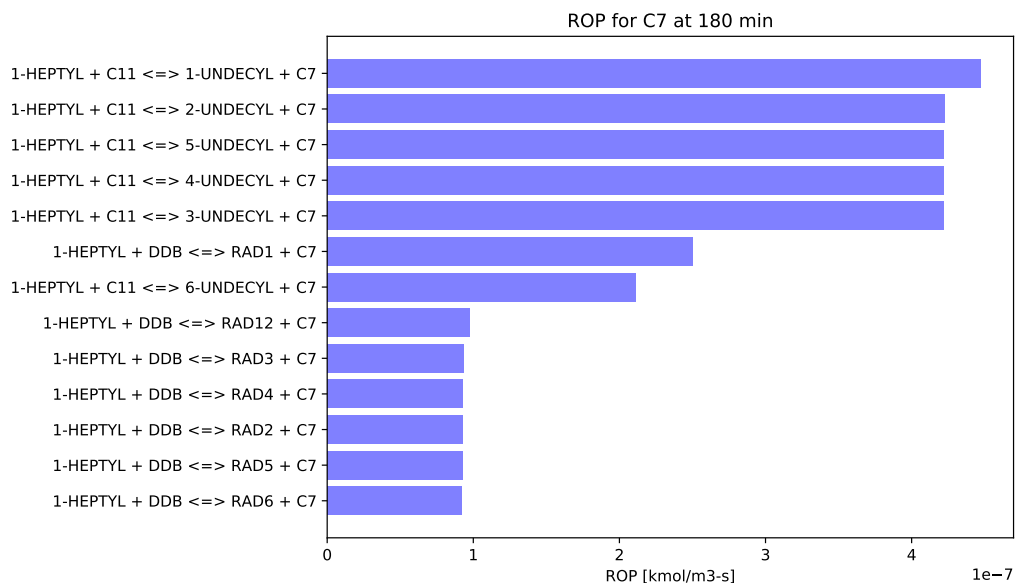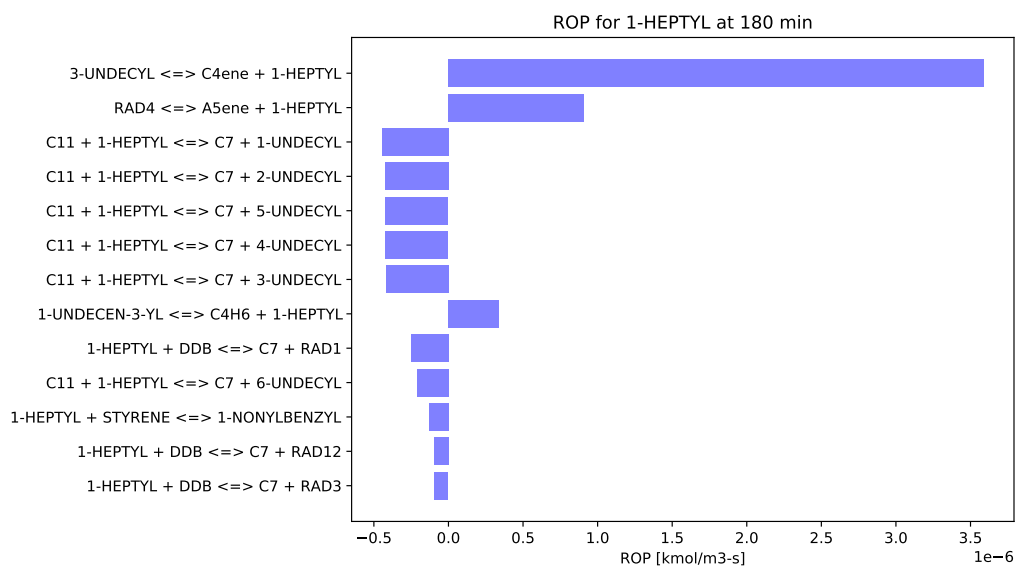
Figure 1.1. Rate-of-production of the highest flux reactions at 3 h for producing (a) heptane and (b) the preceding 1-heptyl radical. Note that only two of the radicals formed by hydrogen abstraction of the feed molecules, namely 3-undecyl and 1-phenyldodecan-4-yl (RAD4), can β-scission to form 1-heptyl. The other radical isomers β-scission to different products following the reaction templates from the main text.

The main pathway for producing straight-chain alkylbenzenes is shown in Figure 1.2. Once a hydrogen is abstracted from the alkyl chain on dodecylbenzene (DDB), that subsequent radical decays to an alkene and a shorter alkylbenzyl radical, which reacts with undecane to produce the alkylaromatic product. Note that pentylbenzene is abbreviated as

BENZ5.
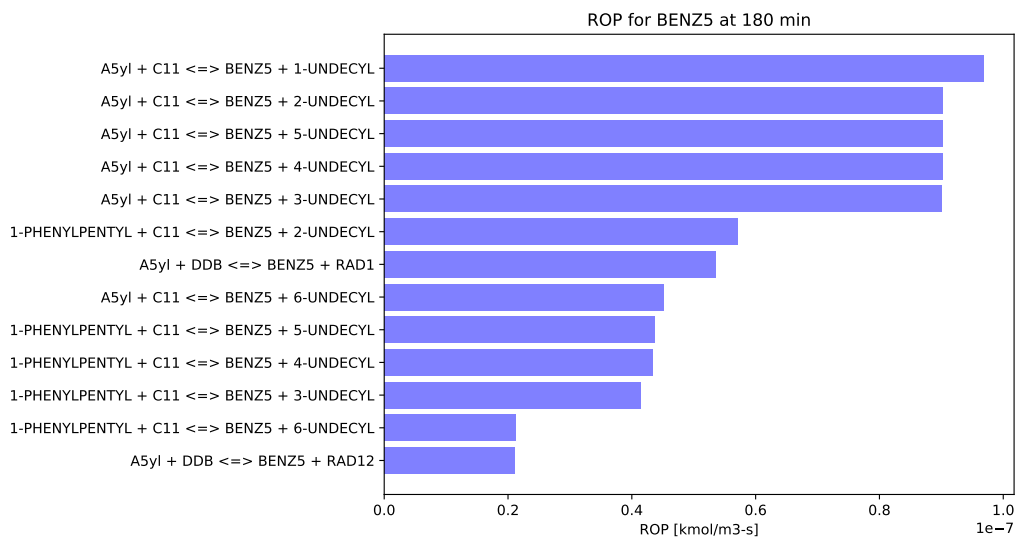
(a)
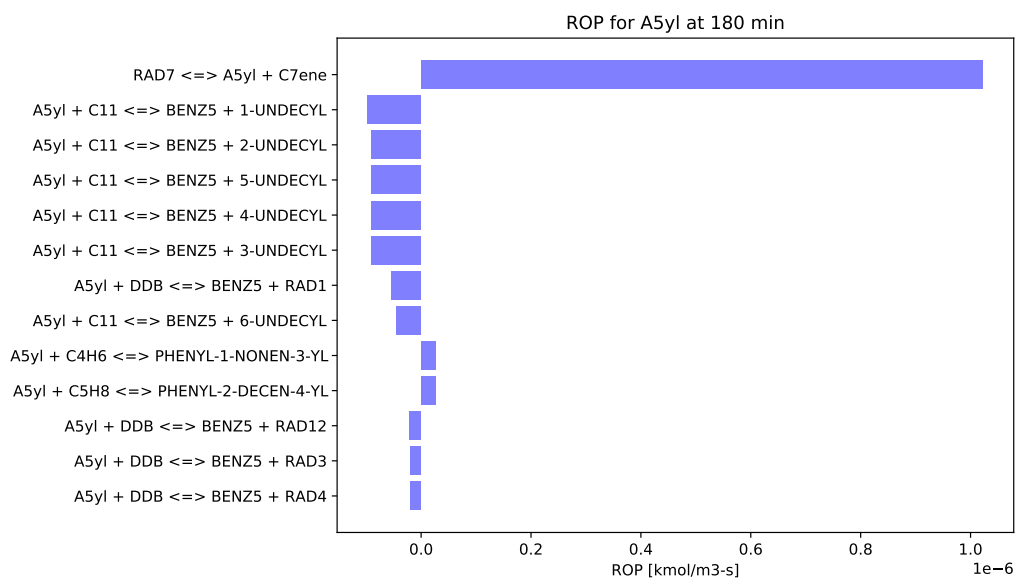
(b)

Figure 1.2. Rate-of-production of the highest flux reactions at 3 h for producing (a) pentylbenzene and (b) the preceding radical. Following the generalized reaction template from the main text, RADX undergoes β-scission to form an alkene along with a primary phenyl alkyl radical, termed "Anyl", such that n = $X - 2$.

Section S4: Model Selectivity

Figures 1.3 to 1.6 show the model's selectivity when simulated at 400 °C and 300 bar with the same initial conditions used in the main text. The model has strong predictive power

130

and agrees well with experimental measurements.
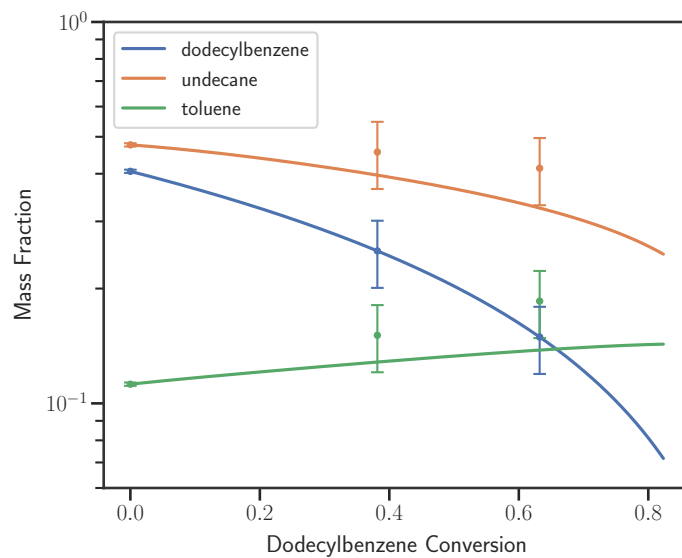
Figure 1.3. Mass fraction of main reactants against dodecylbenzene conversion.
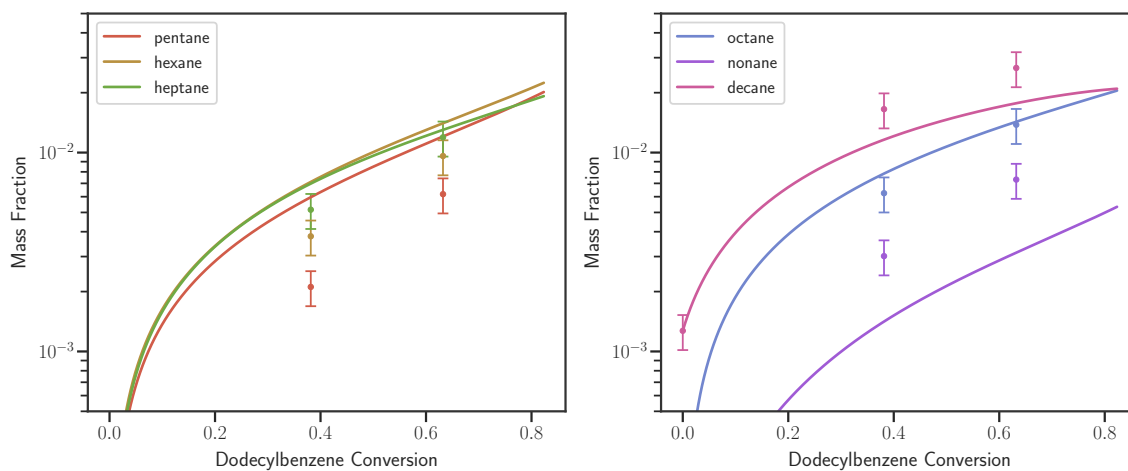
Figure 1.4. Model predictions and experimental measurements for production of alkanes.
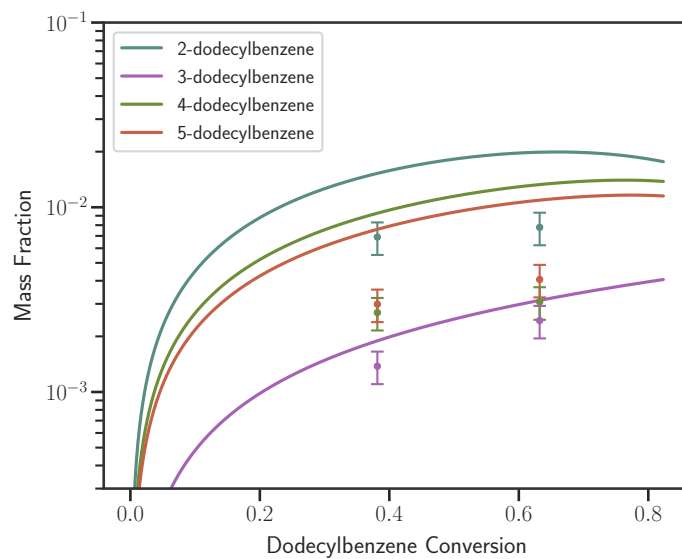
Figure 1.5. Model predictions and experimental measurements for production of dodecylbenzene isomers.
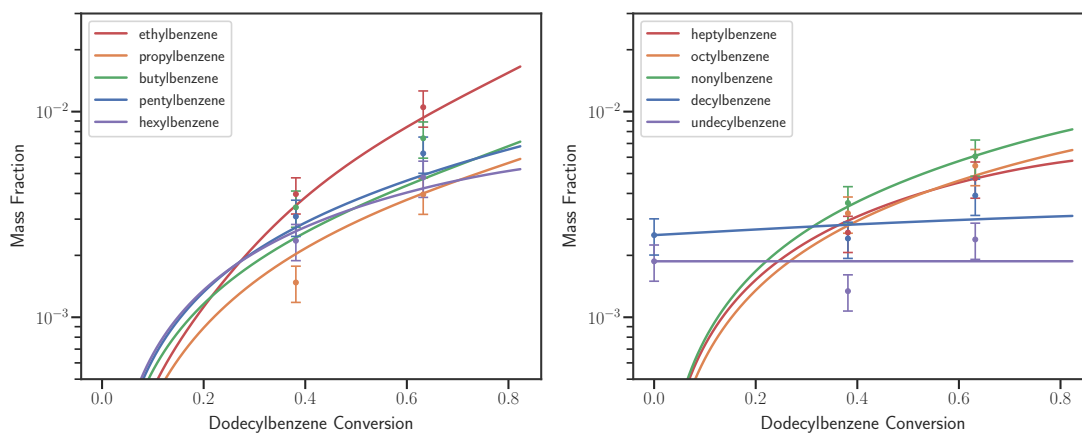
Figure 1.6. Model predictions and experimental measurements for production of alkylaromatic minor products.

Section S5: Experimental Data

Table A.1. Experimental mass fraction for the starting material used for gold tube pyrolysis of the surrogate mixture at 400°C
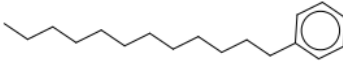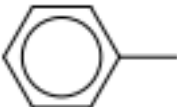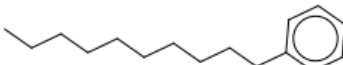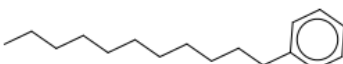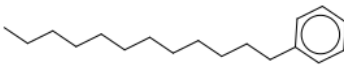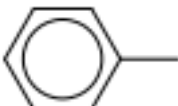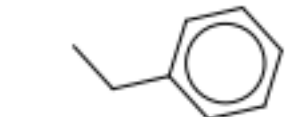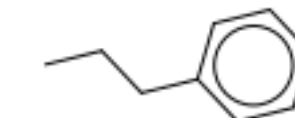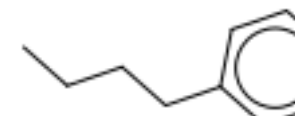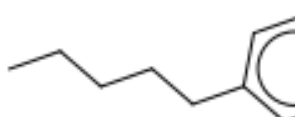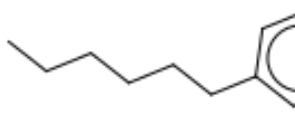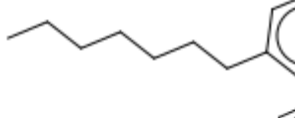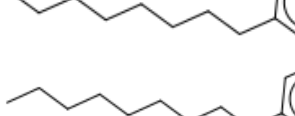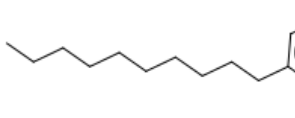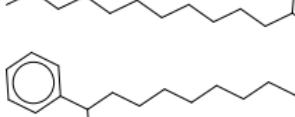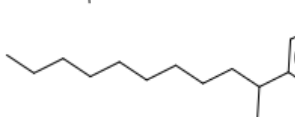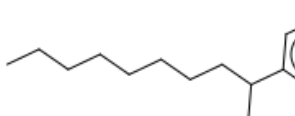
Structure	Label	Mass Fraction [-]
	undecane	4.76e-01
	dodecylbenzene	4.06e-01
	toluene	1.12e-01
	decane	1.27e-03
	decylbenzene	2.51e-03
	undecylbenzene	1.87e-03

Table A.2. Experimental mass fraction for gold tube pyrolysis of the surrogate mixture at 400°C and 3 h

Structure	Label	Trial 1	Trial 2	Avg
	undecane	5.31e-01	3.81e-01	4.56e-01
	dodecylbenzene	2.72e-01	2.30e-01	2.51e-01
	toluene	1.76e-01	1.26e-01	1.51e-01
	pentane	2.25e-03	1.97e-03	2.11e-03
	hexane	4.50e-03	3.09e-03	3.79e-03
	heptane	6.34e-03	3.99e-03	5.16e-03
	octane	7.95e-03	4.54e-03	6.25e-03
	nonane	3.66e-03	2.37e-03	3.02e-03
	decane	2.01e-02	1.30e-02	1.65e-02

ethylbenzene	5.23e-03	2.71e-03	3.97e-03
propylbenzene	1.78e-03	1.17e-03	1.48e-03
butylbenzene	4.20e-03	2.65e-03	3.42e-03
pentylbenzene	3.85e-03	2.33e-03	3.09e-03
hexylbenzene	2.72e-03	1.99e-03	2.36e-03
heptylbenzene	2.96e-03	2.20e-03	2.58e-03
octylbenzene	3.79e-03	2.62e-03	3.20e-03
nonylbenzene	4.16e-03	3.03e-03	3.60e-03
decylbenzene	2.61e-03	2.22e-03	2.42e-03
undecylbenzene	1.49e-03	1.19e-03	1.34e-03
2-dodecylbenzene	8.26e-03	5.53e-03	6.90e-03
3-dodecylbenzene	1.56e-03	1.19e-03	1.38e-03
4-dodecylbenzene	3.14e-03	2.24e-03	2.69e-03
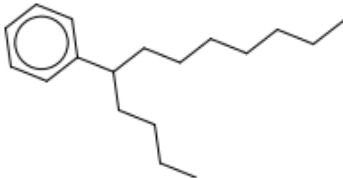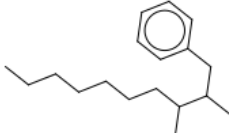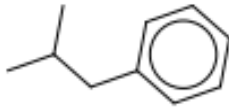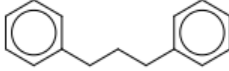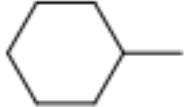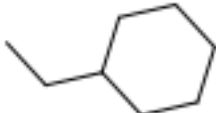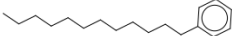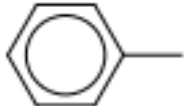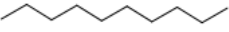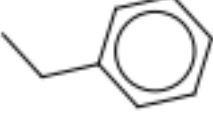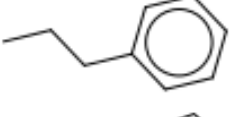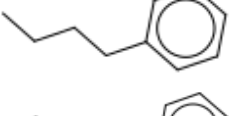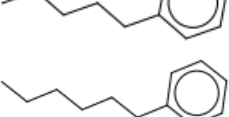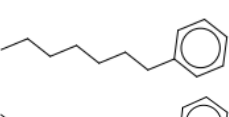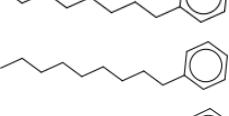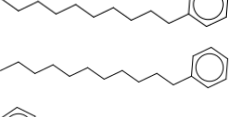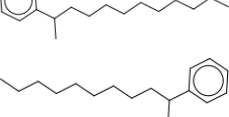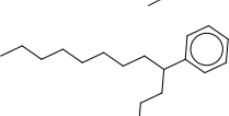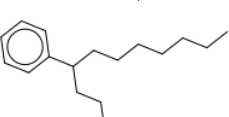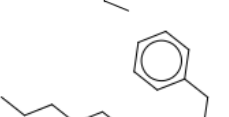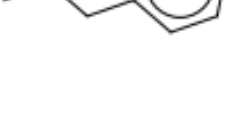
Structure	Label	Trial 1	Trial 2	Trial 3
	5-dodecylbenzene	3.53e-03	2.44e-03	2.99e-03
	2,3-dimethyldecylbenzene	1.07e-03	7.30e-04	8.98e-04
	isobutylbenzene	5.18e-04	3.48e-04	4.33e-04
	diphenylpropane	1.70e-03	1.30e-03	1.50e-03
	naphthalene	5.97e-04	3.77e-04	4.87e-04
	methylcyclohexane	2.71e-04	1.75e-04	2.23e-04
	ethylcyclohexane	3.46e-04	1.19e-04	2.33e-04

Table A.3. Experimental mass fraction for gold tube pyrolysis of the surrogate mixture at 400°C and 6 h

Structure	Label	Trial 1	Trial 2	Trial 3	Avg
	undecane	4.35e-01	3.67e-01	4.38e-01	4.13e-01
	dodecylbenzene	1.46e-01	1.49e-01	1.52e-01	1.49e-01
	toluene	1.96e-01	1.65e-01	1.95e-01	1.85e-01
	pentane	6.46e-03	6.16e-03	5.92e-03	6.18e-03
	hexane	1.02e-02	8.75e-03	9.84e-03	9.61e-03
	heptane	1.26e-02	1.07e-02	1.24e-02	1.19e-02
	octane	1.46e-02	1.25e-02	1.44e-02	1.38e-02

nonane	7.31e-03	7.08e-03	7.56e-03	7.31e-03
decane	2.77e-02	2.44e-02	2.78e-02	2.66e-02
ethylbenzene	1.22e-02	7.92e-03	1.14e-02	1.05e-02
propylbenzene	4.26e-03	3.48e-03	4.12e-03	3.96e-03
butylbenzene	7.99e-03	6.52e-03	7.76e-03	7.42e-03
pentylbenzene	6.24e-03	5.57e-03	7.00e-03	6.27e-03
hexylbenzene	4.77e-03	4.66e-03	4.92e-03	4.78e-03
heptylbenzene	4.62e-03	4.99e-03	4.62e-03	4.74e-03
octylbenzene	5.33e-03	5.61e-03	5.41e-03	5.45e-03
nonylbenzene	6.21e-03	6.13e-03	5.85e-03	6.06e-03
decylbenzene	3.50e-03	4.48e-03	3.75e-03	3.91e-03
undecylbenzene	2.20e-03	2.69e-03	2.28e-03	2.39e-03
2-dodecylbenzene	7.47e-03	7.70e-03	8.19e-03	7.79e-03
3-dodecylbenzene	2.41e-03	2.62e-03	2.28e-03	2.44e-03
4-dodecylbenzene	3.03e-03	3.10e-03	3.09e-03	3.07e-03
5-dodecylbenzene	3.77e-03	4.35e-03	4.07e-03	4.06e-03
2,3-dimethyldecylbenzene	1.38e-03	1.71e-03	1.61e-03	1.57e-03
isobutylbenzene	9.46e-04	9.60e-04	9.69e-04	9.58e-04
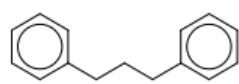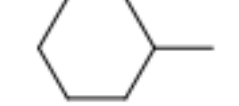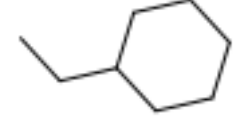
diphenylpropane	1.69e-03	1.88e-03	1.79e-03	1.78e-03
naphthalene	1.26e-03	2.44e-03	1.28e-03	1.66e-03
methylcyclohexane	5.96e-04	5.22e-04	5.30e-04	5.49e-04
ethylcyclohexane	7.51e-04	3.97e-04	6.79e-04	6.09e-04

Table A.4. Experimental mass fraction for the starting material used for gold tube pyrolysis of the surrogate mixture at 350°C
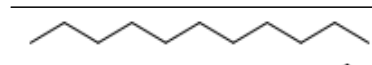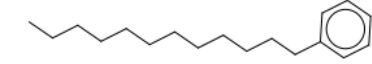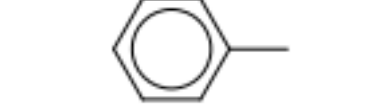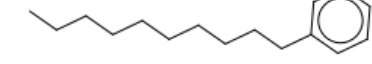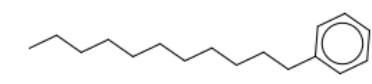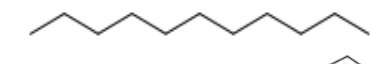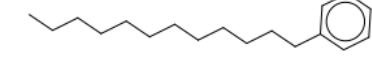
Structure	Label	Mass Fraction [-]
	undecane	4.70e-01
	dodecylbenzene	3.94e-01
	toluene	1.32e-01
	decane	1.26e-03
	decylbenzene	2.43e-03
	undecylbenzene	1.81e-03

Table A.5. Experimental mass fraction for gold tube pyrolysis of the surrogate mixture at 350°C and 116 h

Structure	Label	Mass Fraction [-]
	undecane	4.61e-01
	dodecylbenzene	2.86e-01
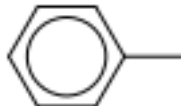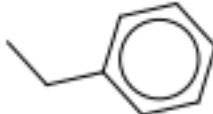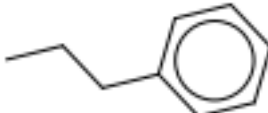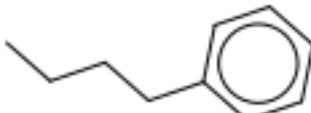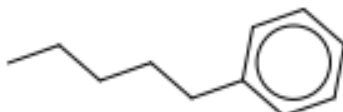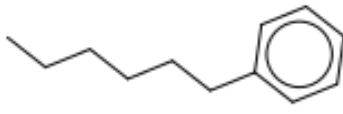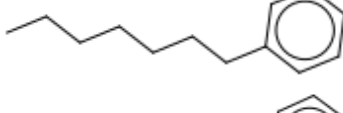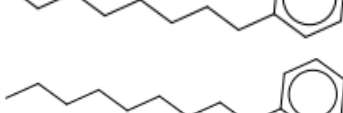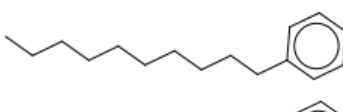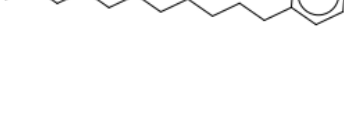
toluene	1.27e-01
pentane	1.27e-03
hexane	2.55e-03
heptane	3.67e-03
octane	4.62e-03
nonane	3.15e-03
decane	1.65e-02
ethylbenzene	3.82e-03
propylbenzene	1.40e-03
butylbenzene	4.13e-03
pentylbenzene	3.91e-03
hexylbenzene	3.29e-03
heptylbenzene	3.94e-03
octylbenzene	5.59e-03
nonylbenzene	6.63e-03
decylbenzene	4.60e-03
undecylbenzene	3.56e-03

Table A.6. Experimental mass fraction for gold tube pyrolysis of the surrogate mixture at 350°C and 239 h

Structure	Label	Mass Fraction [-]
	undecane	3.75e-01
	dodecylbenzene	1.80e-01
	toluene	1.18e-01
	pentane	1.51e-03
	hexane	3.70e-03
	heptane	5.84e-03
	octane	7.76e-03
	nonane	5.52e-03
	decane	2.02e-02
	ethylbenzene	7.73e-03
	propylbenzene	2.71e-03
	butylbenzene	6.03e-03
	pentylbenzene	6.84e-03
	hexylbenzene	5.12e-03
	heptylbenzene	6.19e-03
	octylbenzene	8.92e-03
	nonylbenzene	9.49e-03
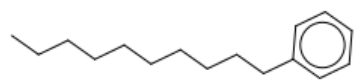
decylbenzene   6.72e-03
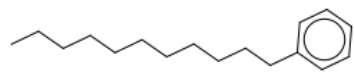
undecylbenzene   4.63e-03

Section S6: Ideal Gas Assumption Error Analysis using the Peng-Robinson Equation of State at 400ºC and 300 bar

There are three primary ways that accounting for real gas behavior will affect kinetic rates: changing the molar volume of the mixture which changes the species concentrations, changing the final equilibrium concentrations, and changing the rate constants themselves due to properly accounting for thermodynamic activity (similar to solvation effects in liquid phase reactions). To see where these effects come from, let's consider the proper thermodynamic and kinetic equations that govern simulating our mechanism. We'll first show the equations of interest in full rigor without any approximations. We'll then discuss which assumptions we made to simplify these equations by assuming ideal gas behavior. Finally, we will estimate the error involved in doing this by comparing to data from a PREOS calculation of the starting material at these conditions.

Equation A.1 shows how to calculate the equilibrium constant, K_a, for a given reaction, which is used to calculate reverse rate constants in RMG. ν_i is the stoichiometric coefficient for species i in the reaction, and $\Delta_f G_i^\circ$ is the standard Gibbs energy of formation of species i. Note that K_a is only a function of temperature, and the standard Gibbs energy of formation sets the reference state for K_a, which in this case is an ideal gas at 1 bar and temperature T. Although the reference state is an ideal gas, the equilibrium constant calculated has the same value even when applied to real gases.

$$\sum_i \nu_i \Delta_f G_i^\circ = -RT\ln(K_a) \tag{A.1}$$

The equilibrium constant for a single reaction is also related to the species mole fractions at equilibrium. Equation A.2 shows how the equilibrium constant is related to thermodynamic activity for a given species, a_i, which in turn is related to the ratio of the fugacity of species i in the real mixture, $\hat{f}_i$ (variables with a hat denotes real mixture properties), to the fugacity of the reference state, f_i°. The fugacity of our reference state, which is an ideal gas at 1 bar, is also equal to 1 bar since the fugacity is equal to the pressure for ideal gasses. The last step in Equation A.2 simply substitutes $\hat{f}_i$ using the definition of the fugacity coefficient in Equation A.3, which allows us to relate K_a to equilibrium mole fractions, x_i^{eq}. Note that the fugacity coefficient is a function of temperature, pressure, and composition.

$$K_a = \prod_i (a_i)^{\nu_i} = \prod_i \left(\frac{\hat{f}_i}{f_i^\circ} \right)^{\nu_i} = \prod_i \left(\frac{x_i^{\text{eq}} \hat{\phi}_i P}{f_i^\circ} \right)^{\nu_i} \tag{A.2}$$

$$\hat{\phi}_i \equiv \frac{\hat{f}_i}{x_i P} \tag{A.3}$$

While so far we have expressed composition in terms of mole fractions, rate equations deal with species concentrations. To rigorously convert between the two, we need to use the compressibility factor, Z, of the real mixture. Equation A.4 shows the definition of the compressibility factor, and Equation A.5 shows how this is used to convert concentrations to mole fractions for an arbitrary species "A".

$$Z \equiv \frac{PV}{NRT} \tag{A.4}$$

$$[A] = \frac{n_A}{V} = \frac{n_A P}{ZNRT} = x_A \frac{P}{ZRT} \tag{A.5}$$

With this, we can now consider what goes in to the rates for a given reaction. Transition state theory (TST) tell us that the forward rate of a reaction is related to the concentration of the transition state, as shown in Equation A.6, which includes a tunneling correction κ. TST also posits that a pseudo-equilibrium exists between the transition state and the reactants, with an equilibrium constant $K_a^\ddagger$. Equation A.7 shows $K_a^\ddagger$ for a bimolecular reaction (with reactants A and B), which is related to the thermodynamic activities of the reactants and the transition state following the same approach as Equation A.2.

$$\text{rate} = \kappa \frac{k_B T}{h} [AB^\ddagger] = \kappa \frac{k_B T}{h} x^\ddagger \frac{P}{ZRT} \tag{A.6}$$

$$K_a^\ddagger = \frac{a^\ddagger}{a_A a_B} = \frac{\hat{\phi}^\ddagger}{\hat{\phi}_A \hat{\phi}_B} \frac{x^\ddagger}{x_A x_B} \frac{f^\circ}{P} \tag{A.7}$$

Rearranging Equation A.7 to solve for the mole fraction of the transition state, $x^\ddagger$, and substituting this into Equation A.6 results in the forward rate of a reaction in terms of useful quantities. Equation A.8 shows the forward rate for a bimolecular reaction in full rigor according to TST, while Equation A.9 shows a similar result but for a unimolecular reaction with reactant C. Note that the units work out to those of concentration per time as

expected ($K_a^\ddagger$, $\hat{\phi}_i$, κ, and Z are all unitless, while f° has units of bar). Also, strictly speaking the left hand side of these equations would be $\frac{d[A]}{dt}$ only in the case of a batch reactor, and we have only included the forward rate of one reaction, but the point we are making here will hold regardless.

$$\frac{d[A]}{dt} = -\kappa \frac{k_B T}{h} K_a^\ddagger \frac{1}{f^\circ} \frac{\hat{\phi}_A \hat{\phi}_B}{\hat{\phi}^\ddagger} x_A x_B \frac{P^2}{ZRT} \tag{A.8}$$

$$\frac{d[C]}{dt} = -\kappa \frac{k_B T}{h} K_a^\ddagger \frac{\hat{\phi}_C}{\hat{\phi}^\ddagger} x_C \frac{P}{ZRT} \tag{A.9}$$

Now we can consider how our assumption of ideal gas behavior affects our calculations. First, ideal gas behavior assumes that $Z = 1$. By assuming this, we are potentially neglecting the effect that Z has on our rate laws as shown in Equations A.8 and A.9. This effect is essentially the fact that a real gas has a different density than an ideal gas, which affects the concentration of the transition state. However, if Z is relatively constant over the course of the reaction, we can pull a factor of $1/Z$ from the left hand side of these rate equations as shown in Equation A.10. This leads to the interesting result in Equation A.11 that suggests that when plotting concentration profiles in terms of mole or mass fractions as we did in this work that Z does not directly affect the rates at all, other than how Z contributes to $\hat{\phi}_i$.

$$\frac{d[A]}{dt} = \frac{d(x_A P/(ZRT))}{dt} = \frac{P}{ZRT}\frac{dx_A}{dt} = -\kappa \frac{k_B T}{h} K_a^\ddagger \frac{1}{f^\circ} \frac{\hat{\phi}_A \hat{\phi}_B}{\hat{\phi}^\ddagger} x_A x_B \frac{P^2}{ZRT} \tag{A.10}$$

$$\frac{dx_A}{dt} = -\kappa \frac{k_B T}{h} K_a^\ddagger \frac{P}{f^\circ} \frac{\hat{\phi}_A \hat{\phi}_B}{\hat{\phi}^\ddagger} x_A x_B \tag{A.11}$$

Second, ideal gas behavior assumes that for all species $\hat{\phi}_i = 1$. This assumption has two consequences. First, as shown in Equations A.8 and A.9, neglecting the fugacity coefficients will also affect our forward rate expressions. This is essentially the same effect as solvation effects in solution phase. It should be noted, though, that for unimolecular reactions that the transition state usually has a similar fugacity coefficient to the reactant, so $\frac{\hat{\phi}^\ddagger}{\hat{\phi}_C} \approx 1$ even if $\hat{\phi}_i \neq 1$. This means that this effect is usually negligible for unimolecular reactions. The second effect of assuming $\hat{\phi}_i = 1$ is that this changes the final equilibrium composition. For example, for a single reaction, Equation A.2 shows that if we include the $\hat{\phi}_i$ terms

Species	Mole Fraction	$\hat{\phi}_i$	ϕ_i
toluene	0.206	0.370	0.362
undecane	0.515	0.324	0.321
dodecylbenzene	0.279	0.0688	0.0676
Z = 1.63			

Table A.7. PREOS calculation of the starting material at 400°C and 300 bar

that x_i^{eq} must change so that their product will still equal the equilibrium constant, which does not change for real gases from ideal gases. For our complex simulation containing 116 045 reactions, calculating the equilibrium composition is not this straight forward, but the general idea holds.

Finally, now that we have discussed the effects of our ideal gas assumption, let's estimate the magnitude of these effects for this system. While it is not possible to calculate the full mixture of 1326 species using PREOS, we can do this for the starting material, where all species are known. The compressibility factor we calculate should be representative of the real mixture for the other time points, as the reactants are always the highest concentration species. The fugacity coefficients should also give us good expectations for the remaining species. Table A.7 shows the results of the PREOS calculation (using its standard mixing rules) for the starting material at 400°C and 300 bar.

The results in Table A.7 confirm that the real mixture does in fact deviate from ideal gas behavior quite a bit. The compressibility factor of 1.63 means that the mixture is more expanded than an ideal gas would otherwise be at these conditions. The values for $\hat{\phi}_i$ are all less than 1, and get smaller as the molecules increase in size. We also calculate the pure species fugacity coefficients, ϕ_i (i.e. $\hat{\phi}_i$ if x_i=1), and they are extremely close to the mixture properties for each species. This means the mixture is behaving as an ideal solution. All of this suggests that the main sources of the non-ideal behavior is that the volume of the molecules themselves cannot be neglected under these conditions. This suggests that the main factor affecting the fugacity coefficient is the size of the molecule, and that differences in intermolecular forces plays a less important role under these conditions.

Putting it all together, the fact that $Z = 1.63$ means that the ideal rates in molar units are a factor of 1.63 faster than they should be just due to the change in density, but that this effect will cancel out for our plots which deal with mole/mass fraction units. As for the remaining effects caused by including $\hat{\phi}_i$, because the values depend mostly on the size of the

molecule, and mass is conserved across reactants, products, and the transition state, ratios of $\hat{\phi}_i$ for products/reactants or transition state/reactants are unlikely to deviate too far from 1. For example, a small fugacity coefficient in the numerator due to a large species is likely being somewhat canceled out by a small fugacity coefficient in the denominator by a large species (to make the mass balance). Bimolecular reactions are probably the most affected by this, as their rate depends on $\frac{\hat{\phi}_A \hat{\phi}_B}{\hat{\phi}^\ddagger}$, which has more coefficients in the numerator. But again, the transition state must have the largest mass, so it will have the smallest fugacity coefficient to somewhat counteract this.

To get an idea of this effect, we also included the pure species fugacity coefficients for select species at these conditions in Table A.8

species	Z	ϕ
methane	1.08	1.06
ethane	1.03	0.97
propane	1.02	0.88
n-butane	1.02	0.77
n-pentane	1.05	0.67
toluene	0.871	0.36
propylbenzene	1.05	0.27
undecane	1.57	0.32

Table A.8. PREOS Analysis

Section S7: Dodecylbenzene Pyrolysis

Briefly, we were interested in comparing our model against the prior literature for dodecylbenzene pyrolysis. Since our model parameters are not fit to any experimental data nor was our mechanism built exclusively for dodecylbenzene pyrolysis, we do not expect the predicted trajectories to go exactly through the data points from previously published work. Instead, we are interested in seeing if our model predictions typically predict the correct trends as well as the order of product concentrations. When simulating our model at 300 bar and 400 °C when dodecylbenzene is the only starting material, the species with largest mole fraction at 9 h include toluene, ethylbenzene, pentane, pentene, hexane, decane, decene, undecene, and

145

dodecylbenzene isomers. Qualitatively, this list of major products agrees with those from published literature, indicating that our model may be useful for dodecylbenzene pyrolysis as well. However, quantitative comparison to our work is difficult since comparing the experimental data from the published works reveals some inconsistencies. For example, Behar et al.[6] mainly reports conversion of dodecylbenzene, toluene, and ethylbenzene. References to other products are bucketed into a group containing seven to fourteen carbons and a group containing over fourteen carbons. Burkle-Vitzthum et al.[7] reports ethylbenzene, octane and nonane and, unlike Behar et al., does not show results for other alkane products, such as the ones discussed in the main body of this work. Their proposed mechanism of homolysis of dodecylbenzene has also been disproven in more recent publications. Neither work shows results for isomers of dodecylbenzene. Savage et al.[8] also does not report any species with fewer than six carbons. Still, their work contains one of the most comprehensive tables, so we plot their experimental data along with our model predictions in Figures 1.7, 1.8, and 1.9. Our model was simulated at 400 °C and constant volume, which matches the experimental conditions from Savage et al. Data points are taken from Table 1 in Savage et al. Their work does not provide experimental error bars. Our model predicts the trends well and is often within a factor of two when predicting the products.
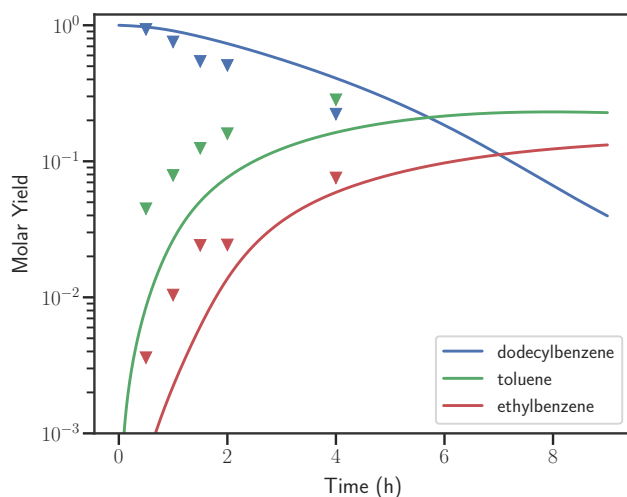
Figure 1.7. Main product formation when pyrolyzing dodecylbenzene. Data points are from Savage et al.[8]
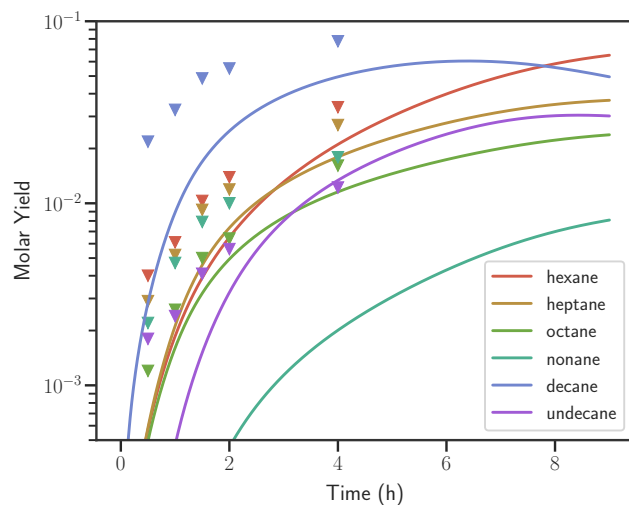
Figure 1.8. Alkane product formation when pyrolyzing dodecylbenzene. Data points are from Savage et al.[8]
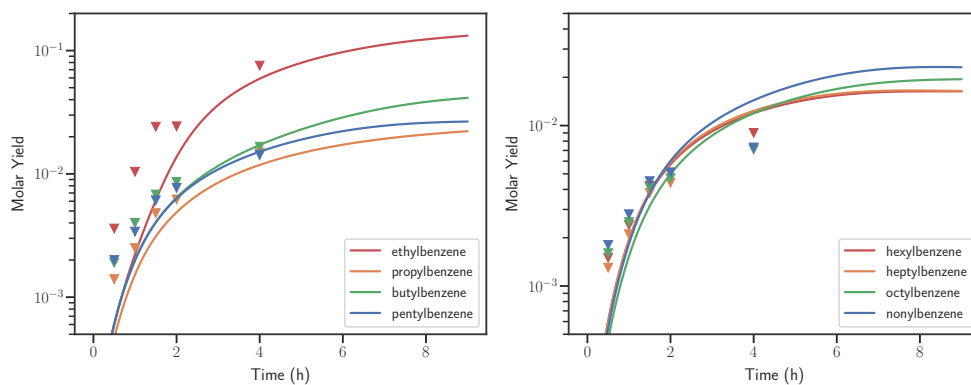
Figure 1.9. Alkylaromatic product formation when pyrolyzing dodecylbenzene. Data points are from Savage et al.[8]

A.1 References

(1) CHEMKIN-PRO, R. 15112, Reaction Design. *Inc., San Diego, CA* **2011**.

(2) Goodwin, D. G.; Speth, R. L.; Moffat, H. K.; Weber, B. W. Cantera: An Object-oriented Software Toolkit for Chemical Kinetics, Thermodynamics, and Transport Processes, https://www.cantera.org, Version 2.4.0, 2018.

(3) Lai, L.; Khanniche, S.; Green, W. H. Thermochemistry and group additivity values for fused two-ring species and radicals. *The Journal of Physical Chemistry A* **2019**, *123*, 3418–3428.

(4) Liu, M.; Chu, T.-C.; Jocher, A.; Smith, M. C.; Lengyel, I.; Green, W. H. Predicting polycyclic aromatic hydrocarbon formation with an automatically generated mechanism for acetylene pyrolysis. *Int. J. Chem. Kinet.* **2021**, *53*, 27–42.

(5) Khanniche, S.; Lai, L.; Green, W. H. Kinetics of Intramolecular Phenyl Migration and Fused Ring Formation in Hexylbenzene Radicals. *The Journal of Physical Chemistry A* **2018**, *122*, 9778–9791.

(6) Behar, F.; Lorant, F.; Budzinski, H.; Desavis, E. Thermal stability of alkylaromatics in natural systems: kinetics of thermal decomposition of dodecylbenzene. *Energy & fuels* **2002**, *16*, 831–841.

(7) Burklé-Vitzthum, V.; Michels, R.; Scacchi, G.; Marquaire, P.-M. Mechanistic modeling of the thermal cracking of decylbenzene. Application to the prediction of its thermal stability at geological temperatures. *Industrial & engineering chemistry research* **2003**, *42*, 5791–5808.

(8) Savage, P. E.; Klein, M. T. Discrimination between molecular and free-radical models of 1-phenyldodecane pyrolysis. *Industrial & engineering chemistry research* **1987**, *26*, 374–376.

Appendix B

Supporting Information for: Implementation and Comparison of Bond Additivity Corrections and Isodesmic Reactions for Thermochemistry Calculations

Section S1: Reference Species

Index	Name	Mult.	Charge	$\Delta_f H^\circ_{298}$	Uncertainty	Source
0	Dihydrogen	1	0	0.0	0.0	ATcT
1	Methylidyne	2	0	596.15	0.11	ATcT
2	Hydroxyl	2	0	37.48	0.03	ATcT
3	Hydroxide	1	-1	-139.07	0.03	ATcT
4	Imidogen	3	0	358.79	0.17	ATcT
5	Sulfanyl	2	0	141.87	0.52	3rd Mil.
6	Hydrogen fluoride	1	0	-272.73	0.05	ATcT
7	Hydrogen chloride	1	0	-92.17	0.01	ATcT
8	Hydrogen bromide	1	0	-35.57	0.16	ATcT
9	Methylene	3	0	391.59	0.12	ATcT
10	Water	1	0	-241.84	0.03	ATcT
11	Amidogen	2	0	186.02	0.12	ATcT
12	Azanide	1	-1	112.02	0.33	ATcT

13	Hydrogen sulfide	1	0	-20.6	0.5	CCCBDB
14	Methyl	2	0	146.41	0.08	ATcT
15	Methylium	1	1	1095.29	0.08	ATcT
16	Methanide	1	-1	137.63	0.3	ATcT
17	Oxonium	1	1	599.09	0.17	ATcT
18	Ammonia	1	0	-45.56	0.03	ATcT
19	Methane	1	0	-74.53	0.06	ATcT
20	Ammonium	1	1	631.73	0.21	ATcT
21	Carbon monoxide	1	0	-110.52	0.03	ATcT
22	Dioxygen	3	0	0.0	0.0	ATcT
23	Nitrilomethyl	2	0	439.99	0.15	ATcT
24	Cyanide	1	-1	67.25	0.1	ATcT
25	Nitric oxide	2	0	91.12	0.07	ATcT
26	Dinitrogen	1	0	0.0	0.0	ATcT
28	Sulfur monoxide	3	0	5.02	0.5	Benson
29	Disulfur	3	0	128.6	0.3	NIST
30	Fluoromethylidyne	2	0	246.74	0.13	ATcT
31	Fluorooxidanyl	2	0	110.89	0.15	ATcT
32	Hypofluorite	1	-1	-108.68	0.48	ATcT
33	Difluorine	1	0	0.0	0.0	ATcT
34	Chlorine monoxide	2	0	101.72	0.04	ATcT
35	Chlorine fluoride	1	0	-55.72	0.06	ATcT
36	Dichlorine	1	0	0.0	0.0	ATcT
37	Bromine monoxide	2	0	123.61	0.28	ATcT
38	Hypobromite	1	-1	-103.63	0.54	ATcT
39	Bromine monofluoride	1	0	-58.85	1.0	CCCBDB
40	Bromine monochloride	1	0	14.44	0.06	ATcT
41	Dibromine	1	0	30.89	0.12	ATcT
42	Oxomethylium	1	1	827.22	0.1	ATcT
43	Formyl	2	0	41.8	0.1	ATcT
44	Formyl anion	1	-1	12.02	0.44	ATcT
45	Dioxidanyl	2	0	12.21	0.15	ATcT

46	Dioxidanide	1	-1	-91.53	0.35	ATcT
47	Hydrogen cyanide	1	0	129.28	0.09	ATcT
48	Hydrogen isocyanide	1	0	192.4	0.38	ATcT
49	Nitrosyl hydride	1	0	106.96	0.11	ATcT
50	Diazynium	1	1	1038.93	0.68	ATcT
51	Diazenyl	2	0	249.25	0.47	ATcT
52	Fluoromethylene	1	0	148.62	0.46	ATcT
53	Hypoflorous acid	1	0	-87.3	0.19	ATcT
54	Chloromethylene	1	0	320.52	0.89	ATcT
55	Hypochlorous acid	1	0	-76.81	0.03	ATcT
56	Hypobromous acid	1	0	-62.18	0.59	ATcT
57	Acetylene	1	0	228.27	0.13	ATcT
58	Vinylidene	1	0	412.17	0.33	ATcT
59	Formaldehyde	1	0	-109.19	0.1	ATcT
60	Hydroxymethylene	1	0	108.93	0.28	ATcT
61	Formaldehyde cation	2	1	941.27	0.1	ATcT
62	Hydrogen peroxide	1	0	-135.48	0.06	ATcT
63	Methyleneamidogen	2	0	238.3	0.67	ATcT
64	Iminomethyl	2	0	272.22	0.71	ATcT
65	Methyleneamidogen anion	1	-1	188.97	0.96	ATcT
66	Nitroxyl	2	0	64.52	0.9	ATcT
67	(E)-Diazene	1	0	199.95	0.44	ATcT
68	Chloromethyl	2	0	115.39	0.87	ATcT
69	Vinyl	2	0	296.94	0.34	ATcT
70	Vinyl anion	1	-1	232.59	0.87	ATcT
71	Ethylidyne	2	0	504.86	0.87	ATcT
72	Methoxy	2	0	21.62	0.29	ATcT
73	Hydroxymethyl	2	0	-16.8	0.28	ATcT
74	Methoxide	1	-1	-130.21	0.31	ATcT
75	Hydroxymethylium	1	1	709.8	0.18	ATcT
76	Methanimine	1	0	88.34	0.62	ATcT
77	Hydroxylamine	1	0	-43.62	0.5	ATcT

78	Hydrazino	2	0	224.28	0.78	ATcT
80	Methyl fluoride	1	0	-235.47	0.24	ATcT
81	Chloromethane	1	0	-82.1	0.25	ATcT
82	Methyl bromide	1	0	-35.58	0.27	ATcT
83	Ethylene	1	0	52.39	0.12	ATcT
84	Ethylidene	3	0	354.3	0.91	ATcT
85	Methanol	1	0	-200.84	0.15	ATcT
86	Aminomethyl	2	0	149.06	0.44	ATcT
87	Aminomethylium	1	1	751.58	0.66	ATcT
88	Methylamidogen	2	0	176.56	0.49	ATcT
89	Methylamidogen anion	1	-1	134.12	0.68	ATcT
90	Hydrazine	1	0	97.41	0.49	ATcT
91	Methanethiol	1	0	-22.84	0.59	CCCBDB
92	Ethyl	2	0	119.87	0.28	ATcT
93	Methylamine	1	0	-21.72	0.31	ATcT
94	Ethane	1	0	-83.96	0.13	ATcT
95	Carbon dioxide	1	0	-393.48	0.01	ATcT
96	Ozone	1	0	141.73	0.04	ATcT
97	Cyanato	2	0	127.5	0.35	ATcT
98	Nitrogen dioxide	2	0	34.05	0.07	ATcT
99	Nitrite	1	-1	-185.44	0.46	ATcT
100	Nitronium	1	1	958.22	0.2	ATcT
101	Nitrous oxide	1	0	82.56	0.1	ATcT
102	Carbonoxidesulfide	1	0	-142.0	0.3	CATCH
103	Sulfur dioxide	1	0	-296.81	0.2	CCCBDB
104	Carbon disulfide	1	0	116.6	0.8	CATCH
105	Disulfur monoxide	1	0	-56.04	1.4	CCCBDB
106	Fluoroformyl	2	0	-176.03	0.39	ATcT
107	Fluorodioxidanyl	2	0	25.09	0.26	ATcT
108	Cyanic fluoride	1	0	9.11	0.71	ATcT
109	Difluoromethylene	1	0	-193.47	0.39	ATcT
110	Oxygen difluoride	1	0	24.55	0.24	ATcT

111	Chloroformyl	2	0	-20.52	0.49	ATcT
112	Chlorodioxidanyl	2	0	102.82	0.37	ATcT
113	Cyanogen chloride	1	0	135.73	0.47	ATcT
114	Nitrosyl chloride	1	0	52.55	0.07	ATcT
115	Dichloromethylene	1	0	230.67	0.65	ATcT
116	Chloro hypochlorite	1	0	77.97	0.36	ATcT
117	2-Propynylidyne	2	0	722.29	0.81	ATcT
118	Oxoethenyl	2	0	178.09	0.61	ATcT
119	Ethynol anion	1	-1	-47.26	0.89	ATcT
120	Hydroxyformyl	2	0	-184.16	0.5	ATcT
121	Formate	1	-1	-466.55	0.56	ATcT
122	Hydroxyoxomethylium	1	1	597.6	0.44	ATcT
123	Formyloxidanyl	2	0	-127.23	0.56	ATcT
124	Trioxidanyl	2	0	23.32	0.13	ATcT
125	Isocyanic acid	1	0	-118.84	0.31	ATcT
126	Isofulminic acid	1	0	169.46	0.49	ATcT
127	Cyanic acid	1	0	-14.91	0.49	ATcT
128	Fulminic acid	1	0	233.68	0.48	ATcT
129	Nitrous acid	1	0	-79.17	0.08	ATcT
130	Hydrazoic acid	1	0	291.61	0.49	ATcT
131	Fluoroacetylene	1	0	105.6	0.41	ATcT
132	Formyl fluoride	1	0	-382.23	0.33	ATcT
133	Chloroacetylene	1	0	228.9	0.77	ATcT
134	Formyl chloride	1	0	-183.18	0.77	ATcT
136	Peroxyhypochlorous acid	1	0	-1.34	0.95	ATcT
137	Dichloromethyl	2	0	90.11	0.88	ATcT
138	Propadienylidene	1	0	555.57	0.41	ATcT
139	Cyclopropenylidene	1	0	496.11	0.46	ATcT
140	Propynylidene	3	0	546.38	0.63	ATcT
141	Ketene	1	0	-48.49	0.13	ATcT
142	Dioxymethyl	1	0	104.9	0.63	ATcT
143	Formic acid	1	0	-378.36	0.22	ATcT

144	Dioxirane	1	0	1.78	0.55	ATcT
145	Trioxidane	1	0	-90.55	0.72	ATcT
146	Methylene fluoride	1	0	-450.58	0.36	ATcT
147	Dichloromethane	1	0	-94.41	0.49	ATcT
149	2-Propynyl	2	0	351.43	0.38	ATcT
150	1-Propynyl	2	0	526.03	0.75	ATcT
151	1,2-Propadienyl anion	1	-1	262.44	0.73	ATcT
152	1-Propynyl anion	1	-1	260.98	0.76	ATcT
153	Cycloprop-1-enyl	2	0	522.98	0.69	ATcT
154	Cycloprop-2-enyl	2	0	485.66	0.64	ATcT
155	Acetyl	2	0	-10.03	0.36	ATcT
156	Vinoxy	2	0	15.58	0.77	ATcT
157	Vinoxide	1	-1	-160.63	0.77	ATcT
158	Methylperoxy	2	0	12.66	0.5	ATcT
159	Methylperoxy anion	1	-1	-99.96	0.68	ATcT
160	Vinyl fluoride	1	0	-142.43	0.48	ATcT
161	Vinyl chloride	1	0	21.86	0.31	ATcT
162	Vinyl bromide	1	0	73.97	0.61	ATcT
163	Propyne	1	0	185.76	0.25	ATcT
164	Allene	1	0	189.89	0.26	ATcT
165	Cyclopropene	1	0	283.61	0.55	ATcT
166	Acetaldehyde	1	0	-165.44	0.28	ATcT
167	Oxirane	1	0	-52.55	0.38	ATcT
168	Ethenol	1	0	-123.95	0.84	ATcT
169	Methyl hydroperoxide	1	0	-127.85	0.76	ATcT
170	Thiirane	1	0	82.0	1.3	CATCH
171	Allyl	2	0	167.82	0.55	ATcT
172	1-Methylethenyl	2	0	252.46	0.78	ATcT
173	1-Propenyl	2	0	267.2	0.79	ATcT
174	Ethoxide	1	-1	-178.29	0.53	ATcT
175	1-Hydroxyethyl	2	0	-55.43	0.65	ATcT
176	2-Hydroxyethyl	2	0	-26.11	0.63	ATcT

177	Ethoxy	2	0	-12.07	0.52	ATcT
178	Fluoroethane	1	0	-272.07	0.38	ATcT
179	Chloroethane	1	0	-111.38	0.2	ATcT
180	Ethyl bromide	1	0	-63.07	0.26	ATcT
181	Propene	1	0	19.98	0.21	ATcT
182	Cyclopropane	1	0	53.63	0.47	ATcT
183	Ethanol	1	0	-234.64	0.21	ATcT
184	Dimethyl ether	1	0	-184.04	0.45	ATcT
185	(Methylamino)methyl	2	0	151.1	0.97	ATcT
186	Ethanethiol	1	0	-46.3	0.6	CATCH
187	Dimethyl sulfide	1	0	-37.5	0.6	CATCH
188	iso-Propyl	2	0	88.18	0.56	ATcT
189	Propyl	2	0	100.87	0.6	ATcT
190	Dimethylamine	1	0	-18.11	0.48	ATcT
191	Propane	1	0	-105.03	0.19	ATcT
192	Nitrate	1	-1	-306.77	0.62	ATcT
193	Nitrooxidanyl	2	0	74.13	0.19	ATcT
194	Cyanogen	1	0	310.18	0.43	ATcT
195	Dinitrogen dioxide	1	0	171.12	0.14	ATcT
196	Sulfur trioxide	1	0	-395.9	0.7	CCCBDB
197	1,2-Difluoroacetylene	1	0	5.59	0.66	ATcT
198	Difluorophosgene	1	0	-606.5	0.48	ATcT
199	Difluorodioxidane	1	0	31.57	0.41	ATcT
200	Trifluoromethyl	2	0	-467.82	0.49	ATcT
202	Nitryl chloride	1	0	12.5	1.0	CCCBDB
203	Dichloroacetylene	1	0	233.39	0.98	ATcT
204	Phosgene	1	0	-219.13	0.27	ATcT
205	Chlorooxy hypochlorite	1	0	131.32	0.56	ATcT
206	Trichloromethyl	2	0	71.43	0.86	ATcT
207	Dibromophosgene	1	0	-113.89	0.37	ATcT
208	Nitric acid	1	0	-134.2	0.18	ATcT
209	Peroxynitrous acid	1	0	-14.54	0.36	ATcT

210	Fluoroform	1	0	-695.74	0.44	ATcT
211	Chloroform	1	0	-102.34	0.52	ATcT
212	Bromodifluoromethane	1	0	-424.05	0.49	ATcT
213	1,3-Butadiyne	1	0	460.07	0.82	ATcT
214	Glyoxal	1	0	-212.55	0.52	ATcT
215	Carbonic acid	1	0	-612.91	0.79	ATcT
216	1,1-Difluoroethene	1	0	-350.66	0.94	ATcT
217	1,1-Dichloroethene	1	0	2.92	0.51	ATcT
218	cis-1,2-Dichloroethene	1	0	-2.54	0.56	ATcT
219	Nitromethane	1	0	-74.75	0.46	ATcT
220	Methyl nitrite	1	0	-67.36	0.45	ATcT
221	Acetyl chloride	1	0	-241.55	0.33	ATcT
222	Methyl formate	1	0	-357.78	0.59	ATcT
223	Acetic acid	1	0	-432.63	0.52	ATcT
224	Thiourea	1	0	22.9	1.6	CCCBDB
225	1,1-Difluoroethane	1	0	-502.78	0.65	ATcT
226	1,2-Difluoroethane	1	0	-447.87	0.86	ATcT
227	1,1-Dichloroethane	1	0	-132.54	0.51	ATcT
228	1,2-Dichloroethane	1	0	-130.43	0.54	ATcT
229	Acetonyl	2	0	-31.88	0.99	ATcT
230	Acetaldoxime	1	0	-22.55	0.29	CCCBDB
231	Thioacetamide	1	0	12.7	1.2	CATCH
232	2-Chloroethanol	1	0	-266.76	0.63	ATcT
233	2-Bromoethanol	1	0	-220.63	0.57	ATcT
234	Bicyclo[1.1.0]butane	1	0	217.15	0.84	CCCBDB
235	1,2-Butadiene	1	0	162.3	0.6	CCCBDB
236	1,3-Butadiene	1	0	110.57	0.4	ATcT
237	2-Butyne	1	0	145.99	0.59	ATcT
238	1-Butyne	1	0	165.77	0.67	ATcT
239	Cyclobutene	1	0	159.74	0.92	ATcT
240	Propylene oxide	1	0	-93.72	0.63	CCCBDB
241	Oxetane	1	0	-80.5	0.6	CCCBDB

242	Acetone	1	0	-216.25	0.32	ATcT
243	Propionaldehyde	1	0	-186.8	0.25	ATcT
244	Ethylene glycol	1	0	-389.44	0.49	ATcT
245	Methylthiirane	1	0	46.11	2.0	CCCBDB
246	Thietane	1	0	61.1	1.3	CCCBDB
247	Dimethyl sulfoxide	1	0	-151.3	0.8	CATCH
248	1,2-Ethanedithiol	1	0	-9.33	1.09	CCCBDB
249	Dimethyl disulfide	1	0	-24.2	1.0	CATCH
250	1-Butene	1	0	-0.09	0.38	ATcT
251	Isobutene	1	0	-17.12	0.43	ATcT
252	trans-2-Butene	1	0	-11.32	0.41	ATcT
253	Cyclobutane	1	0	27.82	0.41	ATcT
254	1-Propanol	1	0	-255.28	0.25	ATcT
255	2-Propanol	1	0	-272.97	0.32	ATcT
256	Ethylenediamine	1	0	-17.03	0.59	CCCBDB
257	(Methylthio)ethane	1	0	-60.3	1.1	CCCBDB
258	2-Propanethiol	1	0	-76.94	0.63	CCCBDB
259	1-Propanethiol	1	0	-68.58	0.63	CCCBDB
260	n-Butyl	2	0	80.02	0.72	ATcT
261	t-Butyl	2	0	50.14	0.68	ATcT
262	sec-Butyl	2	0	65.42	0.98	ATcT
263	iso-Butyl	2	0	73.02	0.77	ATcT
264	Trimethylamine	1	0	-27.69	0.92	ATcT
265	n-Butane	1	0	-125.96	0.26	ATcT
266	iso-Butane	1	0	-134.76	0.32	ATcT
267	Dinitrogen trioxide	1	0	86.63	1.0	CCCBDB
268	Tetrafluoromethane	1	0	-933.39	0.26	ATcT
269	Chlorotrifluoromethane	1	0	-710.0	0.72	ATcT
270	Sulfuryl chloride	1	0	-354.8	2.1	CCCBDB
271	Difluorodichloromethane	1	0	-495.18	0.97	ATcT
272	Tetrachloromethane	1	0	-95.67	0.45	ATcT
273	Bromotrifluoromethane	1	0	-651.4	0.51	ATcT

274	Bromotrichloromethane	1	0	-42.01	0.6	ATcT
275	Malononitrile	1	0	266.3	1.0	CCCBDB
276	Sulfuric acid	1	0	-732.73	2.0	CCCBDB
277	Methyl nitrate	1	0	-122.0	1.0	CCCBDB
278	1H-1,2,4-Triazole	1	0	192.7	0.8	CCCBDB
279	1,1,1-Trichloroethane	1	0	-145.0	0.52	ATcT
280	Furan	1	0	-34.7	0.4	CCCBDB
281	1H-Imidazole	1	0	132.9	0.6	CCCBDB
282	1H-Pyrazole	1	0	179.4	0.8	CCCBDB
283	Thiophene	1	0	114.9	1.0	CATCH
284	Cyclopropanecarbonitrile	1	0	182.7	0.7	CCCBDB
285	Methoxyacetonitrile	1	0	-35.65	0.66	CCCBDB
286	Hydrazinecarbothioamide	1	0	128.2	1.6	CCCBDB
287	Cyclopentadiene	1	0	132.73	0.76	ATcT
288	Vinyl ether	1	0	-12.68	0.84	CCCBDB
289	2,3-Dihydrothiophene	1	0	90.7	1.3	CATCH
290	2,5-Dihydrothiophene	1	0	86.9	1.2	CATCH
291	Propanamide	1	0	-258.94	0.66	CCCBDB
292	Isoprene	1	0	75.44	0.7	ATcT
293	Cyclopentene	1	0	33.82	0.45	ATcT
294	Tetrahydrofuran	1	0	-184.2	0.8	CCCBDB
295	Ethoxyethene	1	0	-140.16	0.96	CCCBDB
296	Dimethoxymethane	1	0	-348.2	0.79	CCCBDB
297	Tetrahydrothiophene	1	0	-34.1	1.3	CATCH
298	1,3-Propanedithiol	1	0	-29.83	1.17	CCCBDB
299	Pyrrolidine	1	0	-3.6	0.92	CCCBDB
300	Cyclobutylamine	1	0	41.0	0.42	CCCBDB
301	2-Chlorobutane	1	0	-166.66	0.99	CCCBDB
302	(Z)-2-Pentene	1	0	-26.3	0.84	CCCBDB
303	Cyclopentane	1	0	-77.61	0.44	ATcT
304	2-Methyl-1-propanol	1	0	-283.8	0.9	CCCBDB
305	(R)-(-)-2-Butanol	1	0	-292.7	0.3	CCCBDB

306	2-Methoxypropane	1	0	-252.04	0.96	CCCBDB
307	Methyl propyl ether	1	0	-238.02	0.85	CCCBDB
308	1,2-Diaminopropane	1	0	-53.68	0.46	CCCBDB
309	1-Butanethiol	1	0	-88.3	1.1	CCCBDB
310	2-Butanethiol	1	0	-96.11	0.79	CCCBDB
311	2-Methyl-2-propanethiol	1	0	-108.74	0.88	CCCBDB
312	2-Methyl-1-propanethiol	1	0	-96.48	0.88	CCCBDB
313	2-(Methylthio)propane	1	0	-89.66	0.75	CCCBDB
314	2-Methyl-2-propanamine	1	0	-120.7	0.84	CCCBDB
315	2-Methyl-1-propanamine	1	0	-98.62	0.54	CCCBDB
316	n-Pentane	1	0	-146.38	0.31	ATcT
317	neo-Pentane	1	0	-167.57	0.4	ATcT
318	iso-Pentane	1	0	-153.31	0.44	ATcT
319	2-Butynedinitrile	1	0	533.46	0.8	CCCBDB
320	Dinitrogen tetraoxide	1	0	10.85	0.14	ATcT
321	Tetrafluoroethylene	1	0	-674.73	0.58	ATcT
322	1,3,5-Triazine	1	0	225.87	0.89	CCCBDB
323	o-Benzyne	1	0	459.13	0.88	ATcT
324	4-Methylene-2-oxetanone	1	0	-190.25	0.54	CCCBDB
325	Succinonitrile	1	0	209.66	0.88	CCCBDB
326	2-Methyl-2H-tetrazole	1	0	328.4	0.7	CCCBDB
327	Ethanedithioamide	1	0	83.0	1.5	CCCBDB
328	1,3-Dithiane-2-thione	1	0	93.8	2.0	CATCH
329	Phenyl	2	0	336.83	0.56	ATcT
330	Phenylium	1	1	1135.67	0.88	ATcT
331	Phenide	1	-1	231.01	0.42	ATcT
332	Pyridine	1	0	140.38	0.25	CCCBDB
333	5-Methylisoxazole	1	0	34.06	0.75	CCCBDB
334	3-Methylisoxazole	1	0	35.65	0.67	CCCBDB
335	4-Methylthiazole	1	0	111.75	0.88	CCCBDB
336	Fulvene	1	0	224.26	0.9	CCCBDB
337	Benzene	1	0	83.07	0.23	ATcT

338	gamma–Butyrolactone	1	0	-366.5	0.8	CCCBDB
339	2,3-Butanedione	1	0	-326.75	0.7	ATcT
340	1,3,5-Trioxane	1	0	-465.76	0.5	CCCBDB
341	3-Methylthiophene	1	0	82.59	0.92	CCCBDB
342	2-Methylthiophene	1	0	83.5	0.9	CATCH
343	Dihydro-2-(3H)-thiophenone	1	0	-196.19	1.97	CCCBDB
344	Dihydro-3-(2H)-thiophenone	1	0	-135.27	1.97	CCCBDB
345	Dimethyl ester sulfurous acid	1	0	-483.25	2.09	CCCBDB
346	1-Methyl-1H-Pyrrole	1	0	103.14	0.54	CCCBDB
347	4,5-Dihydro-2-methyl-oxazole	1	0	-130.46	0.92	CCCBDB
348	Ethoxyacetonitrile	1	0	-69.5	0.56	CCCBDB
349	3,4-Dihydro-2H-pyran	1	0	-112.81	0.9	CCCBDB
350	1,4-Dioxane	1	0	-315.3	0.8	CCCBDB
351	Ethyl acetate	1	0	-444.76	0.42	CCCBDB
352	S-Ethyl thioacetate	1	0	-228.2	0.9	CATCH
353	Butanamide	1	0	-279.17	0.88	CCCBDB
354	2-Methylpropanamide	1	0	-282.6	0.9	CCCBDB
355	cyclohexene	1	0	-4.32	0.98	CCCBDB
356	1-Methylcyclopentene	1	0	-3.6	0.75	CCCBDB
357	1,5-Hexadiene	1	0	83.75	0.4	ATcT
358	3-Methyl-2-butanone	1	0	-262.57	0.87	CCCBDB
359	Tetrahydro-2H-pyran	1	0	-223.84	1.0	CCCBDB
360	1,1-Dimethoxyethane	1	0	-389.74	0.79	CCCBDB
361	1,2-Dimethoxyethane	1	0	-342.8	0.7	CCCBDB
362	Tetrahydro-2-methylthiophene	1	0	-63.89	0.75	CCCBDB
363	Tetrahydro-3-methylthiophene	1	0	-60.54	0.84	CCCBDB
364	Cyclopentanethiol	1	0	-47.78	0.75	CCCBDB
365	Trimethylthiirane	1	0	-21.5	1.8	CATCH
366	Tetrahydro-2H-thiopyran	1	0	-63.5	1.0	CATCH
367	Diethyl sulfoxide	1	0	-205.6	1.5	CATCH
368	1,4-Butanedithiol	1	0	-50.54	1.84	CCCBDB
369	Cyclopentanamine	1	0	-54.86	0.92	CCCBDB

370	Piperidine	1	0	-47.15	0.63	CCCBDB
371	Diethylhydroxylamine	1	0	-121.77	0.73	CCCBDB
372	Methylcyclopentane	1	0	-106.69	0.84	CCCBDB
373	3-Methylenepentane	1	0	-56.07	0.88	CCCBDB
374	Ethylcyclobutane	1	0	-27.7	0.7	CCCBDB
375	Cyclohexane	1	0	-123.2	0.36	ATcT
376	1-Pentanol	1	0	-295.63	0.74	CCCBDB
377	1,2-Butanediamine	1	0	-73.55	0.84	CCCBDB
378	2-Methyl-1,2-propanediamine	1	0	-90.25	0.71	CCCBDB
379	1-Pentanethiol	1	0	-110.83	1.76	CCCBDB
380	3-Methyl-1-butanethiol	1	0	-114.64	1.17	CCCBDB
381	1-(Methylthio)butane	1	0	-102.17	1.09	CCCBDB
382	2-Methyl-1-butanethiol	1	0	-114.73	0.96	CCCBDB
383	3-Methyl-2-butanethiol	1	0	-120.96	0.96	CCCBDB
384	Ethyl propyl sulfide	1	0	-104.73	0.79	CCCBDB
385	2-Methyl-2-(methylthio)propane	1	0	-121.04	0.75	CCCBDB
386	2-Methyl-2-butanethiol	1	0	-126.9	0.92	CCCBDB
387	2,2-Dimethyl-1-propanethiol	1	0	-128.7	0.88	CCCBDB
388	2,3-Dimethylbutane	1	0	-177.8	1.0	CCCBDB
389	n-Hexane	1	0	-166.98	0.34	ATcT
390	Dinitrogen pentoxide	1	0	14.84	0.35	ATcT
391	Sulfur Hexafluoride	1	0	-1220.47	0.8	CCCBDB
392	4-Thiazolecarbonitrile	1	0	293.8	1.3	CATCH
393	Phenolate	1	-1	-161.56	0.93	ATcT
394	Phenoxy	2	0	56.08	0.94	ATcT
395	Fluorobenzene	1	0	-114.77	0.89	ATcT
397	Chlorobenzene	1	0	52.17	0.61	ATcT
398	Phenol	1	0	-93.12	0.68	ATcT
399	Benzenethiol	1	0	112.4	0.9	CATCH
400	Dimethyl sulfate	1	0	-687.0	1.9	CATCH
401	Benzyl	2	0	211.42	0.62	ATcT
402	Benzylide	1	-1	123.1	0.45	ATcT

403	Aniline	1	0	87.03	0.88	CCCBDB
404	Toluene	1	0	50.15	0.34	ATcT
405	Ethyl methyl sulfite	1	0	-524.0	2.1	CATCH
406	Cyclohexanone	1	0	-231.1	0.88	CCCBDB
407	S-Isopropyl thioacetate	1	0	-256.3	1.0	CATCH
408	3-Ethylsulphinyl-1-propene	1	0	-103.7	1.9	CATCH
409	(Methylthio)cyclopentane	1	0	-64.7	1.0	CATCH
410	n-Heptane	1	0	-187.58	0.48	ATcT
411	Octasulfur	1	0	100.42	0.63	CCCBDB
412	Metadifluorobenzene	1	0	-309.2	1.0	CCCBDB
413	1,4-Difluorobenzene	1	0	-306.7	1.0	CCCBDB
414	Orthodifluorobenzene	1	0	-283.0	0.92	CCCBDB
415	Benzaldehyde	1	0	-37.02	0.92	ATcT
416	Succinic acid	1	0	-817.79	0.61	ATcT
417	Phenylethene	1	0	148.33	0.55	ATcT
418	Anisole	1	0	-76.69	0.92	CCCBDB
419	meta-Xylene	1	0	17.2	0.75	CCCBDB
420	Ethylbenzene	1	0	29.72	0.55	ATcT
421	Tetramethylthiourea	1	0	44.9	2.3	CATCH
422	n-Octane	1	0	-208.29	0.67	ATcT
423	Nitrobenzene	1	0	68.53	0.67	CCCBDB
424	n-Propyl benzene	1	0	7.82	0.84	CCCBDB
425	1,3-Benzodithiole-2-thione	1	0	242.0	1.5	CATCH
426	Adamantane	1	0	-133.1	1.26	Cioslowski

Table B.1. Reference species included in the database. The index matches the index in the database. Note that species with index 27, 79, 135, 148, 201, and 396 are not in the reference set, as these species were removed before being used in this study. Data are in kJ/mol, with the lowest uncertainty source and data listed.

Appendix C

Supporting Information for: GCxGC-FID/MS and NMR analysis of Low-Temperature Closed-System Pyrolysis of Type I and II Kerogens for Validating Detailed Kinetic Models

Section S1: GCxGC-FID/MS Quantification

Table C.1. Relative mass fractions (y_{rel}) among liquid products for type I demineralized kerogen pyrolyzed at 270°C and 120 h as determined by GCxGC-FID. Since not all products can be observed in GCxGC-FID, the true relative mass fractions among the liquid products are likely lower. The elution times for the primary (t_1) and secondary (t_2) columns are also included.

Species	t_1 [min]	t_2-$t_2^{C_{20}}$ [sec]	Volume [-]	y_{rel} [-]
C33	89.6	2.9	1.82E+07	2.03E-02
C32	87.47	2.1	2.15E+07	2.41E-02
C31	85.6	1.7	2.82E+07	3.16E-02
C30	83.73	1.45	2.67E+07	2.99E-02
C29	81.6	1.4	3.45E+07	3.87E-02
C28	79.47	1.2	2.92E+07	3.27E-02
C27	77.33	1.05	3.05E+07	3.41E-02

C26	75.2	0.85	2.61E+07	2.93E-02
C25	72.8	0.75	2.80E+07	3.14E-02
C20	59.47	0	1.23E+07	1.38E-02
C21	62.4	0.1	1.10E+07	1.23E-02
C22	65.07	0.3	1.93E+07	2.16E-02
C23	67.73	0.45	3.50E+07	3.93E-02
C24	70.4	0.6	4.06E+07	4.55E-02
2,6,10,14-tetramethylheptadecane	56.27	-0.55	4.97E+06	5.57E-03
C19	56.53	-0.15	8.75E+06	9.80E-03
C18	53.6	-0.35	7.41E+06	8.30E-03
phytane	53.6	-0.55	7.89E+06	8.84E-03
C17 + pristane	50.13	-0.65	2.33E+07	2.60E-02
2,6,10-trimethylpentadecane	48.27	-0.65	1.30E+07	1.45E-02
hexadecane	46.4	-0.35	3.23E+06	3.61E-03
2,6,10-trimethyltetradecane	44.8	-0.7	4.17E+06	4.67E-03
pentadecane	42.93	-0.5	5.54E+06	6.21E-03
2,6,10-trimethyltridecane	41.33	-0.75	1.57E+07	1.75E-02
tetradecane	38.93	-0.5	6.05E+06	6.77E-03
2,6,10-trimethyldodecane	38.13	-0.8	7.77E+06	8.70E-03
tridecane	34.93	-0.6	4.75E+06	5.32E-03
7-methyltridecane	33.6	-0.85	9.44E+06	1.06E-02
2,6-dimethylundecane	31.2	-0.8	7.43E+06	8.32E-03
dodecane	30.4	-0.55	5.79E+06	6.48E-03
undecane	25.87	-0.55	8.24E+06	9.23E-03
1,1,2,3-tetramethylcyclohexane	20	0.45	2.18E+07	2.44E-02
3-ethyl-2-methyl-heptane	18.4	-0.55	9.61E+06	1.08E-02
1,1,3-trimethyl-cyclohexane	14.13	-0.25	3.77E+07	4.22E-02
2-methylheptane	11.73	-0.95	8.37E+06	9.37E-03
toluene	12	0.8	4.16E+07	4.66E-02
p-xylene	15.2	1.9	1.90E+07	2.13E-02
6,7-Dimethyl-3,5,8,8a-tetrahydro-1H-2-benzopyran	28.53	3.35	5.20E+06	5.83E-03
1,3-bis(1,1-dimethylethyl)-benzene	33.07	1.35	9.60E+06	1.08E-02

1,2,3,4-tetrahydro-1,1,6-trimethyl-naphthalene	37.6	3.8	9.96E+06	1.12E-02
9,10-dehydro-Isolongifolene	40.53	2.35	7.65E+06	8.57E-03
2-pentyl-2-nonenal	45.33	1.8	5.30E+06	5.93E-03
1,5,9-Trimethyl-1,5,9-cyclododecatriene	45.33	3.2	1.23E+07	1.38E-02
1,1,4,5,6-Pentamethyl-2,3-dihydro-1H-indene	44	4.95	1.70E+07	1.91E-02
trimethylnaphthalenes	44.53	6.25	6.51E+07	7.30E-02
1,7-dimethyl-naphthalene	40.27	6.8	7.20E+06	8.07E-03
diphenyl ether	39.2	7.55	1.23E+07	1.37E-02
6,10,14-trimethyl-2-Pentadecanone	54.67	1.4	4.38E+06	4.91E-03
5-(para-Phenoxyphenyl)pentanal	63.47	6.75	1.22E+07	1.37E-02
4b,8-Dimethyl-2-isopropylphenanthrene, 4b,5,6,7,8,8a,9,10-octahydro-	62.93	7.6	9.98E+06	1.12E-02
1,1'-ethylidenebis[3,4-dimethyl-benzene	64.53	9.35	9.23E+06	1.03E-02
28-Nor-17α(H)-hopane	86.13	8.1	1.05E+07	1.17E-02
17α(H),21β(H)-Hopane	87.73	8.75	1.09E+07	1.22E-02
Cadina-1(10),6,8-triene	47.73	5.1	7.94E+06	8.90E-03
tetramethylnaphthalenes	51.2	7.4	3.35E+07	3.75E-02