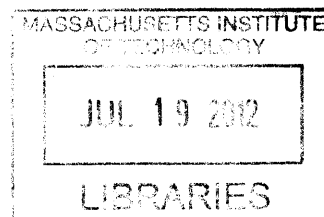


ARCHIVES

**Automated Reaction Mechanism Generation:
Improving Accuracy and Broadening Scope**



by

Gregory Russell Magoon

B.S. in Chemical Engineering, University of Connecticut, 2006

M.S. in Chemical Engineering Practice, Massachusetts Institute of Technology, 2008

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

June 2012

©Massachusetts Institute of Technology 2012. All rights reserved.

Author
.....

Department of Chemical Engineering
May 18, 2012

Certified by
.....

William H. Green
Professor of Chemical Engineering
Thesis Supervisor

Accepted by
.....

Patrick S. Doyle
Professor of Chemical Engineering
Chairman, Committee for Graduate Students

Automated Reaction Mechanism Generation:
Improving Accuracy and Broadening Scope

by

Gregory Russell Magoon

Submitted to the Department of Chemical Engineering on May 18, 2012
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Chemical Engineering

Abstract

Chemical kinetic modeling plays an important role in the study of reactive chemical systems. Thus, an automated means of constructing chemical kinetic models forms a useful tool in the engineering and science surrounding such systems. This document describes work to further develop one such tool, known as RMG (Reaction Mechanism Generator). Focus is placed on improving the accuracy of parameter estimation in the mechanism generation process and expanding the scope of applicability of the tool. In particular, effort has targeted the generation and use of explicit three-dimensional molecular structures for chemical species considered during reaction mechanism generation. This work has resulted in the generation of a software system integrated with RMG that can automatically generate and use such structures with quantum chemistry or force field codes to obtain more reliable thermochemistry estimates for cyclic structures without human intervention. Ultimately, the result of these updates is improved usefulness and reliability of the software system as a predictive tool.

An application of the tool to the high temperature oxidation of JP-10, a jet fuel often used in military applications, is described. Using the newly refined RMG system, a detailed chemical kinetic model was constructed for this system. The resulting model represents a significant improvement upon existing work for JP-10 oxidation by capturing detailed chemistry for this system. Simulations with this model have been found to produce results for ignition delay and product distribution that compare favorably with experimental results. The successful application

of the refined RMG software system to this system demonstrates the practical utility of these updates.

Thesis Supervisor: William H. Green

Title: Professor of Chemical Engineering

Dedication

I dedicate this thesis to my family, and in particular, my parents and my sister.

Acknowledgements

Firstly, I'd like to thank my advisor, Prof. William H. Green who has provided excellent guidance throughout my years in his group and has been extremely helpful when things didn't go right.

I'd also like to express sincere thanks to my thesis committee members, Prof. Bruce Tidor, Prof. Klavs Jensen, and Prof. Arup Chakraborty. My discussions with them have been extremely fruitful in propelling my thesis work forward.

Also, I'd like to express great appreciation to Barbara Balkwill for all her assistance over the years. Also, thanks to Jean Belbin, Jim Hardsog, and Rebecca Hailu for IT support and to the rest of the Chemical Engineering staff, particularly Suzanne Maguire.

Thanks to my classmates in the Chemical Engineering graduate program, particularly David Couling and Geoff Oxberry, in the Green Group.

I'd like to thank all Green Group members and RMG developers, past and present. Also I want to thank RMG users who have provided helpful feedback.

I would particularly like to thank Michael Harper, Sandeep Sharma, Franklin Goldsmith, and Rob Ashcraft for helping me to get started when I first joined the Green Group.

Special thanks to Richard West who has played an integral role in steering efficient, next-generation software development of RMG and who has been extremely helpful as I have worked to develop my programming and modeling skills.

I'd also like to express my great appreciation to Josh Allen who has been invaluable in providing programming, Linux, and LaTeX tips and assistance during much of my time in the Green Group. Thanks also to Connie Gao, who, with Josh, has developed extremely useful web tools in support of reaction mechanism generation.

Thanks to Ray Speth for his great assistance in learning the ropes of Linux system use, design, and administration.

Thank you to Shamel Merchant for helpful discussions regarding CHEMKIN and RMG; I'd particularly like to express my appreciation for your willingness to coordinate with me to share license usage.

Thanks to Amrit Jalan for stimulating discussions. It has been a pleasure to work with you as fellow RMG developer.

Thank you to Luwi Oluwole, Hsi-Wu Wong, David Lewis, Simon E. Albo, and Robin Edwards of Aerodyne Research, Inc. whose sponsorship and collaboration in studying the JP-10 system has been instrumental in a large portion of the work described here.

I'd like to thank Nick Vandewiele for very stimulating discussions, for great feedback, and for playing an integral role in our efforts to better understand and model JP-10 decomposition chemistry.

Thank you to Jorge Aguilera-Iparraguirre and Yu Shi for helpful discussions and assistance on *ab initio* quantum chemistry calculations in support of our JP-10 modeling efforts. (Thanks to Jorge also for his extremely helpful assistance during MOLPRO installation on our new server!)

Thanks to Nathan Yee for his invaluable and timely assistance in the final phases of JP-10 model refinements.

Thanks to UROP student Ben Ruiz-Yi for his assistance in the critical first generation of JP-10 model refinements.

I would particularly like to thank Richard Burnes of the Naval Air Warfare Center for his sponsorship of the project and for his insightful questions and comments.

I'd also like to thank the support staff at ERDC, MHPCC, and AFRL DoD Supercomputing Resource Centers, for their great efforts to provide useful and reliable high-performance computational resources.

Thanks to Green Group post-docs Yury Suleymanov (for Russian-to-English translation assistance) and John Alecu (for helpful discussions regarding conformational flexibility and symmetry numbers).

I'd like to thank Dr. Markus Kraft and the rest of the Computational Modelling Group, particularly Raphael Shirley and Markus Sander, for hosting me during my visit to Cambridge University in Summer 2008.

I'd also like to thank the Combustion Energy Frontier Research Center, particularly Prof. Chung Law and Lillian Tsang, for hosting a very helpful and informative Combustion Summer School session at Princeton University in 2010.

Thanks to my collaborators in my study of intramolecular disproportionation, Prof. Piotr Piecuch and Jesse Lutz of Michigan State University, for working with me on a very interesting project.

I'd like to thank Greg Landrum, Noel O'Boyle, Geoffrey Hutchinson, Egon Willighagen and the rest of the open-source cheminformatics community, without whom this work would have been much more difficult.

Thanks to the technical support teams for Gaussian (particularly Douglas Fox), MOLPRO (particularly Andy May), MM4 (Jenn-Huei Lii), and CHEMKIN (particularly Cheng Wang).

In addition to program support by the Naval Air Warfare Center and Aerodyne Research, Inc., I'd like to thank the Department of Energy and the National Science Foundation for financially supporting the development of RMG.

Finally, I'd like to acknowledge helpful discussions with Professor Linda Broadbelt of Northwestern University, Laurent Catoire of CNRS, Dr. William J. Pitz of Lawrence Livermore National Laboratory, and Larry Harding and Jim Miller of Argonne National Laboratories over the course of this work.

Table of Contents

1	Chapter 1: Introduction.....	13
1.1	Background.....	13
1.1.1	Automated reaction mechanism generation.....	13
1.1.2	RMG.....	13
1.1.3	Rate-based model construction algorithm.....	14
1.1.4	Parameter estimation in automated reaction mechanism generation.....	14
1.2	Overview.....	15
2	Chapter 2: Connectivity-based thermodynamic parameter estimation.....	17
2.1	Introduction.....	17
2.2	Evaluation of connectivity-based approaches to account for steric effects.....	18
2.2.1	Heat of formation for 2,2,4,4,6,8,8-heptamethylnonane.....	21
2.2.2	Benson vs. “methyl repulsion” approach for other branched alkanes.....	25
2.3	Implementation of connectivity-based steric corrections in RMG.....	26
2.4	Concluding remarks.....	29
3	Chapter 3: Generation of explicit 3D structures from connectivity representations.....	33
3.1	Introduction.....	33
3.2	Background.....	33
3.3	Preliminary screening.....	34
3.4	RDKit testing.....	36
3.5	Summary.....	37
4	Chapter 4: Design and implementation of explicit-3D-geometry-based on-the-fly species thermochemistry.....	38
4.1	Background.....	38
4.2	Design and implementation of the QMTP system.....	39
4.2.1	Design overview.....	39
4.2.2	Three-dimensional geometry structure generation.....	40
4.2.3	Format conversion and output parsing.....	41
4.2.4	Calculation method.....	41
4.2.5	Failure checking and recovery.....	41
4.2.6	Storage of results and use of modified InChI and InChIKey.....	44

4.2.7	Connectivity checking	46
4.2.8	External symmetry number and chirality calculation	49
4.2.9	Force field and rotor scan capabilities	49
4.2.10	Scope of QMTP calculations	50
4.3	Testing of the QMTP system	51
4.3.1	Accuracy of estimates	51
4.3.2	Effect on kinetic models	52
4.3.3	Robustness and speed	54
4.4	Opportunities for improvement and applying the approach to other areas	56
4.4.1	Selective use of more accurate methods	56
4.4.2	Improvements to treating conformational flexibility	57
4.4.3	Polarizability estimates	57
4.4.4	Solvation property estimation	58
4.4.5	Standalone thermodynamic property estimation	58
4.4.6	Kinetic parameter estimation	58
4.5	Summary	59
5	Chapter 5: Practical considerations of explicit-3D-geometry-based on-the-fly species thermochemistry	60
5.1	Introduction	60
5.2	Source code and dependencies	60
5.2.1	Modules	60
5.2.1.1	Modifications to <i>cclib</i>	62
5.2.1.2	Modifications to CanTherm	62
5.3	RMG input file considerations	62
5.4	Symmetry and chirality statistical effects	64
5.4.1	External symmetry number and chirality	64
5.4.1.1	Special considerations for radicals	65
5.4.2	Rotor symmetry number	66
5.5	Directories created and used by the QMTP system	67
5.5.1	Description of directories and their contents	67
5.5.2	Managing QMfiles library	69

6	Chapter 6: Application of explicit-3D-geometry-based on-the-fly species thermochemistry to JP-10 oxidation	70
6.1	Background	70
6.2	Mechanism development	71
6.2.1	1 st generation mechanisms: capturing initial decomposition to C ₅	72
6.2.2	2 nd generation mechanisms: avoiding suspicious cyclic C ₅ decomposition routes..	79
6.2.3	3 rd generation mechanism: capturing multiple reaction conditions, pressure-dependence, and aromatic chemistry	86
6.2.4	4 th generation mechanisms: refinement of chemistry based on experimental comparisons and creation of a new comprehensive (pyrolysis+combustion) mechanism ...	93
6.2.4.1	Ignition delay.....	96
6.2.4.2	Pyrolysis speciation	98
6.2.4.3	Flux analysis.....	108
6.3	Model strengths, limitations, and next steps	112
6.4	Conclusions	114
7	Chapter 7: Towards on-the-fly kinetic parameter estimation for improving intra-molecular disproportionation estimates	115
7.1	Introduction / motivation.....	115
7.2	Methodology	120
7.3	Results and Discussion.....	123
7.3.1	Intramolecular disproportionation	123
7.3.1.1	C ₃ H ₈ System	123
7.3.1.2	C ₆ H ₁₂ System.....	125
7.3.1.3	C ₁₀ H ₁₆ System	129
7.3.1.4	Discussion / Kinetic modeling of the C ₁₀ H ₁₆ system	134
7.3.2	Ring-opening pathways with concerted hydrogen transfer	136
7.3.2.1	Background	136
7.3.2.2	DFT calculations	136
7.3.2.3	Kinetics implications.....	139
7.4	Conclusions	141
8	Chapter 8: Opportunities for further investigation and conclusions.....	143

8.1	Introduction	143
8.2	Improvement to reaction mechanism generation algorithm	143
8.3	Better approaches for accounting for conformational flexibility	144
8.3.1	Models to account for conformational flexibility	144
8.3.2	Enumeration of conformers	146
8.3.2.1	Conformational equivalence testing	146
8.3.2.1.1	Background	146
8.3.2.1.2	The MoleCoor algorithm	149
8.3.2.1.3	Results	156
8.3.2.1.4	Implementation	158
8.3.2.1.5	Summary	158
8.4	Improved kinetic parameter estimates based on explicit three-dimensional structures	159
8.5	Main contributions	162
8.6	Conclusions	163
9	Appendix I: Glossary for ambiguous species names	164
10	Appendix II: Derivation of matrix formulation for variable moment of inertia treatment of hindered rotors	167
	References	169

1 Chapter 1: Introduction

1.1 Background

1.1.1 Automated reaction mechanism generation

Detailed chemical kinetic models aim to capture all the relevant chemistry for a system of interest, including relevant elementary reaction pathways and associated kinetic and thermodynamic parameters. Detailed chemical kinetic models of reacting systems have been useful in many applications, including combustion, pyrolysis, and atmospheric chemistry. A typical detailed chemical kinetic model will consist of hundreds of molecular species and thousands or tens of thousands of elementary reactions. Once constructed, such a model can be used to gain fundamental insight into the underlying chemistry of the system or to perform predictive simulations (*e.g.* reactive computational fluid dynamics simulations for combustor design).

The availability of a tool to automate the process of kinetic model generation bypasses the process of constructing kinetic models by hand, which can be very time-consuming and error-prone. In an ideal scenario, such a tool is a black box that takes information from the user regarding reactant structures and conditions of interest (temperature and pressure, and initial concentrations) and returns a detailed chemical kinetic model containing important intermediates, products, byproducts, the important reactive steps, and estimates of the relevant kinetic and thermodynamic parameters. As such, automated reaction mechanism generation software systems can be an extremely useful predictive tool for both engineers and scientists.

1.1.2 RMG

RMG (Reaction Mechanism Generator) is a Java software system for automated reaction mechanism generation.¹ It was originally developed by Dr. Jing Song, under the guidance of Prof. William Green in the Department of Chemical Engineering at the Massachusetts Institute of Technology.² To generate a kinetic model for a system of interest, the user specifies reaction conditions (*i.e.* temperature, pressure, reactant structures and concentrations) and the program constructs kinetic models composed of elementary chemical reaction steps, while determining the associated species concentration trajectories. In this sense, RMG is similar to other

automated reaction mechanism generation programs such as EXGAS.³ However, several unique features of RMG distinguish it from similar tools, including the use of a rate-based model construction algorithm,⁴ described further below. Additionally, integrated consideration of pressure dependence has been incorporated into RMG.⁵ Other recent RMG development efforts have focused on data collaboration with the PrImE community and expanding flexibility of atom types beyond carbon, hydrogen, and oxygen.⁶ Previous applications of RMG include the (steam-cracking) pyrolysis of *n*-hexane⁷ and the chemistry of various hexadiene⁸ and butanol⁹ isomers in diffusion flames.

1.1.3 Rate-based model construction algorithm

RMG uses a version of the rate-based model-construction algorithm described by Susnow *et al.*⁴ In this algorithm, the model is partitioned into a “core” and an “edge”, with the initial reactants serving as seed to the core. The edge consists of products of reactions amongst core species. At each iteration of the algorithm, the core is simulated and when flux to an edge species is found to exceed a user-defined numerical tolerance, that species is moved into the core. This process repeats until the core can be simulated to a target time or conversion without flux thresholds being exceeded for any of the edge species.

1.1.4 Parameter estimation in automated reaction mechanism generation

The RMG software system brings together a number of components, including routines for reaction enumeration and cheminformatics, ODE solvers for model simulation, general program structure and logic, and parameter estimation routines. RMG’s parameter estimation routines are particularly important in the process of mechanism generation; not only are the resulting kinetic and thermodynamic parameters directly reflected in the resulting detailed chemical kinetic model, but also they affect the mechanism generation process itself, influencing which species will be incorporated and which pathways are explored. Moreover, these estimation routines present a clear opportunity for improvement as the scope and accuracy of these routines is one of the biggest obstacles to black-box automated reaction mechanism generation.

Obviously these parameter estimates should be as accurate as possible, but practical considerations suggest that uncertainties on the order of a factor of ten in rate coefficients and equilibrium constants are a reasonable goal. This level of accuracy should be sufficient to produce a model that is “in the right ballpark” and sensitivity analysis can be used to determine

where further effort should be placed for parameter refinement. (In most cases, only a small fraction of the parameters must be estimated accurately for accurate modeling and it is sufficient to use rough estimates for the remainder.)

Originally, RMG used Benson's group additivity method to estimate thermodynamic values used during mechanism generation. Benson's group additivity method works very well when appropriate group values are available. However, accuracy can severely suffer when groups are chosen that are not the best for the system at hand or will not work at all if such groups are not available, limiting the scope of thermodynamic parameter estimation. In particular, for cyclic species, special *ad hoc* ring corrections must be applied. It would be impractical to exhaustively include such specialized ring corrections for all the cases that could arise. As a result of such shortcomings, a more general means of estimating thermodynamic properties "on-the-fly" is desired. This would enable RMG to handle a broader range of compounds, including species with interesting ring structures.

Similarly, kinetic parameters are estimated in RMG using "rate-libraries" which enumerate rules for estimating kinetic parameters, such as activation energy, based on the nature of the groups involved in reaction. Each of RMG's reaction families has its own rate library. In certain cases, kinetic parameters estimates will be based on experimental values from carefully-studied reactions. In other cases, however, parameters must be estimated for unstudied reactions, and values must be extrapolated from well-studied cases. As with the original thermochemistry estimates, kinetic parameters are currently estimated using groups based on molecular connectivity. First order ring-effects for reactions such as intramolecular hydrogen abstraction are taken into account by considering the size of the cycle formed in the transition state.

1.2 Overview

This chapter, Chapter 1, introduces automated reaction mechanism generation, its implementation through RMG, its uses, and the importance of parameter estimation.

Chapter 2 provides a discussion of connectivity-based thermodynamic parameter estimation in RMG. Different approaches for accounting for non-nearest neighbor steric effects in branched hydrocarbons are discussed and the implementation of one such approach in RMG is described.

Chapter 3 describes an evaluation of various tools for three-dimensional molecular structure generation. The selection and testing of one such tool, *RDKit*, is described.

Chapter 4 describes the implementation of a software framework within RMG, termed the QMTP system, for performing on-the-fly thermochemistry estimates using explicit three-dimensional structures.

Chapter 5 follows with a description of some of the practical considerations and implementation details associated with the QMTP system.

Chapter 6 presents a discussion of the application of RMG (with QMTP system features) to the case of JP-10 combustion and work on development of a comprehensive, validated JP-10 combustion model is described.

Chapter 7 presents high-level *ab initio* calculations performed on disproportionation systems towards improving intra-molecular disproportionation estimates.

Chapter 8 discusses opportunities for obtaining further insights and improvements, summarizes the main contributions of this work, and presents concluding remarks.

2 Chapter 2: Connectivity-based thermodynamic parameter estimation

2.1 Introduction

Accurate estimation of ideal-gas thermochemical parameters is a key consideration in many engineering applications. Thermodynamic parameters (enthalpy of formation, entropy, and heat capacity as a function of temperature) are particularly important in kinetic modeling. Over the years, many approaches have been developed to estimate thermodynamic quantities in a manner that is accurate, general, and fast or “inexpensive” (using as few human or computer resources as possible). The often competing nature of these aims leads to a wide range of viable methodologies that are each, in some sense, near-optimal for a given application.

In the case of kinetic modeling for gas kinetic applications (covering pyrolytic and combustion systems), accurate thermodynamic parameters (with errors on the order of kcal/mol) covering a wide temperature range (typically at least 300 K – 1500 K) must be estimated for hundreds of chemical species. Frequently, these chemical species can be branched or cyclic (both aromatic and non-aromatic) and can also include radical sites (unpaired electrons). The process of automatically building kinetic models using a software system such as RMG (Reaction Mechanism Generator)¹ is even more demanding, as tens or hundreds of thousands of chemical species are enumerated in the process of determining which are suitable for inclusion in the final kinetic model, and thermodynamic parameters must be estimated for these species as well. These needs call for fast, general thermodynamic estimation procedures with reasonable accuracy. Traditionally, group additivity approaches, such as the method designed by Benson,¹⁰ have been used. These approaches require group values obtained through regression analysis of “known” target property values (e.g. standard heat of formation at 298.15 K) for a representative set of molecules. These pre-computed group values may then be extended to estimate properties for molecules outside the training set. Despite the long history of these methods, they continue to be an active area of research.¹¹ These approaches are not without limitations, as will be described later, but often perform satisfactorily for a wide range of molecular species.

Originally, RMG was implemented using the Benson group additivity scheme based on 1976 parameter values. However, the implementation did not consider steric effects present in

branched alkanes. In this chapter, two competing approaches to account for such steric effects are evaluated, with focus on highly-branched alkanes, particularly those related to 2,2,4,4,6,8,8-heptamethylnonane, which serves as a benchmark for the cetane scale for diesel fuels in addition to being a primary reference fuel component. The implementation of suitable steric corrections in RMG is then discussed, including a description of an approach for implementing bond-centered groups within an atom-centered group framework (as used by RMG). The chapter closes with a discussion of connectivity-based approaches in the context of other estimation approaches discussed in this thesis as well as opportunities for further refinement.

2.2 Evaluation of connectivity-based approaches to account for steric effects

Steric effects arising from non-nearest neighbor effects require special considerations in these approaches. These effects are particularly pronounced in highly branched alkanes. The Benson group additivity approach¹⁰ and the “methyl repulsion” group additivity approach of Domalski and Hearing¹² are two of the most popular approaches for obtaining quick connectivity-based estimates using group additivity while taking these non-nearest-neighbor effects into account. The Benson and “methyl repulsion” methods are similar in principle, but differ in their approach to treating non-nearest neighbor steric effects, as well as in the raw group values.

The Benson group additivity approach has a long history, being first proposed and parametrized, including destabilizing non-nearest neighbor “gauche interactions”, in 1958;¹³ it was then extended in 1969.¹⁴ The values from their 1976 update are amongst the most widely used, and included the introduction of new non-nearest-neighbor effects in the form of destabilizing “1,5-interactions.”¹⁵ Cohen and Benson later updated Benson’s gauche interaction counting scheme and updated the group values in 1992.¹⁶

In the Benson scheme, gauche interactions are 1,4-interactions in branched species; the gauche interaction counting schemes of the 1976 and 1992 Benson approaches are summarized in Table 1. Examples of gauche and 1,5-interactions, are illustrated in Figure 1 and Figure 2. The 1976 and 1992 alkane group values are reproduced in Table 2.

Table 1. Comparison of Benson gauche counting schemes. Adapted from Cohen and Benson.¹⁶

Benson (1976) gauche counting scheme			Revised Benson (1992) gauche counting scheme		
C _n atom group	C _m atom group	# gauche interactions	C _n atom group	C _m atom group	# gauche interactions
P	P, S, T, or Q	0	P	P, S, T, or Q	0
S	S (trans/gauche)	0/1	S	S	0
S	T (gauche/syn)	1/2	S	T	1
S	Q	2	S	Q	2
T	T (trans/cis)	2/3*	T	T	3
T	Q	4	T	Q	5
Q	Q	6	Q	Q	8

*TT will have 5 gauche interactions

P=primary; S=secondary; T=tertiary; Q=quaternary

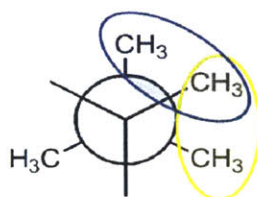


Figure 1. Example illustrating two gauche or 1,4-interactions

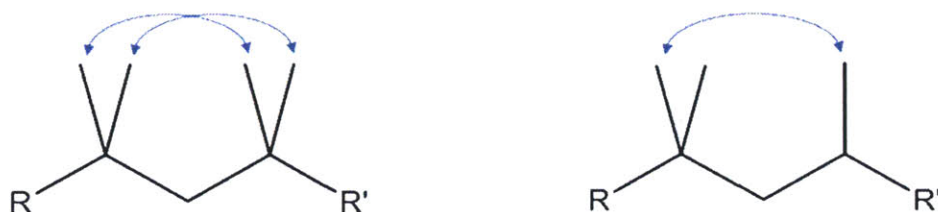


Figure 2. Example branched alkanes with two and one 1,5-interactions, respectively.

Table 2. Benson group additivity group values for alkanes from 1976 and 1992. Adapted from Cohen and Benson.¹⁰

Group	$\Delta\Delta H_{f,298}$ (kcal/mol)	$\Delta\Delta H_{f,298}$ (kcal/mol)
	(1976 values)	(1992 values)
C-(C)(H) ₃ = P	-10.2	-10.00
C-(C) ₂ (H) ₂ = S	-4.93	-5.00
C-(C) ₃ (H) = T	-1.90	-2.40
C-(C) ₄ = Q	0.50	-0.10
gauche (1,4) correction = G	0.80	0.80
1,5-correction = F	1.5	1.60

In 1988, Domalski and Hearing attempted to address limitations in Benson's 1976 approach, particularly with regard to steric effects in branched hydrocarbons.¹² They proposed the "methyl repulsion" method. The name is somewhat of a misnomer as, in this approach, methyl ligands actually provide a stabilizing effect relative to larger ligands. Rather than applying *post hoc* steric corrections as in the Benson approach, the "methyl repulsion" method appears to incorporate the steric effects in the raw group values, which take the degree of the carbon center (i.e. primary, secondary, tertiary, or quaternary) into account; stabilizing corrections are then added when the substituent is small (i.e. methyl). The Domalski and Hearing approach, with its "methyl repulsion" method of accounting for steric effects in branched hydrocarbons is widely used; it is, for example, the recommended approach for estimating heat of formation in the latest *Perry's Chemical Engineer's Handbook, 8th ed.*¹⁷ Alkane group values for Domalski and Hearing's "methyl repulsion" method are summarized in Table 3.

Table 3. Gas-phase heat of formation group values in Domalski and Hearing's "methyl repulsion" method. Adapted from Domalski and Hearing.¹⁸

Group	$\Delta\Delta H_{f,298}$ (kcal/mol)
C-(H) ₃ (C)	-10.10
C-(H) ₂ (C) ₂	-4.931
C-(H)(C) ₃	-0.280
-CH ₃ corr (tertiary)	-0.540
C-(C) ₄	+4.589
-CH ₃ corr (quaternary)	-1.090
-CH ₃ corr (tert/quat)	-0.430
-CH ₃ corr (quat/quat)	-0.153

A comparison of the raw group values between Benson and "methyl repulsion" methods shows that the most significant difference is for quaternary carbons, which likely arises mainly from the different approach to steric corrections.

The appropriate interpretation of the "tertiary", "quaternary", "tert/quat", and "quat/quat" is not immediately obvious for certain molecules. In particular, the questions arise: 1) Should a methyl on a tertiary carbon always use the "tertiary" value or use the "tert/quat" value if a quaternary carbon is nearby? and 2) Do the "tert/quat" and "quat/quat" designations refer to quaternary carbons directly adjacent, or are carbons one spot removed considered as well? The table format and text of Domalski and Hearing¹² seem to imply that methyls on a tertiary carbon should always use the "tertiary" value and that "tert/quat" and "quat/quat" refer to tertiary or

quaternary carbons directly bonded to a quaternary carbon. However, careful examination of the estimated values and a worked example for 2,2,4,4-tetramethylpentane suggests an alternative counting scheme, wherein methyls on a ternary carbon can use the “tert/quat” value if a quaternary carbon is nearby and “tert/quat” and “quat/quat” can refer to carbons either immediately adjacent or one spot removed. In the text that follows, the counting scheme suggested by the tables/text will be termed A, while a counting scheme in the spirit of the treatment implied by the alkane estimates and example in Domalski and Hearing¹² will be termed B.

The heat of formation for 2,2,4,4,6,8,8-heptamethylnonane presents an interesting case study for these alternative approaches to treating steric effects.

2.2.1 Heat of formation for 2,2,4,4,6,8,8-heptamethylnonane

2,2,4,4,6,8,8-heptamethylnonane (otherwise known as isocetane, and henceforth referred to as HMN) is a highly-branched alkane, most widely known for its role as a primary reference fuel component and benchmark for the cetane scale for diesel fuels. There have been recent efforts to construct detailed chemical kinetic models for HMN combustion.¹⁹ The structure of this highly branched hydrocarbon is shown in Figure 3.

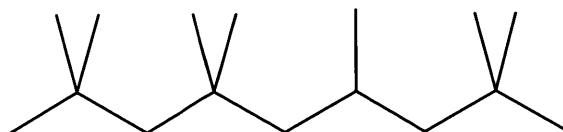


Figure 3. Structure of 2,2,4,4,6,8,8-heptamethylnonane (HMN)

Several approaches were used to estimate the standard heat of formation of HMN at 298.15 K, as shown in Table 4. In addition to using the previously-discussed Benson and “methyl-repulsion” methods, we found an estimate published by DIPPR²⁰, reportedly based on Benson group additivity (though the value suggests that older group values were used or some modifications were applied). Additionally, we have included an estimate²¹ from the program Thergas (which uses a group additivity approach, likely based on the “methyl repulsion” method). We also estimated the heat of formation by applying the “2C” correlation from Somayajulu and Zwolinski, a connectivity-based approach for acyclic alkanes, which takes non-local connectivity effects into account.²² Additionally, the MM4 force field of Allinger *et al.*²³ was applied to the problem. The MM4 approach has been found to reliably reproduce alkane heats of formation.²³

A conformation search with the MM4 program with default parameters (2000 “pushes”) identified 76 distinct low-energy conformations. The minimum energy MM4 conformation was used as a reference onto which corrections were applied. A B3LYP/CBSB7-optimized version of this conformation is shown in Figure 4. MM2 rotor scans in Chem3D (part of the ChemBioOffice suite) suggested that one of the (non-methyl) rotors had a barrier that was below 5 kcal/mol, prompting a “TOR” correction of +0.5715 kcal/mol. Following the recommended approach of Allinger *et al.*,²⁴ contributions from the higher energy conformations were taken into account; through this, we estimated a “POP” contribution of +0.4098 kcal/mol. (Note that for some of the conformations, MM4 did not include contributions from low frequency vibrational modes; we expect that the error introduced by this is very probably less than 0.25 kcal/mol, and likely much less, so we did not make an effort to correct for it.) For the approach labeled CBS-QB3-BAC, we first applied the CBS-QB3 method²⁵ in Gaussian03²⁶ to the minimum energy conformation from the MM4 conformation search; for this relatively large C₁₆H₃₄ molecule, the calculation is quite computationally demanding. We then applied an atomization scheme with bond-additivity corrections, as determined by Sabbe *et al.*²⁷ to arrive at the final value.

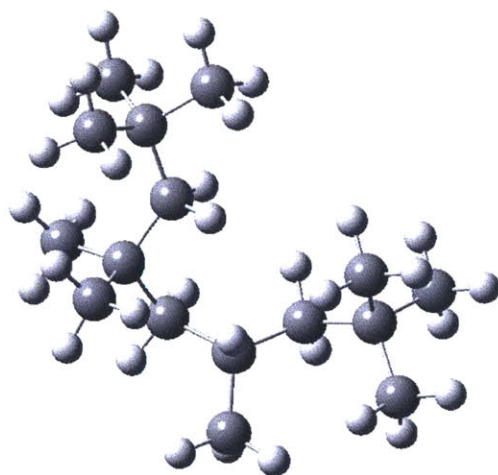


Figure 4. Minimum energy configuration for HMN from MM4 search, subsequently optimized with B3LYP/CBSB7

Table 4. Estimates for standard heat of formation of HMN

Method/Source	$\Delta H_{f,298}$ (kcal/mol)
Benson group additivity (1992 rev.; <i>no steric corrections</i>)	-107.7
“methyl repulsion” method (A/B)	-101.5/-94.7
DIPPR estimate ²⁰	-98.8
Thergas estimate ²¹	-96.3
TRC Web Thermo Tables ²⁸	-90.6 ± 3.1
Benson group additivity (1992 rev.)	-93.3
Somayajulu and Zwolinski correlation ²²	-93.7
MM4	-93.5
CBS-QB3-BAC	-93.2

The fact that the last four approaches give results that agree within about 0.5 kcal/mol provides a strong indication that they all offer a reliable estimate of the true value. More specifically, based on these four results, -93.5 ± 1.0 kcal/mol seems to be a reasonable estimate for the true heat of formation. The difference in Benson values in the table also illustrates the significant magnitude (over 10 kcal/mol) that the destabilizing steric effects have in this molecule. The results suggest that the Benson method is significantly more accurate than the “methyl repulsion” (A) method in this case. It is also noted that the value from the curated and critically evaluated Thermodynamics Research Center (TRC) Web Thermo Tables (WTT) overlaps with the suggested estimate, given the uncertainties; the source of the TRC WTT estimate is not clear, however. Furthermore, the heat of formation for HMN used in the combustion mechanism by Oehlschlaeger *et al.* is -99.0 kcal/mol;^{19a,29} this coincides most closely with the DIPPR estimate in Table 4, though it is more than 5 kcal/mol lower than both the central value of the estimate suggested here and the TRC WTT value.

Although the agreement of the last four approaches appears to provide strong validation for these values and the Benson approach, we were not aware of an experimental value for the heat of formation for HMN for comparison, and we decided to seek further validation from two structurally similar molecules with experimental data for heat of formation: 2,2,4,6,6-pentamethylheptane and 2,2,4,4,6-pentamethylheptane. These are shown in Figure 5, below:

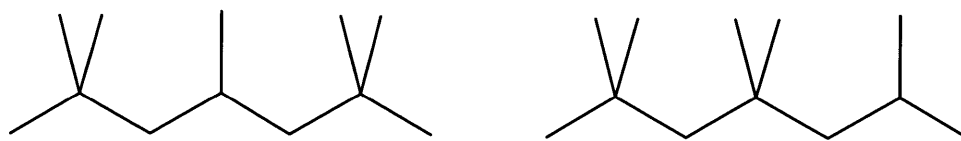


Figure 5. Structures of 2,2,4,6,6-pentamethylheptane and 2,2,4,4,6-pentamethylheptane

Results of application of various approaches, along with experimental data from Desai³⁰ and TRC recommended values (which are apparently based on both experiment and a Somayajulu-based estimate)³¹, are shown in Table 5.

Table 5. Estimates for standard heat of formation of structural analogs for HMN

Method/Source	$\Delta H_{f,298}$ (kcal/mol)	
	2,2,4,6,6-pentamethylheptane	2,2,4,4,6-pentamethylheptane
“methyl repulsion” method (A/B)	-78.7/-74.7	-78.2/-73.3
Benson group additivity (1992 rev.)	-74.6	-72.2
Somayajulu and Zwolinski correlation ²²	-74.9	-72.15
experiment ³⁰	-75.45	-72.1
TRC ³¹	-75.4	-71.6

In addition to matching closely with the experimental values, the Benson estimates capture both the direction and most of the magnitude of the difference in heat of formation between 2,2,4,6,6-pentamethylheptane and 2,2,4,4,6-pentamethylheptane.

Although the “methyl repulsion” (A) method also captures the relative instability of 2,2,4,4,6-pentamethylheptane, it significantly underestimates the magnitude of this difference. The “methyl repulsion” (B) method appears to be somewhat better, though still not as good as the Benson approach, at capturing this difference.

It is noted that the differences between Benson and “methyl repulsion” (A) approaches for these three molecules are of the right direction and magnitude to suggest that large portions of the apparent errors in the “methyl repulsion” approach may be due to the lack of destabilizing 1,5-interactions. It is therefore hypothesized that a refined version of the “methyl repulsion” approach, reparametrized with an additional “group” for 1,5-interactions, may produce significantly improved results, assuming the “A” counting scheme is used.

Without such reparametrization, however, it would appear that counting scheme B is more appropriate than counting scheme A. A summary of the methyl group counts for these alternative counting schemes is included in Table 6.

Table 6. Methyl correction group counts used in “methyl repulsion” counting schemes A and B, for cases where the two schemes differ

Molecule	Scheme A				Scheme B			
	-CH3 (tert)	-CH3 (quat)	-CH3 (quat/tert)	-CH3 (quat/quat)	-CH3 (tert)	-CH3 (quat)	-CH3 (quat/tert)	-CH3 (quat/quat)
2,2,4,4,6,8,8-heptamethylnonane	1	8	0	0	0	0	4	5
2,2,4,6,6-pentamethylheptane	1	6	0	0	0	0	7	0
2,2,4,4,6-pentamethylheptane	2	5	0	0	0	0	2	5
2,2,3-trimethylbutane	2	0	3	0	0	0	5	0
2,2,3,4-tetramethylpentane	3	0	3	0	0	0	6	0
2,3,3,4-tetramethylpentane	4	0	2	0	0	0	6	0
2,2,4,4-tetramethylpentane	0	6	0	0	0	0	0	6

2.2.2 Benson vs. “methyl repulsion” approach for other branched alkanes

Although a comprehensive comparison between the Benson and “methyl repulsion” methods is beyond the scope of this work, a comparison of a few cases where the “methyl repulsion” method did better than the original 1976 Benson approach is warranted, as such cases provided a key original motivation for the use of the “methyl repulsion” approach.

Table 7 shows estimates for the 16 branched alkanes considered in Appendix C of Domalski and Hearing¹², where they argued that the “methyl repulsion” approach outperformed Benson’s 1976 parametrization in accuracy.

Table 7. Comparison of heat of formation estimates for 16 branched alkanes. (Values reported as A/B when the “methyl repulsion” counting approaches A and B differ; Table 6 includes the differing group counts for these cases)

Molecule	Standard heat of formation at 298.15 K (kcal/mol)				Error (experiment – estimate) (kcal/mol)		
	experiment ¹²	Benson (1976)	Benson (1992)	“methyl repulsion”	Benson (1976)	Benson (1992)	“methyl repulsion”
2-methylpropane	-32.1	-32.5	-32.4	-32.2	0.4	0.3	0.1
2,2-dimethylpropane	-40.2	-40.3	-40.1	-40.2	0.1	-0.1	0.0
2-methylbutane	-36.7	-36.6	-36.6	-36.6	-0.1	-0.1	-0.1
3-methylpentane	-41.1	-40.8	-40.8	-41.0	-0.4	-0.3	-0.1

2,2-dimethylbutane	-43.5	-43.6	-43.5	-44.0	0.1	0.0	0.5
2,3-dimethylbutane	-42.6	-43.0	-42.4	-43.1	0.4	-0.2	0.5
3-ethylpentane	-45.3	-44.9	-45.0	-45.4	-0.4	-0.3	0.1
2,2,3-trimethylbutane	-48.9	-48.4	-48.5	-48.6/-48.3	-0.5	-0.4	-0.3/-0.5
3,3-diethylpentane	-55.5	-53.6	-53.7	-55.5	-1.9	-1.8	0.0
2,2,3,3-tetramethylbutane	-53.9	-55.4	-53.8	-52.3	1.5	-0.1	-1.6
2,2,3,3-tetramethylpentane	-56.6	-58.7	-57.2	-57.1	2.2	0.6	0.5
2,2,3,4-tetramethylpentane	-56.6	-58.2	-56.9	-59.5/-59.2	1.6	0.3	2.9/2.5
2,3,3,4-tetramethylpentane	-56.4	-59.7	-56.9	-59.6/-59.2	3.3	0.5	3.2/2.7
2,2,4,4-tetramethylpentane	-57.8	-58.9	-58.8	-62.9/-57.3	1.2	1.0	5.1/-0.5
3,3-dimethylpentane	-48.1	-47.0	-46.9	-47.9	-1.1	-1.2	-0.2
3-ethyl-3-methylpentane	-51.4	-50.3	-50.3	-51.7	-1.1	-1.1	0.3
Root-mean-squared error					1.33	0.72	1.74/1.06

The table shows that four of the five largest errors from the 1976 Benson approach (2,2,3,3-tetramethylbutane, 2,2,3,3-tetramethylpentane, 2,2,3,4-tetramethylpentane, and 2,3,3,4-tetramethylpentane) are significantly mitigated when the updated group values and gauche counting scheme of the 1992 Benson approach are used. In fact, the root-mean-squared deviation between estimate and experimental values for these 16 molecules is lower with the 1992 Benson approach than with the “methyl repulsion” approaches. Even so, the “methyl repulsion” approach still outperforms the 1992 Benson approach in several cases, most notably 3,3-diethylpentane, 3,3-dimethylpentane, and 3-ethyl-3-methylpentane. The fact that the “methyl repulsion” approach outperforms the Benson approach in these cases with ethyl branching suggests that the general principle of treating methyl differently than larger ligands (e.g. ethyl) has a solid empirical basis. As mentioned earlier, further improvement on the “methyl repulsion” approach may be possible by reparametrizing with the inclusion of 1,5-repulsion effects.

Although further study is needed, the results for HMN and other branched molecules discussed here provide evidence that the Benson method (with the updated 1992 group values and gauche counting scheme) is at least comparable in accuracy to “methyl repulsion” (B) for estimating the heat of formation of branched alkanes.

Additionally, based on these findings, we suggest that when using the “methyl repulsion” method, care should be taken to implement appropriate methyl corrections, using an approach like counting scheme B. Alternatively, with counting scheme A, a refinement to the “methyl repulsion” method that considers 1,5-corrections may be appropriate.

2.3 Implementation of connectivity-based steric corrections in RMG

In order to account for the destabilizing effects of branching in RMG, the Benson approach to steric effects has been implemented in RMG in “15_” and “Gauche_” group databases to

supplement the existing “Group_”, “Radical_” and “Ring_” group databases. Some of the practical considerations associated with this implementation are discussed here.

Firstly, the gauche and 1,5 corrections are only applied to acyclic compounds, as these interactions will not be correctly counted in strained (cyclic) molecules and also, some of the interactions may already be taken into account in the ring-corrections. In any case, alternative approaches newly implemented in RMG, as discussed in later chapters, are suitable for cyclic molecule thermochemistry estimation.

Also, it has been noted that consistency between group values is important,³² that is, all group values should be fitted simultaneously, and trying to piece together group values from multiple sources can produce unexpected results or inaccuracies. Investigation suggested that RMG’s existing group values were based on Benson’s 1976 parametrization. Although it was found above that the 1992 parametrization was slightly more reliable for alkanes, applying such corrections to the 1976 values would produce inconsistencies. Therefore, group values and group counting were implemented in RMG based on the 1976 parametrization. (The alternative of reparametrizing all group values is discussed later.)

Additionally, as shown in Table 1, some of the 1976 gauche counting scheme relies on information (*i.e.* trans, gauche, syn, and cis configurations) not available through connectivity-based representation of the molecule. In these cases, we have RMG count based on the most stable configuration (with the lowest count of gauche interactions); in these cases, the counting scheme used also agrees with the 1992 counting scheme. On the other hand, some of the more complicated gauche counting, such as the TTT case which should have 5 gauche interactions, does follow from connectivity and is taken into account in the RMG implementation. The implemented counting scheme for alkanes is shown in Table 8.

Table 8. Gauche counting scheme for alkanes implemented in RMG

C _a atom group	C _b atom group	# gauche interactions
P	P, S, T, or Q	0
S	S	0
S	T	1
S	Q	2
T	T	2*
T	Q	4
Q	Q	6

*TTT will have 5 gauche interactions

Finally, we need to implement these corrections within the atom-centered group framework used by RMG. The 1,5-corrections are readily implemented in RMG's atom-centered framework by using the middle ("3") carbon atom as the group center. The gauche corrections, on the other hand, are basically bond-centered corrections, and mapping these to the atom-centered framework is not as straightforward. The distinction between atom-centered groups and bond-centered groups is visually summarized by example in Figure 5.

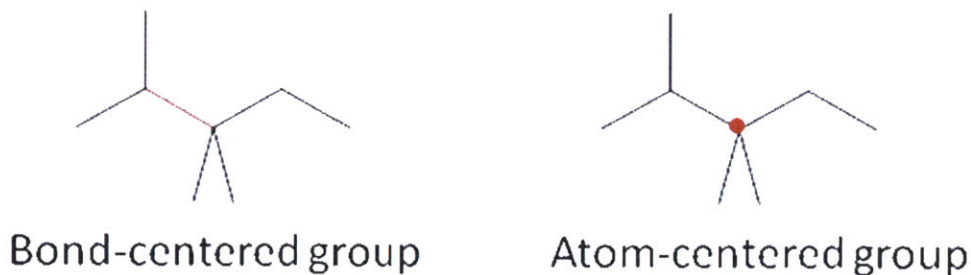


Figure 5. Bond-centered and atom-centered group example

An approach was developed to perform this mapping, and is visually summarized in Figure 6.

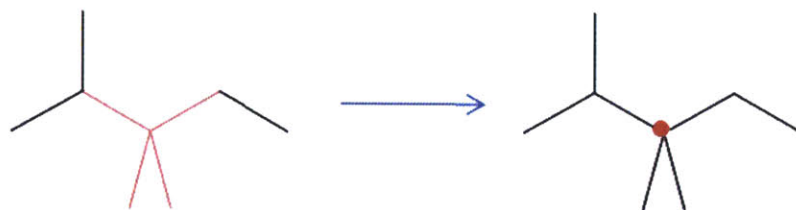


Figure 6. Diagram summarizing the mapping of bond-centered groups to atom-centered groups by combining all bonds values around a given center into a group value for that center

The basic idea of the approach is to create an atom-centered group for each possible combination of bond-centered groups. In the case of alkane gauche corrections, each such bond group will be

counted twice in a given molecule (once for each atom in the bond). Thus, in order to avoid double-counting, the total of all the bond group values for each bond-group combination should be divided by two. Using this approach, the ten* possible alkane bond groups were mapped to 70† different atom-centered gauche groups. It should be possible to extend this approach for mapping bond-centered groups to atom-centered groups for other bond-centered schemes (such as the bond-centered approach for polycyclic aromatic hydrocarbon thermochemistry estimation from Yu *et al.*³³).

Although the previous discussion has focused on alkanes, steric effects due to branching can also be important in alkenes and ethers. Benson's 1976 parametrization includes gauche corrections for alkenes and ethers as well as 1,5-interactions for ethers. For each ether 1,5-interactions, a 3.5 kcal/mol destabilizing correction is added to the enthalpy of formation. Each ether gauche correction is worth 0.5 kcal/mol. Because the ether corrections may be readily localized by using a (atom-centered) group centered on the oxygen, the corrections are not double-counted and the sum of the two bond increments is not divided by two. For alkenes, a destabilizing correction of 0.5 kcal/mol per gauche interaction is used and the counting scheme discussed in Benson *et al.* (1969),¹⁴ (including neglecting the single gauche correction for "secondary" Cs) was applied. As in the ether case, the alkene gauche groups will not be double-counted and the division by two is not applied.

2.4 Concluding remarks

As described above, fast routines for steric corrections for branched hydrocarbons and ethers have been implemented in RMG's atom-centered group framework. There are still opportunities for improvement on the existing framework, however. Firstly, there may be opportunities to refine the groups used in these steric corrections, as discussed previously. Also, there is the opportunity to improve upon the group values. As discussed previously, we have decided to use older Benson parameters based on practical considerations. Ideally, one could

$$* \binom{n+k-1}{k} \text{ with } n=4, k=2 \text{ (4}^{\text{th}} \text{ triangular number)}$$

$$† \sum_{k=0}^n \binom{n+k-1}{k} = \binom{2n}{n} \text{ with } n=4$$

produce a utility that would be able to quickly train a self-consistent and up-to-date set of group values, given an arbitrary group counting scheme and training set. The main limiting step for implementing such an approach is the creation of a curated set of training data that is accurate and comprehensive in terms of both the groups that are covered and the available thermodynamic property data (*i.e.* data is available for enthalpy, entropy, and heat capacity at the necessary temperature). Finally, there is the opportunity to expand the scope of the groups to consider heteroatoms and other types of interactions not considered here. This would require the acquisition of a large amount of data (either from the literature or through computational means) for compounds with the groups of interest. In general, however, the degree of improvement from many of the above approaches is seen to be rather small, and there is, in some sense, a point of diminishing returns. There are likely some areas of “low-hanging fruit” (which could become apparent from further use of these methods and application to new areas) that could provide opportunities for significant improvement for connectivity-based approaches.

Taking a step back from focusing on connectivity-based approaches, we now consider the problem of thermochemistry estimation more generally. As the examples above demonstrate, group additivity can produce quite reliable estimates in many cases; however, in some cases, the approach is ill-suited to the task at hand. Two noteworthy limitations of group additivity procedure are: 1) it relies on a connectivity based molecule representation and 2) it requires availability of accurate group values to cover molecules of interest.

Although the fact that group additivity relies on a connectivity-based representation of a molecule may also be interpreted as a strength of the method, it also limits the application to structures without significant steric effects (outside the previously discussed branching considerations); in particular, group additivity will produce quite inaccurate results for cyclic species, unless *ad hoc* ring corrections are applied. Appropriate corrections, are not, in general, available, and accurate application of group additivity approaches is limited to simple (e.g. cyclopentane) or well-studied (e.g. adamantane) ring structures.

The fact that group additivity requires accurate pre-computed group values is not typically a significant issue for the well-studied classes of molecules involving only carbon, hydrogen, and/or oxygen. However, as soon as heteroatoms (e.g. sulfur, silicon, nitrogen, and halogens) are added to the set of target molecules, the limited availability of appropriate group values can become a significant issue.

An approach that, to a large extent, bypasses these two limitations is to generate explicit three-dimensional molecular structures and perform on-the-fly force field or quantum mechanics calculations on them to obtain the desired thermodynamic quantities.

Broadbelt and co-workers first introduced this approach in 1994.³⁴ However, use of their approach has been limited by factors including robustness and speed.³⁵ We have developed a next-generation system within RMG that attempts to circumvent some of these limitations. In addition to making extensive use of open-source tools, the approach takes advantage of recent developments from several fields, including progress in three-dimensional geometry embedding, force fields, and chemical structure representation, along with enhanced robustness of quantum chemistry codes, and improvements in computer speed and disk storage capabilities. This approach will be more completely described in later chapters.

It should be noted, however, that this is not the only approach that can be imagined for addressing the limitations of connectivity-based approaches. For example, the shortcomings of these methods for cyclic molecules may be partially addressed by expanding the scope of the *ad hoc* ring corrections to encompass all the ring templates needed for a particular set of molecules. However, the set of such molecules that will be considered, will not, in general, be known in advance, particularly in the context of automated reaction mechanism generation. Also, this somewhat defeats the purpose of these group-based approaches, which is to extrapolate to new compounds based on known values for reference compounds. In the existing implementation, the ability to extrapolate for such ring compounds is fairly limited (including only cases with acyclic ligands bound to the ring template), such that the amount of work to obtain the ring correction is likely to be comparable to the effort to obtain an accurate estimate for the compound of interest through other means. Even so, opportunities for expanding the scope of extrapolation may be possible by adjusting the implementation of ring corrections. In the existing approach, a maximum of one ring correction is applied to a given molecule, regardless of the number of rings that are present (though some ring corrections, such as adamantane, may encompass multiple cycles). One can imagine an alternative approach wherein multiple ring corrections can be applied to the same molecule. This approach is likely to work fairly well for cases where the rings are largely independent (*e.g.* cyclopentylcyclopentane), but its performance for fused rings (*e.g.* bicyclo[3.3.0]octane) is likely to produce less reliable estimates. Such an approach still has a good chance at being more reliable than the current group-additivity approach, and doesn't

suffer from the speed limitations of the alternative (“QMTP”) approach to be discussed; Nick Vandewiele of Ghent University is currently investigating the feasibility of this approach.³⁶

3 Chapter 3: Generation of explicit 3D structures from connectivity representations

3.1 Introduction

A critical part of the process of performing on-the-fly quantum mechanics or force field calculations is getting some reasonable initial guess for the three-dimensional structure. Converting connectivity representations to three-dimensional structures has been the subject of considerable investigation, often in the context of studying protein or pharmaceutical compound structure. A number of different approaches exist, and many of these have been considered in this investigation for use in automated reaction mechanism generation.

3.2 Background

Three-dimensional structure generation procedures tend to fall into two general categories: rule-based approaches and numerical approaches.³⁷

Rule-based approaches make use of heuristics and pre-computed ring templates to construct an initial guess for the geometry. In such an approach, chemical structure knowledge of experts is implemented as rules within a computer program; the program may then be used to analyze a compound's connectivity, including factors such as the presence of functional groups and rings, and return a three-dimensional structure. Such methods are relatively fast but are limited in terms of their ability to handle a wide range of species. The program will only handle species for which rules have been programmed.³⁸ This can limit, for instance, the types of rings that may be handled by the program. Examples of codes that use these approaches include Frog³⁸ and CORINA³⁹.

An alternative to rule-based approaches is to use distance geometry, a numerical technique. Examples of programs implementing distance geometry methods include Key3D,⁴⁰ Balloon,⁴¹ RDKit,⁴² and smi23d.⁴³ Distance geometry refers to the use of specified distance bounds between points to "embed" those points in n -dimensional space.⁴⁴ In the context of chemical structure conversion, distance geometry methods apply this technique to atoms in a molecule in order to produce reasonable atom coordinates in three dimensional space. This embedding is often followed by an energy minimization using a force field to refine the

structure. One of the main advantages of these methods is their flexibility; they do not, for instance, need to be preprogrammed with ring structures.

There are several distinct distance geometry approaches. Partial metrization is one of the more common methods for chemical structure conversion. Partial metrization refers to smoothing of some of the distance bounds using the triangle inequality.⁴⁵ Various other metrization schemes also exist.⁴⁵

A relatively recent distance geometry variant is the Stochastic Proximity Embedding (SPE) algorithm introduced by Agrafiotis.⁴⁶ The general idea of the algorithm is to repeatedly choose two points at random and update their distance to more closely match the desired distance. This algorithm has been applied to chemical structure conversion (and conformational sampling)⁴⁷ and has been found to be faster and more effective at probing conformational space than typical distance geometry techniques.⁴⁸ However, this approach is protected by patent⁴⁹ and no tools that implement this approach are presently publicly available (though there have been indications that such a tool might be forthcoming⁵⁰).

3.3 Preliminary screening

Due to the large amount of existing research in this area and the complexities of implementing a fast and robust structure conversion utility, it was decided that the most straightforward approach would be integration of an existing third-party tool. In addition to the method that is used, there were other factors to consider, such as availability (including platform/operating system considerations) and robust handling of various molecular species that can arise in automated reaction mechanism generation, including radicals. A number of software packages were considered, and these are summarized in Table 9.

Table 9. Software packages considered for generating initial three-dimensional molecular structure

Name	URL
RDKit	http://rdkit.org
Marvin/Clean3D	http://www.chemaxon.com/marvin/index.html
BUILD3D	http://sourceforge.net/projects/xdrawchem/files/build3d/
Balloon	http://web.abo.fi/~mivainio/balloon/index.php
smi23d	http://www.chembiogrid.org/cheminfo/smi23d/

FROG (Free Online druG 3D conformation generator)	http://bioserv.rpbs.jussieu.fr/Frog.html
Vconf	http://www.verachem.com/vconf.html
DGEOM95	http://www.ccl.net/chemistry/resources/messages/1995/03/09.008-dir/index.html
Rubicon	http://www.daylight.com/dayhtml/doc/rubicon/index.html
Key3D	http://www.immd.co.jp/en/product_2.html
PRODRG	http://davapc1.bioch.dundee.ac.uk/programs/prodrg/
Converter	http://www.accelrys.com/products/insight/protein_modules.html
Corina	http://www.molecular-networks.com/products/corina
Concord	http://www.tripos.com/index.php?family=modules,SimplePage,sybyl_concord
ICM	http://www.molsoft.com/chemistry.html
HyperChem	http://www.hyper.com/
Omega	http://www.eyesopen.com/products/applications/omega.html
dgsol	http://www-unix.mcs.anl.gov/~more/dgsol/

In addition to the utilities in the table above, some additional utilities were found in the literature, but were not found to be readily available; these include: MOLBUL,⁵¹ ChemDBS-3D,⁵² WIZARD,^{37, 51} COBRA,^{37, 51-53} AIMB,^{37, 51} MOLGEO,^{51, 53}, STRFIT,^{51, 54} Alchemy 2000,⁵⁵ Chem-X,^{37, 53} and Alcogen.⁵³ Of the software utilities investigated, RDKit and smi23d were the most promising on the basis of method and availability, as they both provided implementations of the distance geometry method and both were open-source codes under active development at the time of the evaluation. However, testing showed that smi23d suffered from limitations in radical handling capabilities; in particular, the version tested added hydrogens to radical sites to form saturated three-dimensional structures. So, ultimately, *RDKit* was chosen for its ability to handle radicals, its open-source nature, and its use of flexible distance geometry methods. (It should be noted that since the time of this evaluation, the OpenBabel cheminformatics toolkit⁵⁶ has been endowed with three-dimensional structure creation capabilities via the "--gen3d"

option; although not considered here, this feature could also prove useful for generating three-dimensional structures in the context of automated reaction mechanism generation.)

3.4 *RDKit* testing

Performance tests of *RDKit* gave generally quite satisfactory results. *RDKit* was tested with several thousand species from the PrIME database. The tests suggested that the program worked as expected for most cases, including some “tricky” cases for which a rule-based approach might encounter difficulty. For example, the *RDKit*-based approach produces the desired structure for a challenging helical polycyclic aromatic hydrocarbon, as shown in Figure 7, whereas a typical rule-based approach would produce a nonsensical planar geometry for this case.

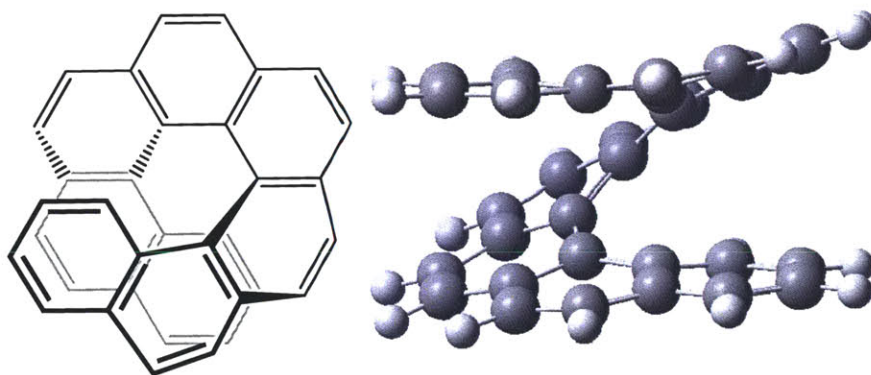


Figure 7. Three dimensional structure for a helical polycyclic aromatic hydrocarbon (octahelicene) obtained with *RDKit*-based approach (including UFF refinement)

For problematic cases that arose during this initial testing, bug reports were submitted to Greg Landrum, the primary developer of *RDKit*, via the *RDKit* bug tracker on Sourceforge.net.⁵⁷ Similar issues arose sporadically in later application of *RDKit* in the context of automated reaction mechanism generation and similar bug reports were warranted. In all cases, *RDKit* developer Greg Landrum promptly addressed the issues and included the fixes in subsequent (typically quarterly) releases of the *RDKit* program.

More recently, systematic testing of open-source three-dimensional molecular structure generation utilities has been performed by an independent group of researchers,⁵⁸ the researchers concluded that *RDKit* was one of the best tools considered in their analysis, providing *post hoc* support to the choice of this tool for integration with RMG.

3.5 Summary

The availability of explicit three-dimensional molecular structures is a key component of efforts to implement on-the-fly force field or quantum mechanics calculations in automated reaction mechanism generation. A large amount of research has been done in the area of three dimensional structure generation, and distance geometry methods offer a relatively flexible and robust approach to the problem. An investigation uncovered many utilities that are available to generate three-dimensional structures using connectivity representations as inputs. Evaluation of the available utilities found that the open-source cheminformatics toolkit, *RDKit*, met the robustness, accuracy, speed, and availability needs of our intended application, and it was decided that RMG would be interfaced with this utility to quickly and reliably obtain three-dimensional structures for molecular species considered during automated reaction mechanism generation. Such integration is discussed in greater detail in the next chapter.

4 Chapter 4: Design and implementation of explicit-3D-geometry-based on-the-fly species thermochemistry

4.1 Background

The estimation of thermochemical parameters (enthalpy, entropy, and heat capacity) is a key component of automated reaction mechanism generation programs like RMG. Such thermochemical parameters affect reaction equilibrium constants, kinetic parameter estimates, and thermal effects, influencing both the mechanism generation process and the behavior of the final resulting model. The scale of the problem is significant, as parameters must be estimated for tens or hundreds of thousands of chemical species in the process of generating a typical detailed chemical kinetic model with hundreds of species.

Traditionally, automated reaction mechanism generation software relies on a group additivity approach, such as the approach developed by Benson^{10, 13, 15-16}, for estimating required thermodynamic quantities (enthalpy, entropy, and head capacity). This process is very fast and usually quite accurate. However, the approach relies on the availability of appropriate parameters. For many classes of compounds, including acyclic species with C, H, and O, the Benson approach can provide accurate results with a relatively small number of parameters. However, when cyclic species are considered, specialized ring corrections are required; these corrections are not as extensible and *ad hoc* ring corrections specific to a particular polycycle are often required; examples include specific “norbornadiene” and “quadricyclane” ring corrections applied to species with these particular polycyclic structures. Therefore, an approach that bypasses the need for specialized parameters, such as *ad hoc* ring corrections, is highly desirable.

One approach to enhance the generality of the thermodynamics parameter estimation process and reduce the need for specialized parametrization is to utilize explicit three-dimensional molecular structure representations (rather than connectivity-based molecular structure representations). Explicit three-dimensional molecular structures can be used with quantum mechanics or force field calculations to obtain estimates of the desired thermodynamics quantities. A proof-of-principle for on-the-fly quantum calculations in automated reaction mechanism generation was developed by Broadbelt et al. in 1994.³⁴ The system described here extends the general approach of Broadbelt et al., incorporating recent developments in several

fields, to make on-the-fly quantum and force field calculations more suitable for routine use in automated reaction mechanism generation. We here refer to this as the quantum mechanics thermodynamic property (QMTP) system.

4.2 Design and implementation of the QMTP system

4.2.1 Design overview

An overview of the QMTP system is shown in Figure 8. The workflow starts by estimating a three-dimensional molecular structure using RDKit, followed by calls to an outside program to perform quantum mechanics or force field calculations to refine that geometry and compute its enthalpy and vibrational frequencies; the results of the calculation are then read and used to calculate the desired thermodynamic properties using standard statistical mechanical relationships within the framework of the rigid rotor/harmonic oscillator approximation.⁵⁹

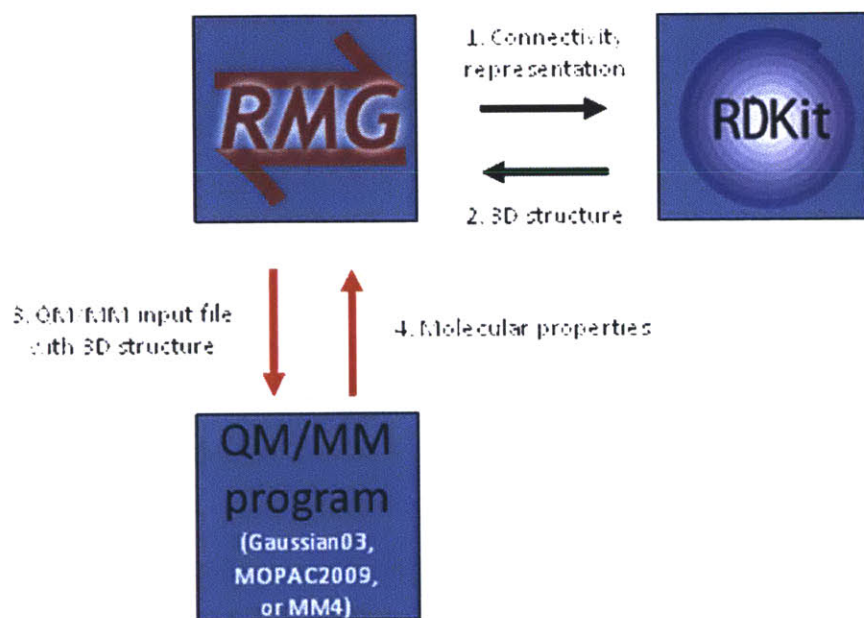


Figure 8. Overview of QMTP system.

The overview shows that the QMTP system leverages the large body of existing work in relevant areas of cheminformatics and computational chemistry. In particular, system was designed to make extensive use of free and open-source modules, when possible, to mitigate licensing issues and avoid limiting operation to one platform or operating system.

4.2.2 Three-dimensional geometry structure generation

As discussed in the previous chapter, *RDKit* was chosen to serve as a third party utility for conversion of connectivity representations into three-dimensional molecular structures due to its ability to handle radicals, its open-source nature, and its use of flexible distance geometry methods. Once the tool for three-dimensional structure generation was chosen, several other aspects of the “0D->3D” conversion process needed to be considered. One issue is obtaining an appropriate reference conformation of the molecule for performing the calculations; ideally, we would use the global minimum energy conformation of the molecule; the force field structure refinement will produce a local minimum, which may or may not also be the global minimum. A number of different approaches exist to search for global energy minima, but the distance geometry methods previously described also offer a simple and straightforward means for conformational exploration.⁴⁵ With such an approach, different random seeds are used to initialize the distance geometry algorithm; after each of the resulting structures are refined with a force field, the structure with the lowest force field energy following force field refinement represents the most stable conformation identified. As the number of random seeds (iterations) increases, it becomes more likely that the global minimum energy conformation will be identified.⁴⁰ Unfortunately, with this approach there is no way to tell how many iterations will be required to reach the global minimum, nor will it be obvious that the global minimum has been obtained once it is enumerated (this issue is shared by most conformational search approaches, with the exception of deterministic global optimization using branch-and-bound approaches⁶⁰). We have chosen here to use a heuristic for the number of embeddings ($\max[1, 5 \times (N_{\text{atoms}} - 3)]$) that scales linearly with the number of atoms in the molecule; this was chosen to avoid potentially burdensome computational cost associated with exponential scaling, while still accounting for the generally greater conformational flexibility of larger molecules. It is noted that this will not guarantee that the global minimum has been identified, though we expect that this will produce a reasonable low-lying minimum for most of the cases where this approach would be applied (for purposes of automated reaction mechanism generation for decomposition of small- or medium-sized molecules).

Another consideration is the design of the interface between the programs involved in the process. To implement the “0D->3D” structure conversion, a Python script was constructed to provide an interface between RMG and the *RDKit* program. After RMG creates a “two-

dimensional” MDL MOL file with the appropriate connectivity, *RDKit* is instructed to read the structure into memory, perform the three-dimensional embedding (a number of times, as discussed above), refine the three-dimensional coordinates using the UFF force field, and write out a MOL file with the three-dimensional coordinates for the UFF-refined structure with the lowest UFF energy, along with a MOL file for the corresponding unrefined structure (the use of this unrefined structure will become apparent in the next section).

4.2.3 Format conversion and output parsing

The OpenBabel command-line utility⁵⁶ is used extensively throughout the QMTP system to create input files and convert between various formats. Python code based on version 1.0 of the cclib libraries⁶¹ is used to parse output files.

4.2.4 Calculation method

With the exception of the MM4 method discussed below, the initial incarnation of the system has been designed around using the semi-empirical PM3 method⁶², as implemented in Gaussian03²⁶ and MOPAC2009⁶³. The PM3 method was chosen for its relative speed compared to other quantum mechanics-based calculations, for its accuracy relative to similar semi-empirical methods (e.g. AM1), and for its wide availability in popular computer codes. Despite the focus on this method for the initial implementation, and for the discussion in this paper, it is noted that the framework described here can be readily extended to use additional methods and/or quantum mechanics programs. It is expected that as computer hardware improves and as electronic structure calculation/force field methods are further developed, it will become practical to perform on-the-fly calculations using more accurate and robust methods with the QMTP system during automated reaction mechanism generation.

4.2.5 Failure checking and recovery

Although the codes used to perform quantum mechanics or force field calculations are relatively robust, they are not error proof. When calculations of this nature are performed manually, troubleshooting is often required. We have incorporated automated troubleshooting into the design for the QMTP system, since to automatically generate large kinetic models, QMTP must successfully return a sensible estimate of the thermochemistry for every molecule considered.

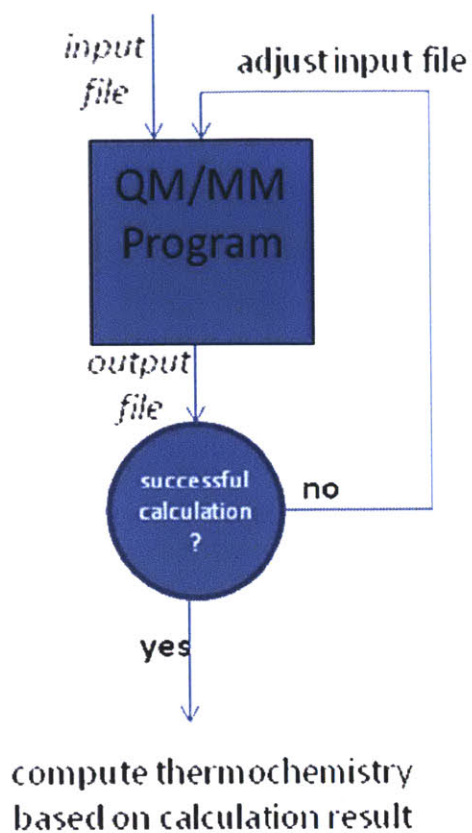


Figure 9. Overview of the automated troubleshooting implemented in the QMTP system

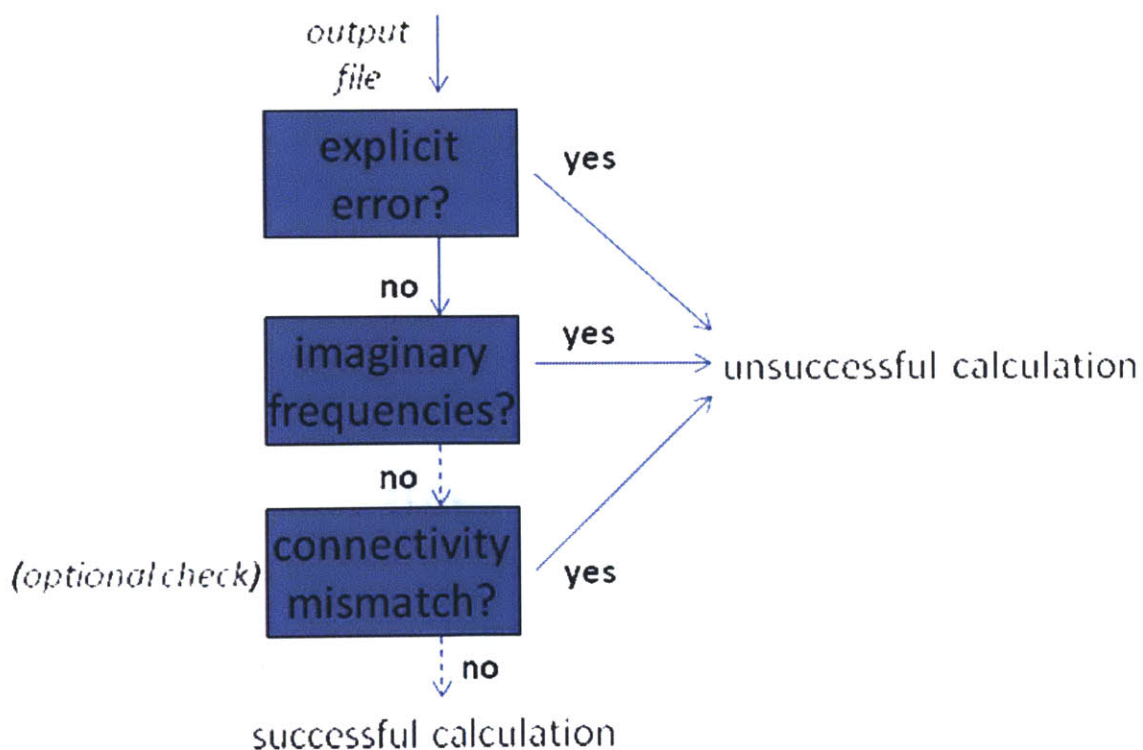


Figure 10. Overview of calculation failure checking in the QMTP system

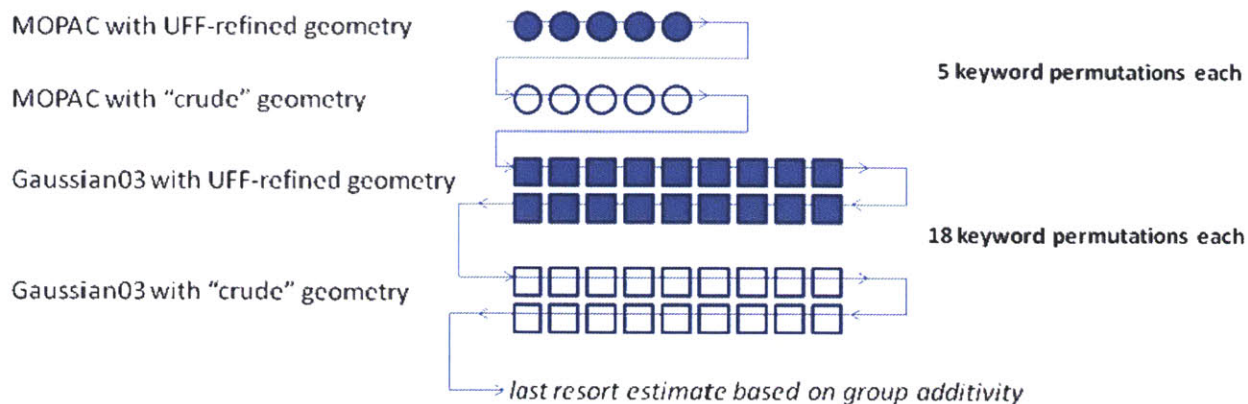


Figure 11. Sequence of input file adjustments currently implemented in QMTP system for PM3 calculations

The QMTP system employs several approaches for checking and dealing with calculation failures of various types. The QMTP system has been programmed with a list of alternative inputs to provide to the quantum mechanics/force field program for cases in which the initial default inputs do not produce a successful result; this is depicted schematically in Figure 9.

For each calculation, the QMTP system will check whether the calculation has completed without an obvious (explicit) error from the called program. Assuming this passes, the QMTP system will also confirm that there are no imaginary frequencies, which would indicate that a saddle point (rather than a minimum) had been obtained. If requested by the user, the QMTP system can also check that the apparent connectivity of the optimized structure matches the desired connectivity (this is discussed further below). These checks are summarized schematically in Figure 10. If all these checks pass, the key results from the output file are parsed and used to compute the desired thermochemical properties. On the other hand, if any of these checks fail, the next input combination in the list will be attempted. The process is repeated until a successful result is obtained or all of the programmed input combinations have been attempted. (In the latter case, the QMTP system will print a warning to the user and fall back to the conventional group-additivity based estimates for the molecule.)

The different input combinations include variations on keywords to adjust the geometry optimization algorithm, change the how the Hessian is guessed or updated, or change the self-consistent field (SCF) algorithm; other variations in input include attempts to use initial guess geometry, or even try an alternative program implementing the same method. This is schematically depicted in Figure 11. The diagram shows how the QMTP system will start with MOPAC, trying five different keyword combinations using the UFF refined geometry. If all these five attempts fail, the QMTP system will retry the same keyword combinations using the

“crude”, unrefined geometry from *RDKit*. If these five new attempts also fail, the QMTP system will move on to trying 18 different keyword combinations with Gaussian03, first with the UFF-refined geometry, and then with the unrefined geometry, should the initial 18 attempts fail. The result is up to 46 different input combinations that are attempted for each molecule.

The different input combinations used have been designed with the goal of increasing the robustness of the system and avoiding the need to fall back to alternative, less reliable estimation procedures. In particular, the default list of keyword combinations that have been programmed into the QMTP system has been developed by manually investigating cases that fail all earlier attempts and trying to find a combination of keywords that would produce a successful result. In the process of investigating such cases, it was also found that the UFF-refined geometry was sometimes unreasonable in some fashion (e.g. by the introduction of a spurious symmetry to the molecule); however, the corresponding “crude”, unrefined coordinates from *RDKit*’s distance geometry routine were more reasonable and readily processed using the default list of keywords. Thus, in addition to storing the UFF-refined coordinates, the QMTP system also stores the corresponding crude, unrefined coordinates. These coordinates may then be used to provide the initial guess geometry, should the initial attempts with the UFF-refined geometry prove to be unsuccessful.

One might imagine the possibility for more sophisticated troubleshooting approaches that attempt to diagnose the actual problem with failed jobs and adapt the input keywords accordingly, or do extensive cross-checking between thermochemistry computed using different methods. Although this might reduce the time needed to obtain a successful result in many cases, the extra layers of complication in developing the appropriate checks and logic were outside the scope of the present study.

4.2.6 Storage of results and use of modified InChI and InChIKey

Advances in disk storage capabilities have made it practical to store a library of calculation input and output files for hundreds of thousands of molecular species. As the time to read to the calculation results is much smaller than the time to actually perform the calculation, it is efficient to store the calculation results between runs so that they may be reused without time-consuming recalculation. To do this, a unique and consistent shorthand notation for each molecule is desired so that results for a particular molecule will be recognized from run to run,

regardless of atom numbering differences or other representation discrepancies between the runs. Recent developments in chemical structure representation, have produced such a unique chemical identifier, known as the InChI (International Chemical Identifier).⁶⁴ We have found that the InChI (with some adjustments described below) is well-suited for labeling of stored quantum chemistry results (at least to the extent that atomic connectivity is unambiguous), as it does not encode electron position or bond type, so, for example, different resonance forms are “correctly” represented by the same InChI. However, as the InChI can be quite long and includes characters such as “/”, it is not well-suited for file names for the stored results. Even so, a hash of the InChI, known as the InChIKey, has more desirable properties for filenames, being only 25 characters in length and consisting only of upper-case letters and a single hyphen (“-“).⁶⁵ Consequently, input and output files for calculations performed by the QMTP system are named using a version of the InChIKey, so they may be readily retrieved during later runs, obviating the need to re-perform the calculation. The stored results are kept in a folder that may be preserved between runs so that a library of hundreds of thousands of results can be accumulated.

As there is a non-zero (though very small) chance for the InChI strings for two molecules in the library mapping to the same InChIKey (an “InChIKey collision”), the QMTP system places the InChI as the molecule name in the input file for the calculation so that it can be checked.

As alluded to above, a couple of special considerations are applied to the InChI/InChIKey strings. Firstly, it is noted that the recently-introduced “standard” InChI employs an option that does not localize tautomeric hydrogen atoms in certain species, and is therefore not well suited to representing species in the gas phase. Consequently, we use a “non-standard” InChI with the “FixedH” layer that localizes such tautomeric hydrogen atoms. Additionally, no versions of InChI that we are aware of currently allow for representing the electron spin multiplicity of the species. In order to represent this, we here employ a modified version of the InChI/InChIKey wherein an additional “mult n ” layer/string is appended to the InChI/InChIKey for species with multiplicity of three or higher. This is best illustrated by example; as shown below, CH₂ singlet and CH₂ triplet would be represented by InChI and InChIKey as shown in Table 10.

Table 10. Illustrative examples of modified InChI/InChIKey system to distinguish molecules with different electron spin multiplicities.

	Modified InChI	Modified InChIKey (v1.02beta)

CH2 (singlet)	InChI=1/CH2/h1H2	HZVOZRGWRWCICA-UHFFFAOYAZ
CH2 (triplet)	InChI=1/CH2/h1H2/mult3	HZVOZRGWRWCICA-UHFFFAOYAZmult3

4.2.7 Connectivity checking

Ideally, the initial guess geometry and optimized geometry will correspond to the intended target molecule. However, it is recognized that there is no inherent guarantee for this in the process described above. It can be imagined that in certain cases, the optimizer will converge to a potential energy surface minimum corresponding to an entirely different molecule than the intended species. The consequence would be non-representative (and thus unreliable and likely inaccurate) thermochemical properties being computed for the molecule of interest. In the QMTP system, we have implemented optional (though recommended) connectivity checking safeguard features to guard against this. The options associated with this feature and process is described in further detail below.

When the “CheckConnectivity” option is set to “confirm”, the QMTP system will attempt to perceive connectivity (using methods described in greater detail below) for calculation results that are otherwise successful; when there appears to be a connectivity mismatch, the calculation will be treated as a failure, similar to jobs with imaginary frequencies or other sorts of errors, and input is adjusted until either a successful result, including an apparent connectivity match with the target molecule, is obtained or all of the various attempt options have been exhausted (in which case, the QMTP system will fall back to the group additivity estimation approach). This is shown schematically in the optional portion of Figure 10. Connectivity checking may also be limited to only providing a warning to the user (without retrying with alternative keywords, geometries, etc.) by setting the “CheckConnectivity” option to “check” or turned off completely (no attempt at connectivity perception) with the “off” option.

Connectivity checking has been implemented with two alternative approaches to connectivity perception. The primary approach is to use the OpenBabel’s⁵⁶ built-in connectivity perception algorithm on the final optimized geometry. This produces a .MOL file with connectivity information for the optimized geometry. For added robustness and greater flexibility (as well as for the ability to process MM4 results), a backup connectivity perception approach (which is used in cases where the primary check fails due to OpenBabel crash or due to

connectivity mismatch) has been implemented via our MoleCoor utility.⁶⁶ The MoleCoor utility perceives connectivity using an algorithm described elsewhere.⁶⁷ The idea of the algorithm is to consider two atoms to be bonded when the distance between them falls below the sum of their covalent radii plus a pre-set tolerance, while also being above a minimum threshold (0.40 Å). (The lower bound prevents atoms that are abnormally close to each other from being considered to be bonded.) In our approach, the tolerance has been set somewhat loosely at 0.50 Å to accommodate apparent O-O bonding in a strained peroxide test case (apparently not recognized by OpenBabel's algorithm); the edge case is depicted in Figure 12. This backup connectivity perception process also produces a .MOL file with connectivity information.

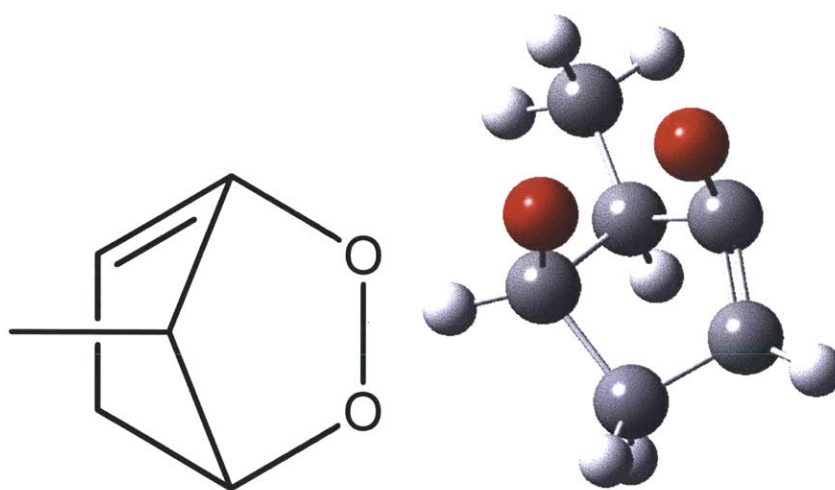


Figure 12. Edge case with strained O-O bond considered in determining appropriate threshold for connectivity perception; the double bond is a key source of the strain; the O-O bond length is approximately 1.84 Å, necessitating a threshold of at least 0.48 Å for the bond to be recognized using the algorithm considered here

In either case, processing this .MOL file through the InChI utility produces an InChI string, which, following removal of stereochemical layers, may be directly compared to the InChI of the target molecule based on RMG's internal connectivity-based representation of the species. An InChI mismatch corresponds to an apparent connectivity mismatch, while an InChI match corresponds to an apparent connectivity match. When operating in the "confirm" mode, in cases where both the primary and backup checks fail, the calculation result is assessed to be a failure, and fallback procedures are followed as described previously.

An example illustrating how this connectivity checking process can help to ensure reliable thermochemical parameter estimation, while maintaining robustness, is shown in Figure 13. The figure shows how the initial attempt using the UFF-refined initial guess geometry can converge to a structure that does not correspond to the desired cyclopentyne target molecule. However, in this case, the “crude” initial guess geometry is much more reasonable and will readily converge to a structure corresponding to cyclopentyne.[‡] Without connectivity checking, the results from the incorrect structure would be used, while with `CheckConnectivity=confirm`, the results that are used will be based on the structure with the desired connectivity.

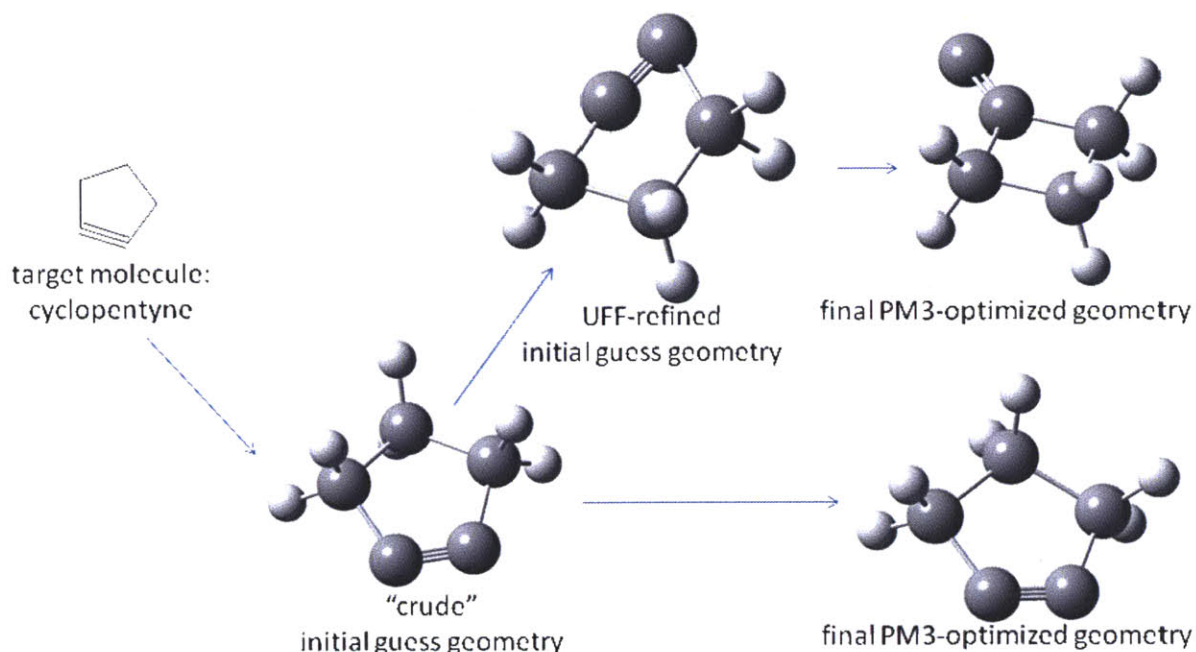


Figure 13. PM3 optimization results for different initial guess geometries for cyclopentyne.

It should be noted, however, that connectivity is not always clear-cut, and ambiguous cases can arise. It is quite possible that connectivity checking approach used here could indicate false positive or false negative matches. Even so, we have made a reasonable attempt to make the connectivity checking process reliable and robust, and our experience thus far suggests that the use of the “`CheckConnectivity=confirm`” option is superior to alternatives.

[‡] In this particular case, the QMTP approach is also able to successfully converge even the UFF-refined geometry to the cyclopentyne target molecule using one of the four backup MOPAC2009 keyword variants.

4.2.8 External symmetry number and chirality calculation

The symmetry number for molecular rotation (also known as the external symmetry number) and chirality are key factors in the entropy of a molecular species, and these contributions can affect the equilibrium constants for reactions by factors often two or greater. Information about the point group of the molecular structure is sufficient to determine both the symmetry number and chirality corrections. Consequently, we make use of the open-source SYMMETRY program⁶⁸ by passing it the optimized three-dimensional geometry as well as a tolerance to allow small deviations from exact symmetry; SYMMETRY then calculates the point group and returns the result to RMG. The symmetry number is determined from the point group based on published relationships.⁵⁹ A chirality contribution of $+R \ln 2$ is included for point groups that lack a superposable mirror image (i.e. point groups lacking σ_h , σ_d , σ_v , and S_n symmetry elements), effectively assuming a racemic mixture of mirror image enantiomers.

4.2.9 Force field and rotor scan capabilities

An interface with the MM4 force field software⁶⁹ has also been implemented in RMG's QMTP system. MM4 force field calculations are much less computationally demanding than semi-empirical electronic structure calculations, and these methods have been found to produce highly reliable thermodynamic property estimates for classes of molecules for which they have been parametrized (e.g. alkanes²³, alkenes⁷⁰, and conjugated hydrocarbons⁷¹).

As alluded to above, the default behavior is to compute thermochemical properties within the framework of the rigid rotor/harmonic oscillator approximation. However, the speed of the MM4 calculations allows us to gather additional information about the potential energy surfaces of molecules with rotors to better account for conformational flexibility without an inordinate amount of added computational cost. In particular, we have implemented rotor scan capabilities when using the MM4 force field. In this approach, a relaxed scan of each rotor is performed in 5° increments; each rotor is considered independently, so that each scan gives a one-dimensional profile of energy and moment of inertia for the mode of interest. Rotor symmetry number is estimated based on connectivity. This information is provided to a Python-language code based on CanTherm v1.0⁷² to compute the desired thermochemical quantities using a separable hindered rotor treatment (and also accounting for variation in the moment of inertia with the rotor angle).

4.2.10 Scope of QMTP calculations

Although on-the-fly PM3 calculations based on explicit three-dimensional geometries are expected to be useful in many cases, it is recognized that there are many cases where alternative approaches, such as the original method based on group additivity, are still desirable. For example, we expect that the traditional group additivity approach to be more reliable (and much faster) than the QMTP approach for acyclic hydrocarbons. Therefore, we have implemented a switch that will further restrict calls to the QMTP estimation routines to cyclic species, for which the traditional group additivity based estimates are expected to be inaccurate. Also, the user may wish to specify “known” thermochemistry parameters for a set of molecular species. Therefore, the QMTP methods will only be called for molecules without user-specified thermodynamic parameters.

Also, it is recognized that many methods, including the previously-discussed PM3 and MM4, may be less accurate and/or less robust for treating radical systems with unpaired electrons. Consequently, we have implemented an option where the user can specify the maximum number of radical sites that a molecule can have for it to be directly processed by the QMTP system. So, if a user sets this option to zero, only closed shell molecules are directly processed by the QMTP system. In other cases (monoradicals, biradicals, etc.) hydrogen bond increments (HBI)⁷³ are applied to QMTP results for the saturated, closed-shell parent molecule. (It is noted that the HBI adjustments applied are the same that are used for conventional group additivity-based estimates for radicals.)

For routine use for systems with only carbon, hydrogen, and oxygen, we currently suggest settings that use QMTP only for cyclic species and closed-shell species (applying HBI corrections for radicals).

It should be noted however, that the approach of using generic HBI corrections does not always achieve a desirable level of accuracy. In particular, it has been found that the use of the generic “C=CCJ=C” radical correction in 1,3-cyclopentadien-2-yl underestimates the HBI by a significant amount, resulting in a result (79.4 kcal/mol) that differs from a CBS-QB3-based result (95.9 kcal/mol) by over 16 kcal/mol. In this case, the error most likely due to the fact that the allenic resonance form of this radical is severely strained by the presence of a ring, in contrast to the unstrained acyclic molecules such as 1,3-butadien-2-yl, which would have been used to determine this particular radical correction. The alternative for these cases would be

either to apply more specific ring HBI corrections, such as those published in the literature for particular cyclic radicals (e.g. Ref. ⁷⁴), or use the QMTP system to directly perform calculations on monoradicals. In the case of 1,3-cyclopentadien-2-yl, the latter approach using PM3 produces an enthalpy of formation at 298.15 K of 86.4 kcal/mol, which is in closer agreement with the CBS-QB3 value, though still deviates by about 9.5 kcal/mol.

4.3 Testing of the QMTP system

In addition to extensive testing during the development of the QMTP system (including development of a library of keywords for failure recovery) several additional tests were performed on the current version of the system, to more formally characterize the accuracy, robustness, and speed of the system, as well as its influence on kinetic models.

4.3.1 Accuracy of estimates

In addition to informal comparisons to ensure accuracy, the QMTP system was tested on a challenging test set of polycyclic hydrocarbons from Osmont et al., who had compiled experimental values for standard enthalpy of formation at 298.15 K for 47 molecules.⁷⁵ The QMTP system (with PM3 and with the MM4 approach with separable hindered rotor treatment) was applied to 43 of these molecules (four of the molecules had two geometric isomers; in these cases, the QMTP result was applied to the isomer with the lower enthalpy of formation.)

Table 11. Accuracy of RMG estimation approaches, using experimental enthalpy of formation values compiled by Osmont *et al.* as reference. Two molecules with three-membered rings failed the attempts with the MM4 approach and were excluded from averaging. N=43, with the exception of the MM4 case.

	Mean abs. error (kcal/mol)	Root-mean-squared error (kcal/mol)
Original RMG group-additivity approach	40.0	55.8
QMTP system with PM3 approach	7.0	10.6
QMTP system with MM4 approach	29.6	53.1

Results of the comparison are shown in Table 11. The results show that the error in the PM3 approach is, on average, more than a factor of five lower than the original RMG approach based on group additivity. However, the MM4 approach, on average, does noticeably worse than

the PM3 approach. A closer analysis of the results from the MM4 approach shows that the largest errors are concentrated in species with three-membered rings, for which the MM4 force field currently lacks parameters.⁷⁶

Table 12. Accuracy of RMG estimation approaches, using experimental enthalpy of formation values compiled by Osmont *et al.* as reference for a subset of molecules without three-membered rings. N=23

	Mean abs. error (kcal/mol)	Root-mean-squared error (kcal/mol)
Original RMG group-additivity approach	22.8	40.9
QMTP system with PM3 approach	8.4	12.9
QMTP system with MM4 approach	1.3	3.0

Comparison on the subset of molecules without three-membered rings produces results that are much more favorable with the MM4 approach, as shown in Table 12. In fact, the MM4 approach significantly outperforms the PM3 approach on this subset.

4.3.2 Effect on kinetic models

To assess the influence of the use of QMTP estimates on an actual detailed chemical kinetic model, we took an existing combustion model, JP-10 combustion mechanism version 0.19⁷⁷ and re-generated thermodynamic properties for the species using two approaches. The first approach (“no QMTP”) is the traditional approach used by RMG with use of group additivity in most cases and with results taken from built-in default and GRI-Mech 3.0 thermodynamic libraries for selected compounds where estimation is not required. The second approach (“QMTP”) relies on the QMTP system described here for cyclic and polycyclic compounds; cyclic radical thermochemistry is computed using hydrogen bond increment corrections to the saturated molecule as computed using the QMTP system, and the CheckConnectivity option is set at “confirm” to ensure that all estimates obtained via QMTP correspond to the desired molecule; as with the first approach, results for selected molecules can be taken from the built-in default and GRI-Mech 3.0 thermodynamic libraries. The mechanism includes a number of polycyclic compounds, including JP-10 itself, for which the estimates obtained by the two approaches are likely to disagree significantly.

Figure 14, below, shows the time to maximum modeled CH concentration (used here as a proxy for ignition delay) versus shock tube results from Davidson et al.⁷⁸ at 45 different conditions.

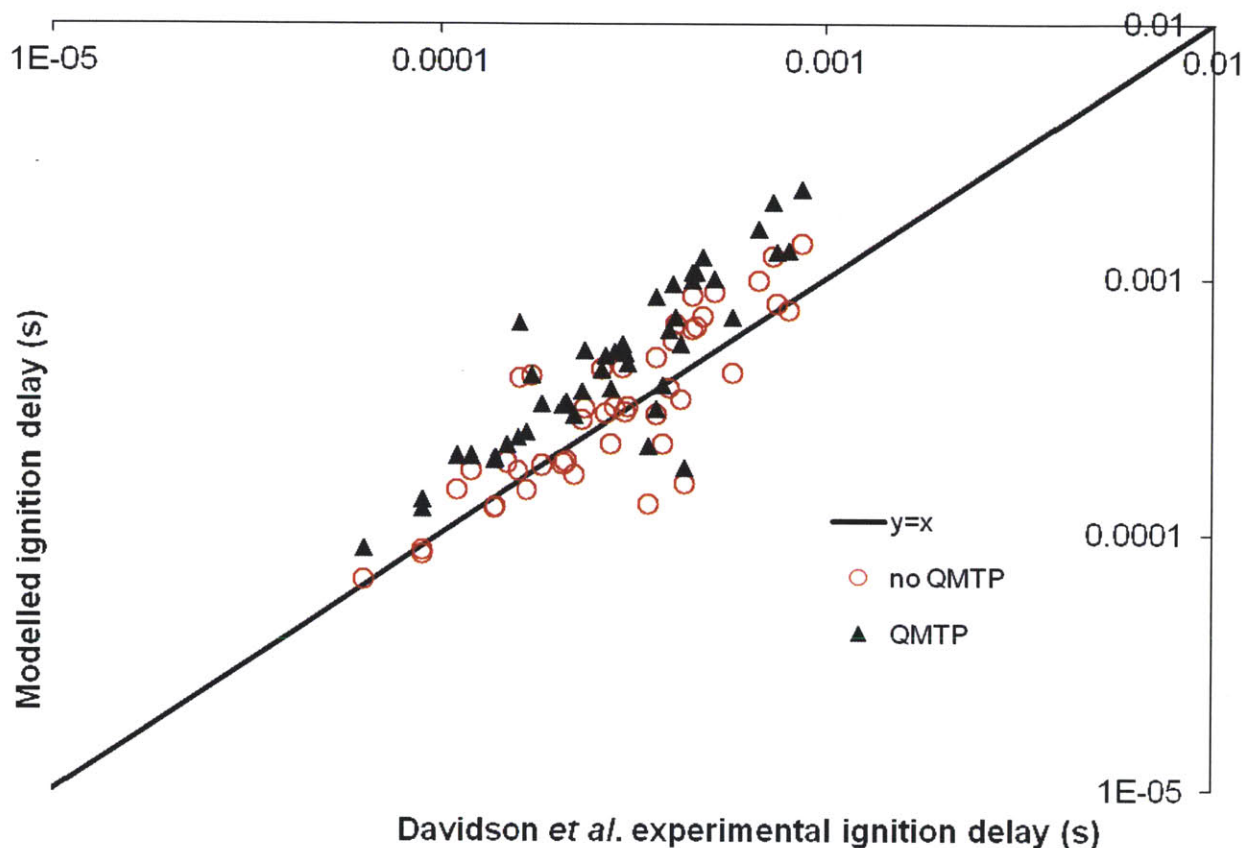


Figure 14. Parity plot showing the time to maximum modeled CH mole fraction (used here as a proxy for the modeled ignition delay) versus experimental ignition delay from shock tube data from Davidson *et al.*⁷⁸

The plot shows that the use of the QMTP thermochemistry (which should, in general, be more reliable) has a significant effect, sometimes changing the time to maximum CH concentration by more than 70%.

It should be emphasized that the underlying mechanism and kinetic parameters are the same in both approaches; the effect of using the QMTP approach should be even more pronounced when considering its use in the context of the entire RMG mechanism generation approach. Because detailed chemical kinetic models generated by RMG using the QMTP approach will involve different thermochemistry estimates for core species and potential intermediates, the actual species and reactions that are important enough to be included in the mechanism will differ. Also, even if the same reaction appears in both models, the kinetic

parameter estimates themselves can be affected through equilibrium and Evans-Polanyi-type considerations (and also potential energy surface effects when obtaining pressure-dependent kinetic estimates via Master Equation calculations).

4.3.3 Robustness and speed

To test the performance of the QMTP system in a mechanism generation setting, RMG was run for 120 hours (5 days) with the QMTP system turned on to use PM3 results for cyclic molecules; the CheckConnectivity option was set to “confirm”[§]. The system that was considered involved high-temperature oxidation of the polycyclic molecule, JP-10.

Over this 120 hour period, RMG used the QMTP system to successfully obtain PM3 results for 22,277 cyclic molecules (22,244 with the primary MOPAC approach and 53 with Gaussian03), in addition to performing traditional mechanism generation functions (e.g. simulating the time evolution of the model, generating reactions, and estimating kinetic parameters). An additional 20 cyclic species were attempted using QMTP but failed all attempts; for all 20 of these cases, the failure of QMTP was due to an apparent connectivity mismatch from the intended species, and without the connectivity checking option turned on, these would have been considered successes. This corresponds to a success rate of greater than 99.9% and demonstrates the robustness of this approach. Graphs for the 20 failures are shown in Figure 15; many of these structures appear to highly strained and there may not exist minima on the PM3 potential energy surface that have the desired connectivity. Among the successful results were 7 molecular species for which results from initial attempts failed the connectivity check but later attempts were able to produce a result with the desired connectivity. An additional 6 cases failed the primary connectivity check, but the backup connectivity check suggested that the molecule had the desired connectivity, and the attempt was considered a success.

[§] In order to avoid an apparent bug specific to the Linux version of Gaussian03 used here, four of the Gaussian03 keyword alternatives involving the use of “opt=calcall” (corresponding to eight of the 36 Gaussian03 attempt options) were removed from the programmed keyword list. It is not expected that the removal of 8 attempts has a significant effect on the results.

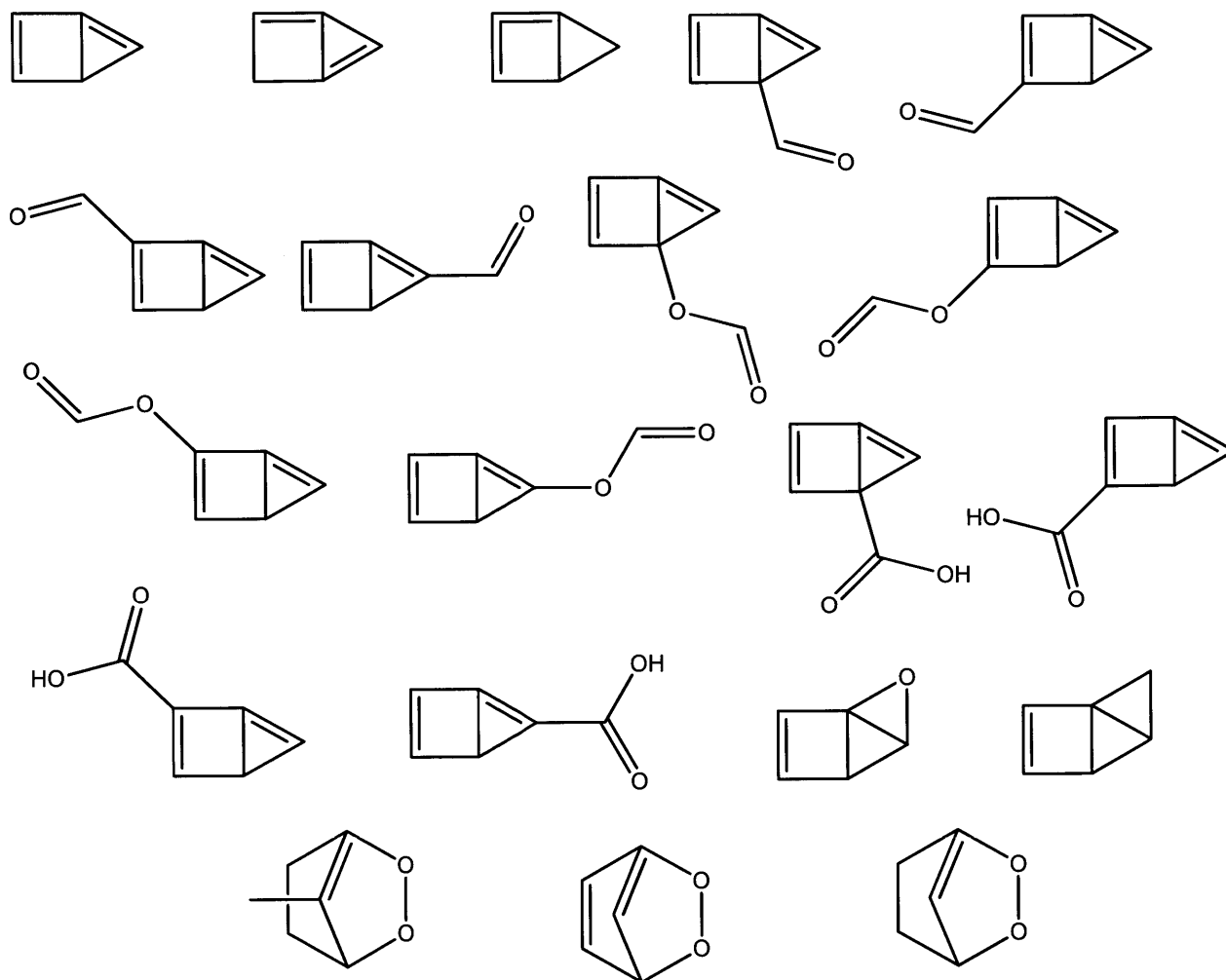


Figure 15. Twenty species that failed all QMTP attempts in JP-10 high-temperature oxidation test run. All failed due to apparent connectivity mismatch between the desired molecule and the PM3-optimized geometry.

Figure 16 shows the number of species (on a logarithmic scale) that first succeeded at each of the attempts. The plot shows that in the vast majority of cases, the first MOPAC attempt was able to produce a successful result. However, the number of cases where the backup attempts proved to be useful still number in the hundreds; the results demonstrate the utility of using the alternative “crude” geometry, as well as the alternative PM3 implementation offered by Gaussian03. Even so, these backup attempts reach a point of diminishing returns, and the graph shows that many of the later attempts did not prove to be useful in this particular test run.

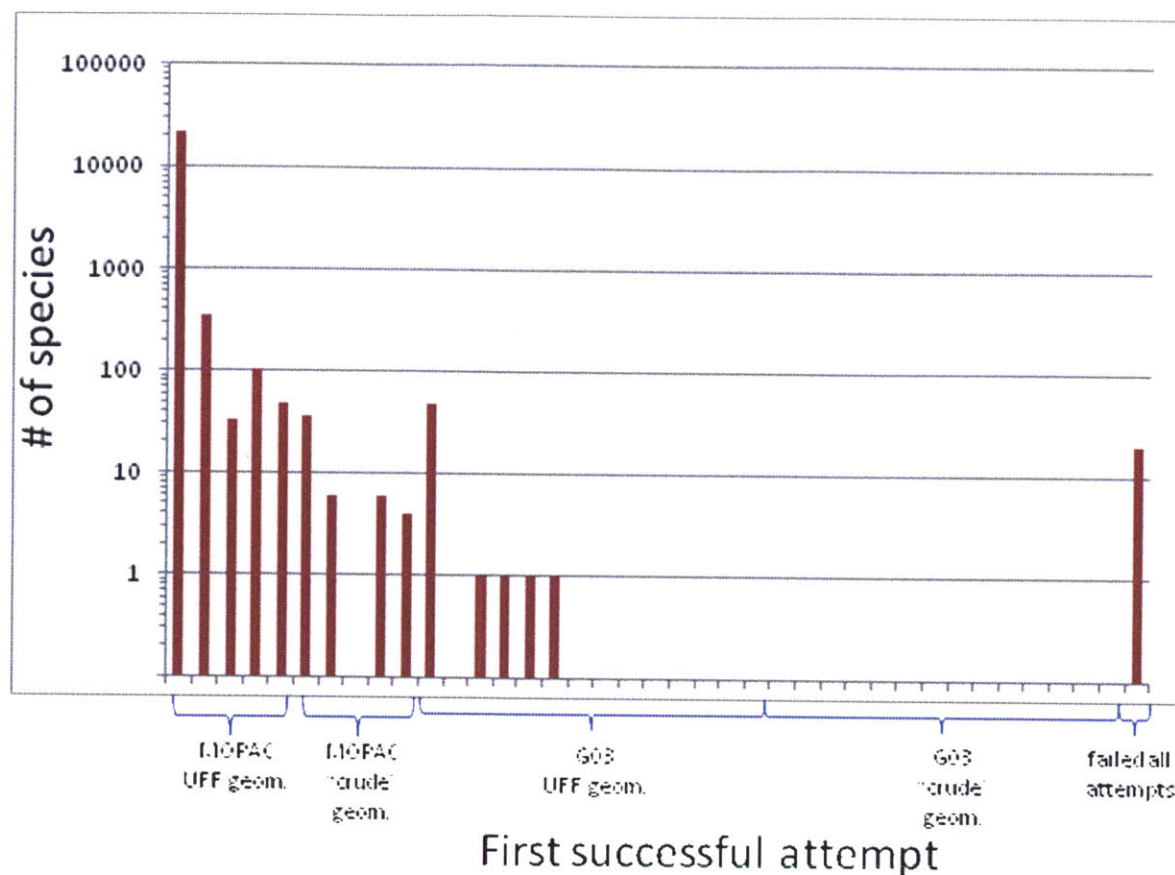


Figure 16. Distribution of outcomes for QMTP system in JP-10 high-temperature oxidation test run. Each bin on the abscissa corresponds to a particular “attempt” by the QMTP system to perform PM3 calculations with various choices of input keywords, initial guess geometries, and PM3 implementations. The first five attempts correspond to the use of the MOPAC implementation of PM3 with the UFF-refined initial guess geometry; the next five use the MOPAC implementation with the “crude” (non-UFF-refined) initial guess geometry; the next 14 correspond to the use of the Gaussian03 implementation of PM3 with UFF-refined geometry; the final 14 attempts correspond to the use of the Gaussian03 implementation with the “crude” geometry.

4.4 Opportunities for improvement and applying the approach to other areas

4.4.1 Selective use of more accurate methods

As shown in Table 12, MM4 is significantly more accurate than PM3 for certain types of molecules. Similarly, in some situations it would be better to run a DFT or high-level quantum chemistry calculation rather than relying on PM3 or MM4. It would be easy to modify QMTP to

use these other types of calculations – the challenge is writing the logic which would decide which type of calculation to perform for each molecule.

4.4.2 Improvements to treating conformational flexibility

As discussed previously, the typical treatment is to assume harmonic behavior for all internal modes; we have implemented rotor scan methods with the MM4 force field that account for conformational flexibility at a higher level (separable hindered rotor treatment), but even this approach may not achieve a sufficient level of accuracy for certain molecules/applications, and some examples of this appear in the literature.⁷⁹ There is opportunity for improvement of this treatment, particularly when using the relatively computationally inexpensive MM4 approach. One of the more promising approaches is the recently introduced multi-structural method of Truhlar and coworkers⁸⁰; this method is reportedly able to account for conformational flexibility via both rotation about bonds as well as other modes such as ring inversion, and obeys desired statistical mechanical limits (e.g. high temperature heat capacity). Implementation of this approach within the QMTP system for use with MM4 calculations would seem to be the next logical step for improving treatment of conformational flexibility. A key aspect of such an implementation would be accurate and robust enumeration of all the minima on the molecule's potential energy surface, without double counting equivalent conformations. Toward this end, we have developed and implemented a novel conformational equivalence algorithm, designed for this particular application, in the MoleCoor utility.⁶⁶ These opportunities are discussed further in Chapter 8.

4.4.3 Polarizability estimates

In addition to being used in estimation of thermodynamic parameters, on-the-fly quantum mechanics calculations can produce estimates of polarizability, with the appropriate keyword choices. Such polarizability estimates can, for example, be incorporated into approaches to estimate transport properties estimation, and, in fact, an RMG script using the QMTP framework has been created and applied to produce polarizability estimates as part of a post-processing transport property estimation step. With additional software development, it should be possible to integrate on-the-fly polarizability estimates directly into RMG's transport property estimation routines.

4.4.4 Solvation property estimation

The availability of explicit three-dimensional molecular structures also presents the opportunity for more accurate treatments of solvation properties. Such properties are highly dependent on volumetric and surface area properties of the molecules; a number of volumetric and surface area metrics have been developed that can be quickly computed based on explicit three-dimensional structures.⁸¹ Alternatively, the three-dimensional molecular structures could serve as the starting point for more advanced (quantum mechanical) treatments of solvation, such as polarizable continuum model (PCM) calculations.⁸² Work on treatment of liquid-phase systems and solvation effect parameter estimation is currently underway through separate RMG development efforts in the Green Group.

4.4.5 Standalone thermodynamic property estimation

Although “black box” *ab initio* methods have enabled reasonably straightforward estimation of thermodynamic properties for a wide range of compounds, the application of these methods requires some initial training and mistakes can easily creep into calculation results. Also, the amount of “human time” associated with input file construction, job monitoring, result analysis, and troubleshooting is non-trivial. As such, a more automated approach for setting up and performing calculations and analyzing the results is highly desirable. The QMTP system described previously is well-suited to this level of automation, as it performs all necessary steps, from generation of an initial guess three-dimensional structure, to construction of calculation input file, to troubleshooting, to processing the calculation results into the desired thermodynamic property estimates, without human intervention. In fact, the QMTP features have recently been incorporated into a standalone thermodynamic property estimation tool distributed with RMG. Though this tool currently uses the same PM3 and MM4 methods as used by RMG, it should be possible to extend this to use other, more accurate and time-consuming *ab initio* methods, such as the popular CBS-QB3²⁵ or Gaussian-*n*⁸³ composite methods.

4.4.6 Kinetic parameter estimation

Currently, the results from the QMTP system affect kinetic parameters somewhat indirectly through the properties of reactants and products. It could be possible to further improve certain kinetic parameter estimates using information that may be obtained from on-the-

fly calculations using explicit three-dimensional geometries for species minima. For example, a scheme for kinetic parameter estimation for intramolecular disproportionation reactions based on ring strain of stable molecules has been proposed by Herbinet *et al.*⁸⁴

One can imagine even more direct means of using on-the-fly calculations with explicit three-dimensional geometries to obtain kinetic parameters. In particular, there are opportunities to adapt the approach to locate first-order saddle points; calculations using saddle-points could be used to produce kinetic parameter estimates within the framework of transition-state theory (TST). If realized, these types of capabilities could be particularly useful for reactions with cyclic transition states (e.g. intramolecular hydrogen abstraction) for which the existing connectivity-based estimates can be unreliable. Such opportunities are discussed further in Chapter 7 and Chapter 8.

4.5 Summary

A system for performing on-the-fly quantum mechanics or force field calculations in the context of automated reaction mechanism generation has been described. Testing has demonstrated the impact, accuracy, and robustness of the system, which make it suitable for routine use during mechanism generation. The system is particularly useful for obtaining more-reliable thermochemical parameters for cyclic species (for which alternative automated group additivity-based approaches are prone to significant error). Possible avenues for further improvements and alternative applications have also been described.

5 Chapter 5: Practical considerations of explicit-3D-geometry-based on-the-fly species thermochemistry

5.1 Introduction

The previous chapter provided a general overview of the design and implementation of the QMTP system as well as some general testing results. This chapter extends the discussion of the QMTP system with focus on various practical considerations and provides additional implementation details.

5.2 Source code and dependencies

The bulk of the QMTP system is implemented in the *QMTP* class within the Java implementation of RMG; created as an analog to the *GATP* group-additivity class, it is located in the *jing.chem* package. A *QMData* class has also been created as a data structure for in-memory storage of calculation results. A *molFile* class has been introduced to allow creation of pointers to MOL files stored on disk. Several *python* scripts (and batch script files for Windows operation) used to interface the *QMTP* Java class with modules are located in the */scripts* folder of the RMG distribution.

5.2.1 Modules

As mentioned in the previous chapter, the QMTP system leverages several modules, some of them developed by third-parties and distributed with permissive licensing. A brief description of each of these, how they are used by the QMTP system, and how they are integrated with RMG follows.

- *cclib*: A modified version of *cclib* v1.0 is used by the QMTP system for parsing desired data from Gaussian, MOPAC, and/or MM4 files. (Further details about modifications are described later.) The *cclib* license allows modification and redistribution with RMG (in the */source/cclib* folder), so no additional action by the user is required to obtain this module.
- *MoleCoor*: *MoleCoor* is a *python* program developed by the author. It is used by the QMTP system for connectivity perception in three-dimensional molecular structures, for

MM4 input file creation, and for parsing of geometries from MM4 output during connectivity checking. It is distributed as a submodule within RMG in the */source/MoleCoor* folder, and no additional action by the user is required to obtain this module.

- **CanTherm:** CanTherm is a *python* package developed within the Green Group. It is used by the QMTP system when using the MM4 feature for calculating desired thermodynamic quantities based on one-dimensional rotor scans; it can also be used to analyze the force constant matrix from MM4 calculations for more reliable frequencies when using harmonic oscillator treatment. CanTherm has been modified by the author to implement desired functionality and this is distributed with RMG in the */source/CanTherm* directory; the modifications are described in greater detail later.
- **OpenBabel:** The OpenBabel package is used by the QMTP system during generation of Gaussian03 and MOPAC2009 input files; it also provides primary connectivity perception functionality. OpenBabel is free, open-source software distributed separately; it must be installed by the user.
- **InChI:** InChI software (v1.02 beta) is needed to determine the InChI and InChIKey for unique representation of molecules, which is important in the file names and file titles used by the QMTP system. InChI may also be used elsewhere in RMG and is not solely a dependency of the QMTP system. InChI software is (freely) distributed separately and the InChI binary must be placed in RMG's */bin* directory.
- **SYMMETRY:** The SYMMETRY package is used for point group calculation. This is an open-source academic code freely available on the Internet. The compiled binary must be placed in RMG's */bin* directory.
- **RDKit:** As discussed previously, RDKit provides the QMTP system with the functionality to convert connectivity representations into explicit three-dimensional molecular structures. RDKit is free, open-source software distributed separately; it must be installed by the user. RDKit also has its own dependencies, including, for example, NumPy and Python.
- **Gaussian03, MM4, and/or MOPAC2009:** As discussed previously, Gaussian03 and MOPAC2009 are used to perform PM3 calculations while the MM4 software package provides capabilities for performing calculations with the MM4 force field. These

programs are distributed separately and must be installed by the user, including the setting of appropriate environment variables. Unlike the other modules, Gaussian03 and MM4 are not freely available. An alternative, which is free for not-for-profit academic use, is MOPAC2009.

5.2.1.1 Modifications to *cclib*

Several modifications were made to *cclib* v1.0. The original version parsed some portions of Gaussian03 output files. Modifications included the addition of molecular mass, rotational symmetry number, and rotational constants from Gaussian03 output. Additionally, functionality for parsing MOPAC output files and MM4 output files was added. In these cases, parsing capabilities were added for atomic numbers, molecular mass, Cartesian coordinates, energy, vibrational frequencies, rotational constants, and the number of atoms. Additionally, steric energy parsing capability was added for the case of MM4 output parsing. Finally, a bug in orbital symmetry parsing and HOMO determination for triplet oxygen atom PM3 results from Gaussian03 was addressed.

5.2.1.2 Modifications to CanTherm

A number of modifications were made to the CanTherm software. Several of these are relevant to the QMTP system. In particular, functionality for working with MM4 cases, including, for example, the ability to read in MM4 rotor scan output and MM4 force constant matrix output, was added. The output was also adjusted to allow easier parsing by the QMTP system in RMG. Additionally, code was added to consider the variation in reduced moment of inertia with rotor angle, and this variation was incorporated into the solution of the Schrodinger equation by fitting the inverse of the reduced moment of inertia to a Fourier series in the dihedral angle. Appendix II includes a derivation of the matrix formulation for the kinetic energy term for the Hamiltonian for this approach with variable moment of inertia.

5.3 RMG input file considerations

Various QMTP options were mentioned in the previous chapter's discussion of the QMTP system. These options are selected using RMG's main input file, *condition.txt*. If the user desires

to use the QMTP system features, they must include a QMTP block in the input file, as in the example below, following the temperature and pressure and preceding the *InitialStatus* block:

```
ThermoMethod: QM Gaussian03
QMForCyclicsOnly: on
MaxRadNumForQM: 0
CheckConnectivity: confirm
```

The second field in the first line chooses the program/method to use. Options include "MOPAC", "Gaussian03", or "both" (the default if this field is omitted). The "both" option requests that RMG first try to use MOPAC and if all MOPAC attempts for a particular species are unsuccessful, then it will try Gaussian03. In all three cases, these will use PM3 calculations. Two other (more experimental) options are "MM4" and "MM4hr"; the former uses MM4 with the rigid-rotor, harmonic oscillator (RRHO) approximation, while the latter does one-dimensional rotor scans to treat rotor modes in a separable manner. The next line determines whether RMG will use the QMTP approach for all species or just cyclic ones. It is often desirable to run with this option on, as the QM calculations can be time-consuming and may be less accurate than group-additivity for acyclic species, whereas they are more likely to improve accuracy for cyclic species. Then, the user must specify the maximum radical number for species that will be fed to the QM/force field program. For molecules with more radicals than the value specified here, QM/force field calculations will be performed on the saturated molecule (with added hydrogens) and hydrogen bond increment (HBI) corrections will be applied. Finally, the user must specify an option for connectivity checking. Options here are "off", "check", and "confirm". As mentioned in the previous chapter, the "check" option only uses connectivity checking to determine whether to print a warning to the user; the "confirm" option uses connectivity checking to determine whether a given calculation should be considered successful and used to compute thermochemistry.

An example condition file using the QMTP system is included in the RMG distribution under */examples/RMG/cyclopropane_QM/*.

5.4 Symmetry and chirality statistical effects

5.4.1 External symmetry number and chirality

As discussed previously, the QMTP system makes use of the SYMMETRY program for point group calculation. Typically, the optimized molecular structure that should have a given symmetry will not have exactly this symmetry due to rounding of the Cartesian coordinates and the fact that the optimizer typically won't converge to the exact optimum. The SYMMETRY program is able to take these deviations from perfect symmetry into account, and the user can specify thresholds that the program uses in the process of testing for symmetry elements. The current implementation of symmetry number estimation is performed first using a primary/initial criterion at the built-in default of 0.05 and a final criterion at 0.02. In a small fraction of cases, this will not result in a recognized point group. In these cases, fallback approaches are applied until a recognized point group is obtained. The first fallbacks apply looser final and/or primary thresholds, while the last resort approach requires exact symmetry (final threshold of 0.0) and will most likely result in identification of the C1 point group. In cases where the last resort approach is used, a warning is printed to the user, indicating that the symmetry may be underestimated. The values used are summarized in Table 13.

Table 13. Thresholds used by QMTP system when invoking SYMMETRY for point group calculation

Attempt #	Primary threshold (built-in default: 0.05)	Final threshold (built-in default: 0.0001)
1	default	0.2
2	default	0.1
3	0.2	0.1
Last resort	default	0.0

The point groups recognized by SYMMETRY and the QMTP system (and the associated chirality and external symmetry number corrections to entropy) are summarized in Table 14.

Table 14. Point groups recognized by SYMMETRY and the QMTP system, and associated chirality and external symmetry number corrections to entropy; $n=2-8$, $m=4, 6$, or 8 ; based partly on point group to symmetry number conversion tables in the literature⁵⁹

Point group	$\frac{\Delta S_{symm}}{R}$	$\frac{\Delta S_{chiral}}{R}$	Point group	$\frac{\Delta S_{symm}}{R}$	$\frac{\Delta S_{chiral}}{R}$
C1	0	+ ln 2	T	- ln 12	+ ln 2
Cs	0	0	Th	- ln 12	0
Ci	0	0	Td	- ln 12	0
C _n	- ln <i>n</i>	+ ln 2	O	- ln 24	+ ln 2
D _n	- ln 2 <i>n</i>	+ ln 2	Oh	- ln 24	0
C _{nv}	- ln <i>n</i>	0	C _{in} fv	0	0
C _{nh}	- ln <i>n</i>	0	D _{in} fh	- ln 2	0
D _{nh}	- ln 2 <i>n</i>	0	I	- ln 60	+ ln 2
D _{nd}	- ln 2 <i>n</i>	0	Ih	- ln 60	0
S _m	- ln $\frac{m}{2}$	0	Kh	0	0

In addition to determining statistical corrections, the point group is also used to determine linearity, which affects which formulas are used to compute rotational contributions to thermodynamic quantities. The C_{∞v} (“C_{in}fv”) and D_{∞h} (“D_{in}fh”) point groups correspond to linear arrangements of atoms.

5.4.1.1 Special considerations for radicals

As discussed in the previous chapter, the currently recommended approach to compute thermochemistry for radicals is to first compute thermochemistry for the molecule with the radical sites saturated with hydrogens and then apply hydrogen bond increments (HBI). We want to correct the entropy for radical species using the external symmetry number for the radical. However, we don't explicitly have this number available as we don't have an explicit optimized molecular structure for the radical. To partially address this complicating factor, the currently-implemented approach is to compute the symmetry number effects due to removal of hydrogens as a perturbation to the three-dimensional structure symmetry number based on the relative

group-additivity symmetry number for saturated and unsaturated (radical) molecules, as shown in the expression below.

$$\sigma_{ext,rad} = \sigma_{ext,sat}^{3D} \frac{\sigma_{tot,rad}^{graph}}{\sigma_{tot,sat}^{graph}} \quad \text{Eq. 1}$$

An implicit assumption here is that the internal symmetry numbers for the radical and saturated molecule are the same. This is not guaranteed to be the case, particularly if the radical site is located on a rotor atom. Ideally, the full expression below should be used:

$$\sigma_{ext,rad} = \sigma_{ext,sat}^{3D} \frac{\sigma_{tot,rad}^{graph}}{\sigma_{tot,sat}^{graph}} \frac{\sigma_{int,sat}}{\sigma_{int,rad}} \quad \text{Eq. 2}$$

At this time, this refinement hasn't been implemented, though estimating the ratio on the right may be possible using graph-based methods, possibly using some of the same functionality used to estimate rotor symmetry number (discussed next).

5.4.2 Rotor symmetry number

When performing calculations with MM4 hindered rotor capabilities, rotor (or “internal”) symmetry effects must be considered for accurate computation of partition function and entropy. In principle, RMG’s graph-based total symmetry number, including the product of internal and external symmetry numbers could be used, in conjunction with the external symmetry number from the three-dimensional structure, could be used to compute the net internal symmetry number for all rotors. However, colleagues in the Green Group have found that the graph-based total symmetry number estimation approach currently used by RMG for cyclic species is unreliable. Therefore an alternative approach was desired, particularly in view of the fact that the main use of the QMTP system, at least initially, would be for cyclic species.

Toward this end, a graph-based approach was developed for non-radical species that appears to work for all cases considered thus far. The approach is implemented in *ChemGraph.calculateRotorSymmetryNumber(Node p_node1, Node p_node2)*. First, a “fragment” rotor symmetry number is calculated for each node in the rotor using *ChemGraph.calculateRotorFragmentSymmetryNumber(p_nodei)*. The rotor fragment symmetry number code is based off of the existing *calculateAtomSymmetryNumber* code, as the idea is similar, but it must be adjusted to handle incomplete fragments, triple bonds, and aromatic structures. This function returns an integer, either 1, 2, or 3. With the rotor fragment symmetry

numbers calculated, the rotor symmetry number is calculated by combining the two fragment symmetry numbers as shown in Table 15 to determine the overall rotor symmetry number.

Table 15. Determination of overall rotor symmetry number from rotor fragment symmetry numbers

Larger rotor fragment symmetry number	Smaller rotor fragment symmetry number	Overall rotor symmetry number	Example
3	3	3	Ethane
3	2	6	Toluene
3	1	3	Methanol
2	2	2	Biphenyl
2	1	2	Phenol
1	1	1	hydrogen peroxide

5.5 Directories created and used by the QMTP system

5.5.1 Description of directories and their contents

Several subdirectories are created and/or used by the QMTP system within the RMG working directory. A description of each follows:

- */2Dmolfiles*: The */2Dmolfiles* folder is used to store connectivity representations in MOL file format as created by RMG. Files are stored according to *[modified InChIKey].mol*. The Cartesian coordinates for all atoms are all zero. These files serve as the inputs to the three-dimensional molecular structure generation performed using *RDKit*. As the generation of these files is quite fast, the folder is cleared at the start of a new run.
- */3Dmolfiles*: The */3Dmolfiles* folder is used to store three-dimensional molecular structure representations in MOL file format. For each molecule, two files are stored. One of the files is the UFF-refined molecular geometry with minimum UFF energy among all the embeddings tested; this file is named according to *[modified InChIKey].mol*. The other file is the corresponding “crude” or raw molecular geometry (without UFF refinement);

this file is named according to *[modified InChIKey].cmol*. The Cartesian coordinates for all atoms in these files are, in general, non-zero. These files provide the initial guess geometries which will be included in input files for on-the-fly quantum/force-field calculations. As the generation of these files is relatively fast, the folder is cleared at the start of a new run.

- */InChI*: The QMTP system makes extensive use of RMG's InChI interface, and thus uses the */InChI* folder in the working directory.
- */QMfiles*: The */QMfiles* directory stores a number of files used in on-the-fly quantum or force-field calculations. The time involved in generation of each of the output files is on the order of seconds, so storage of these results can result in significant time savings between (and within) runs; therefore, unlike the */2Dmolfiles* and */3Dmolfiles* directories, the contents of this directory are left intact between runs. The files include:
 - *[modified InChIKey].gif* Gaussian03 input file
 - *[modified InChIKey].log* Gaussian03 output file
 - *[modified InChIKey].mop*: MOPAC2009 input file
 - *[modified InChIKey].out* and *[modified InChIKey].arc*: MOPAC2009 output files
 - *[modified InChIKey].com*: MM4 script file
 - *[modified InChIKey].comi*: MM4 rotor scan script file
 - *[modified InChIKey].mm4*: MM4 input file
 - *[modified InChIKey].mm4roti*: MM4 rotor scan input file
 - *[modified InChIKey].mm4out* and *[modified InChIKey].mm4opt*: MM4 output files
 - *[modified InChIKey].mm4rotouti* and *[modified InChIKey].mm4rotopti*: MM4 rotor scan output files
 - *[modified InChIKey].fmat*: MM4 force constant matrix output (when using CanTherm for post-processing)
 - *[modified InChIKey].can*: CanTherm input file (with MM4)
 - *[modified InChIKey].rotinfo*: CanTherm rotor input file (with MM4 rotor scans)
 - *[modified InChIKey].canout*: CanTherm output file (with MM4)
 - *[modified InChIKey].xyz*: XYZ format file containing Cartesian coordinates; used as an intermediate file format in connectivity checking with MM4 jobs

- *[modified InChIKey].symm*: SYMMETRY input file
- *[modified InChIKey].mol*: MOL file for optimized geometry with connectivity perceived (for connectivity checking)
- *[modified InChIKey].hold*: (empty) hold file to prevent other processes from trying to perform calculations for the same molecule at the same time; allows simultaneous RMG runs to perform PM3 calculations using the same */QMfiles* directory (discussed further below)

5.5.2 Managing QMfiles library

As mentioned previously, the contents of the */QMfiles* directory are preserved from run to run. There is a significant time savings when a result can be read in from disk, without first having to compute the result. On a Linux system, a softlink may be created in the working directory of an RMG job to point to a centralized */QMfiles* library that collects calculation results from jobs:

```
ln -s /home/gmagoon/JP10_2011/QMfiles/ QMfiles
```

Effort has been made to name working files with distinct file names based on InChIKey to avoid file conflicts when multiple jobs are using the same */QMfiles* library. Even with this file naming approach, however, it was found that the system could still crash in cases where multiple RMG runs were simultaneously considering the same species. To address this, the QMTP system now creates an empty “hold” file which will prevent other jobs from beginning calculations on the same species until the job finishes and the hold file is removed. If a pre-existing hold file is found, a check is performed every 60 seconds to monitor the existence of this file, and calculations are started when the file is found to have been removed.

6 Chapter 6: Application of explicit-3D-geometry-based on-the-fly species thermochemistry to JP-10 oxidation

6.1 Background

In order to test, debug, and apply the new RMG QMTP capabilities described in previous chapters, the RMG software package was applied to study the high-temperature oxidation of the jet fuel, JP-10.

JP-10 is an important military fuel that finds application in air-breathing propulsion.⁸⁵ It is composed essentially entirely of *exo*-tetrahydrodicyclopentadiene. Other names for this molecule include (*exo*-)THDCPD, *exo*-tricyclo[5.2.1.0^{2,6}]decane, or simply tricyclodecane. The molecular structure for this species is shown in Figure 17.

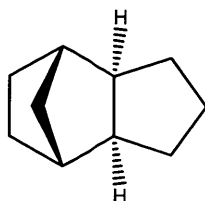


Figure 17. Molecular structure of *exo*-tetrahydrodicyclopentadiene, the primary constituent of JP-10
In the discussion that follows, this primary component of JP-10 fuel will be referred to as JP-10 for the sake of simplicity and brevity.

One of the main advantages of JP-10 as a fuel is its high volumetric energy density; it also offers low freezing point and good heat transfer properties.⁸⁵⁻⁸⁶ A number of ongoing research efforts are focused on the development of propulsion technologies (including ramjet, scramjet and pulse detonation engines) that burn JP-10 fuel.^{78, 84, 86a, 87}

Despite the relative simplicity of the composition of JP-10, its decomposition chemistry is quite complex. Several experimental and computational studies focusing on JP-10 have been conducted to characterize JP-10 pyrolysis and combustion chemistry.^{78, 84, 88} However, many important aspects of its decomposition behavior, particularly initial decomposition steps in the presence of oxygen, are not well-characterized in the existing literature.

In particular, there is a comprehensive pyrolysis mechanism developed by Herbinet *et al.*,⁸⁴ but it lacks oxidation chemistry; furthermore, the mechanism is not presently available to outside researchers. There is also a rough combustion model from Li *et al.*,^{87f} however, this mechanism

includes a significant number of non-elementary reactions such as $\text{JP-10} \leftrightarrow \text{C}_2\text{H}_2 + 2 \text{C}_2\text{H}_4 + \text{C}_4\text{H}_6$, which the authors refer to as “highly speculative”; also, besides JP-10, this model does not include any species larger than C_5 . Thus, when this work began, the biggest gap in modeling was related to the initial decomposition steps to C_5 in the presence of oxygen.

6.2 Mechanism development

Several iterations of mechanism generation using RMG were performed. In each case, various post-processing steps were performed on the raw mechanism generated by RMG. The mechanisms obtained are summarized in Table 16, and a more detailed discussion is provided in the subsequent sections.

Table 16. Summary of mechanism generation progress, with statistics of key raw mechanisms from each generation

	Improvements from previous generation	Example mechanism version	Core species	Core reactions	Additional (edge) species considered	Additional (edge) reactions considered
First generation	N/A	v0.19	317	~7.7k	25k	1.26m
Second generation	<ul style="list-style-type: none"> improved rate parameter estimates improved chemistry library, particularly for C_5 chemistry considered longer residence times 	v0.21	370	~4.8k	56k	2.16m
Third generation	<ul style="list-style-type: none"> comprehensive pressure-dependence consideration of a wider range of conditions during mechanism generation improvements to chemistry library, particularly aromatics 	v0.30	263	~9.5k	61k	1.34m

Fourth generation	<ul style="list-style-type: none"> refinements to chemistry library completed RMG run 	v0.50az	166	~3.2k	15k	0.43m
Comprehensive (combustion and pyrolysis) mechanism	<ul style="list-style-type: none"> incorporates chemistry generated by RMG under pyrolysis conditions 	CombPryl v0.53	930	~13.8k	N/A	N/A

6.2.1 1st generation mechanisms: capturing initial decomposition to C₅

The nominal conditions chosen for development of the first generation JP-10 combustion mechanism using RMG were $T = 1500$ K, $P = 1$ bar, and equivalence ratio (in air) = 1.0, which are relevant to high-temperature JP-10 oxidation applications. Triradicals and higher multiplicity radicals were forbidden from being included in the mechanism, along with species with more than ten carbon atoms or more than two oxygen atoms.


Additionally, several modifications were made to the set of reaction families considered by RMG. The “*1,3_Insertion_CO2*” reaction family was “turned off”, as it was expected that this reaction family would not be important at our (high temperature) conditions and turning it off significantly reduces the number of species that must be considered, hence speeding mechanism generation and reducing memory requirements. On the other hand, the “*Intra_R_Add_Exocyclic*” and “*Intra_R_Add_Endocyclic*” families were “turned on” (they had been “turned off” by default), corresponding to intramolecular radical addition to a double bond with the resulting radical exocyclic and endocyclic to the formed ring, respectively; these reactions are expected to be important for the JP-10 system as it involves a significant number of radical rings; these are intramolecular analogs to “*R_Addition_MultipleBond*” and the reverse reactions involve ring-opening by beta-scission.

Finally, a new “*Intra_Disproportionation*” reaction family for intramolecular disproportionation was implemented and “turned on” for mechanism generation. The expected importance of intramolecular disproportionation in the JP-10 system was first described by Herbinet *et al.*⁸⁴ A first order approximation of the Herbinet kinetic parameter estimation scheme was implemented in the RMG database for this reaction family. The family is implemented for transition state ring sizes ranging from 4 to 8 (3 to 7 non-hydrogen atoms), and reference cycle

activation energies were taken from Herbinet *et al.*⁸⁴; A and n parameter estimates were taken from Equation 1 of Warth *et al.*³ with $\Delta n_{\text{int}} = -1$. Further details and discussion concerning this reaction family are provided in the next chapter.

A combustion core mechanism based on the Leeds methane combustion mechanism⁸⁹ was used as a seed mechanism for mechanism generation. A goal reaction time of 0.001 s was specified, as well as a mechanism generation tolerance of 0.15.

Before terminating due to memory limitations, RMG performed mechanism generation through JP-10 conversions as high as 98.5% and reaction times as long as 11 microseconds. At this point the mechanism consisted of 316 species and 7,704 reactions. In constructing this mechanism, RMG had considered, but not included, 1,264,644 additional reactions and 25,352 additional chemical species.

As described previously, inclusion of triradicals (such as CH) and other high multiplicity radicals was forbidden in the automatic mechanism generation. Instead, CH chemistry (based on the Leeds methane oxidation mechanism) was manually added to the RMG-generated model for ignition modeling. Thermochemical data in the RMG-generated model were updated with values from GRI-Mech 3.0⁹⁰ for selected species appearing in both mechanisms. Additionally, two cyclopentadienyl decomposition reactions from a hexadiene mechanism compiled by Sharma *et al.*^{8, 91} were incorporated to supplement RMG's treatment of these chemistries. Finally, since the mechanism generated automatically by RMG exhibited accumulation of cyc-C₅H₄ ()₂, the reaction cyc-C₅H₄ ↔ HCCCH + HCCH, was manually added to the mechanism with an estimated reverse rate constant of 1×10^{13} mol/cm³-s.

The RMG mechanism with these modifications constitutes mechanism “v0.10”, which is described and analyzed elsewhere.⁹² Including a total of 317 species and 7715 reactions, this mechanism also includes hundreds of species larger than C₅, demonstrating that it is significantly more comprehensive than the Li *et al.* mechanism^{87f} in terms of capturing initial decomposition chemistry. (The mechanism of Li *et al.* contains 36 species, which are all C₅ or smaller, with the exception of JP-10 itself, and 174 reactions.) Carbon and oxygen flux diagrams based on simulations with the v0.10 model (created using Dr. Richard West's RMG-Visualizer) are shown in Figure 18 and Figure 19, respectively, illustrating some of the chemistry detail captured in this RMG-based model.

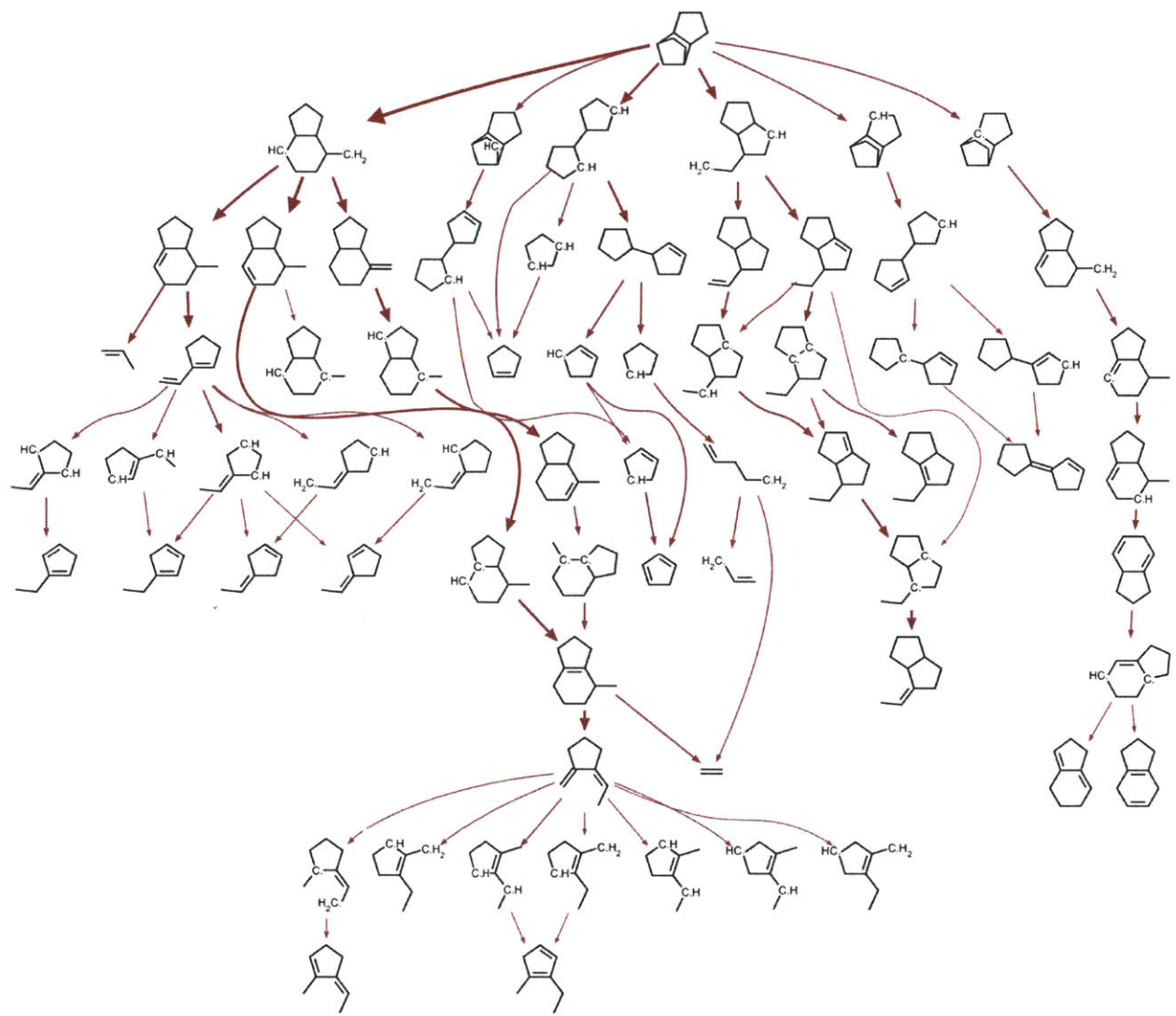


Figure 18. A carbon flux diagram (created using RMG-Visualizer) based on RMG JP-10 model v0.10 (arrow thickness based on flux magnitude, with darker, thicker arrows corresponding to higher flux)

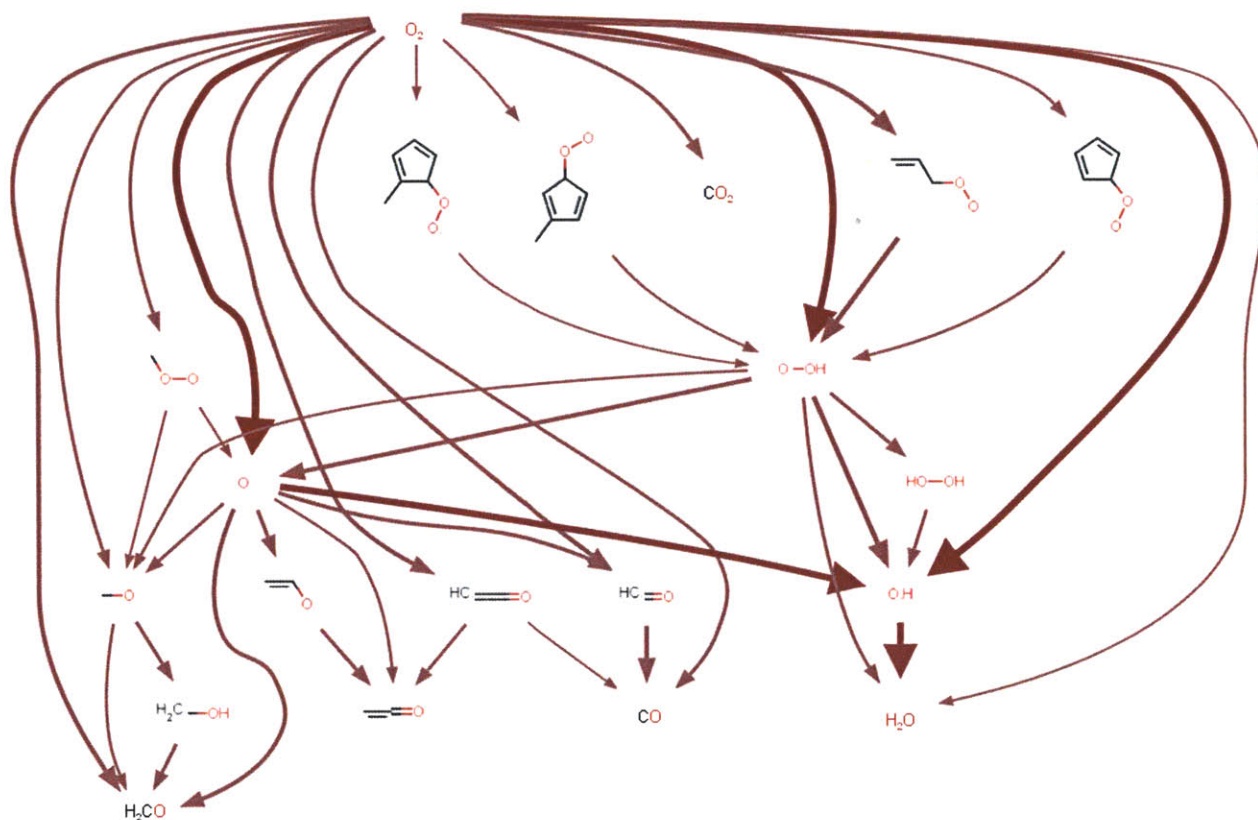


Figure 19. An oxygen flux diagram (created using RMG-Visualizer) based on RMG JP-10 model v0.10 (arrow thickness based on flux magnitude, with darker, thicker arrows corresponding to higher flux)

The progression from the RMG-generated mechanism to the final mechanism in this first-generation series (called “v0.19”) is illustrated in Figure 20.

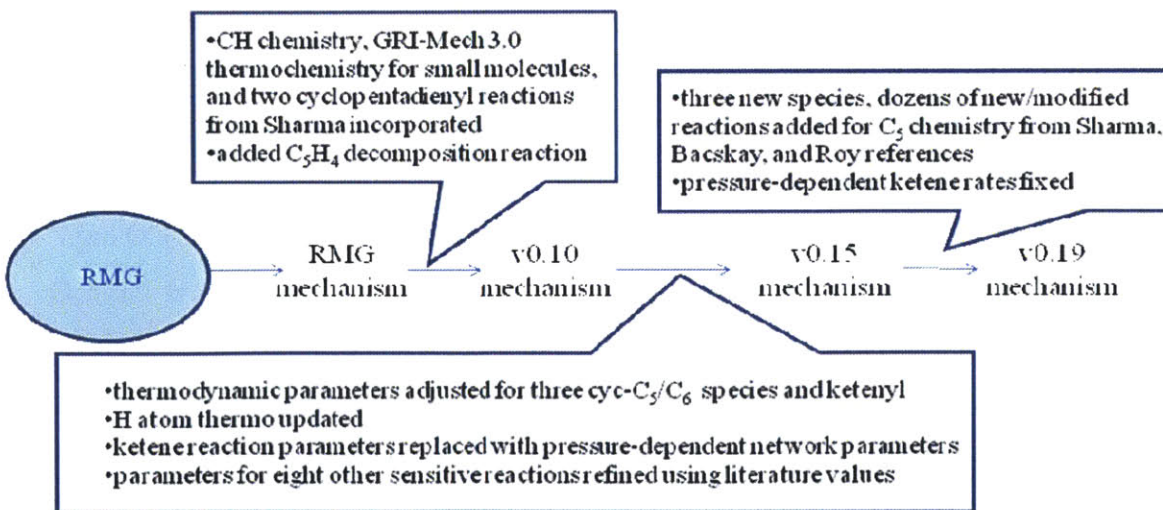


Figure 20. Development of first generation mechanism series from raw RMG mechanism. (some intermediate mechanisms omitted for clarity and brevity)

Mechanism refinements focused on capturing important aspects of cyclic C₅ and ketene chemistry, using experimental ignition delay data (primarily shock tube data from Davidson *et al.*⁷⁸) and sensitivity analysis as a guide.

Thermochemical parameters for ketylenyl, cyclopentadienyl, methylcyclopentadienyl and fulvene were refined (see Table 17) to use more reliable values from the literature (as parameter estimation error was judged to be potentially significant for these species, based on sensitivity analysis and typical errors associated with the estimation techniques originally used in RMG). Hydrogen atom thermochemical properties were also updated by using NASA coefficient values from GRI-Mech version 3.0.

Table 17. Thermodynamic parameters updated in refinement of first-generation RMG JP-10 mechanism

Species	Original G° _{1500K} (kcal/mol)	Updated G° _{1500K} (kcal/mol)	Source of updated parameters
ketylenyl (HCCO)	-63.30	-63.26	Ref ⁹⁰
H atom (H)	4.906	4.908	Ref ⁹⁰
cyclopentadienyl (C ₅ H ₅)	-71.0	-68.9	Ref ⁹¹
1-methylcyclopentadienyl (C ₆ H ₇)	-109.3	-109.7	Ref ⁹¹
fulvene (C ₆ H ₆)	-91.9	-97.1	Ref ⁹¹

Additionally, sensitivity analysis suggested that kinetic parameter values for the ketene reaction $\text{H} + \text{HCCO} \leftrightarrow \text{H}_2\text{CCO}$ were critical and pressure-dependent network calculations were performed using high-pressure limit rates from the literature and the pressure-dependent network calculation features of RMG. To obtain high pressure limit rates as input to the pressure-dependent network calculations, the $\text{H} + \text{CHCO} = \text{CH}_2\text{CO}$ high-pressure rate coefficient was taken from the $\text{H} + \text{CHCO} \rightarrow \text{CH}_2 + \text{CO}$ rate from Baulch⁹³; the $\text{CH}_2(\text{s}) + \text{CO} = \text{CH}_2\text{CO}$ high pressure rate coefficient was taken as the k_{inf} in GRI-Mech for $\text{CH}_2 + \text{CO} (+\text{M}) \rightarrow \text{CH}_2\text{CO} (+\text{M})$.⁹⁴ CH_2 was assumed to be singlet, as the literature suggests about 92% goes to singlet at room temperature in $\text{H} + \text{CHCO} \rightarrow \text{CH}_2 + \text{CO}$.⁹⁵ These calculations produced Chebyshev fits for $k(T,P)$ that were used to replace existing kinetic parameter estimates for the reactions $\text{H} + \text{HCCO} \leftrightarrow \text{H}_2\text{CCO}$ and $\text{H} + \text{HCCO} \leftrightarrow \text{CH}_2(\text{s}) + \text{CO}$ and to add the reaction $\text{CH}_2(\text{s}) + \text{CO} \leftrightarrow \text{H}_2\text{CCO}$ (where “(s)” indicates a singlet state).

Additionally, RMG's original parameter estimates for eight other reactions, which sensitivity analysis suggested were important, were replaced with more reliable values from the literature. Reaction parameters for C₅ chemistry were also incorporated from work by Sharma *et al.*⁸, Bacskay *et al.*⁹⁶, Roy⁹⁷, and Moskaleva and Lin⁹⁸ and were used to either replace earlier estimates with more reliable values or to incorporate additional reaction pathways that had not been originally included. The inclusion of several of these new pathways necessitated the addition of the following three species that were not previously present in the mechanism: CH₂CHCHCH, CH₂CHCHCCH₂, and CH₂CHCH₂CCH. Overall, in addition to the three new species, the refinements in the first-generation series of mechanisms since v0.10, added with the assistance of Ben Ruiz-Yi, included the addition of 24 new reactions and updates to the parameters to 24 other reactions.

This series of mechanisms was found to reproduce experimental ignition delay measurements, particularly those from Davidson *et al.*⁷⁸, well. A parity plot showing modeled ignition delay versus experimental ignition delay from Davidson *et al.* appears in Figure 48. (It should be noted that some of the model simulation results presented here and later in the chapter come from simulations performed by Luwi Oluwole at Aerodyne Research, Inc.) The plot shows significant improvement in the progression from v0.10 to v0.19, particularly for the four outlying points in v0.10, which correspond to fuel rich conditions.

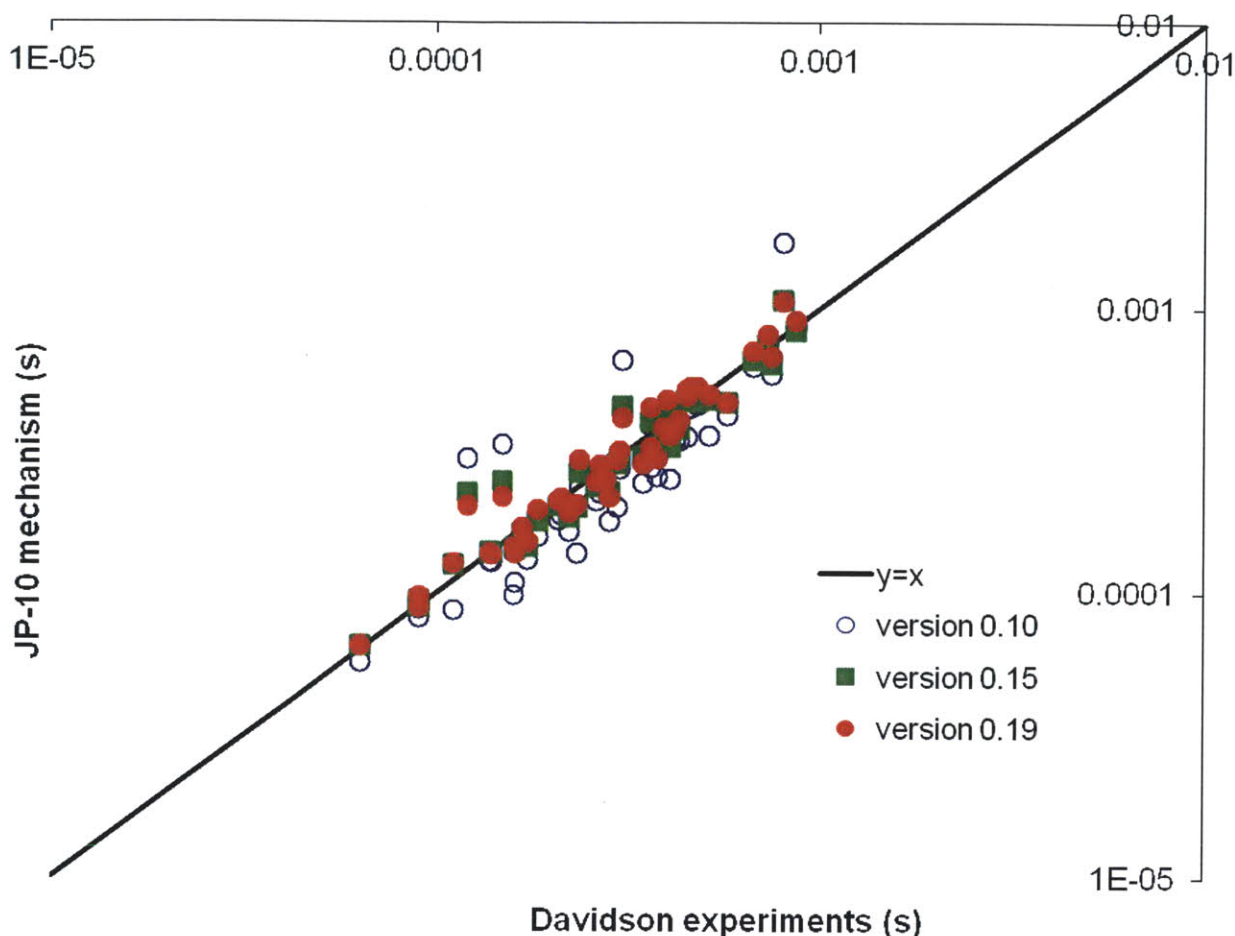


Figure 21. Parity plot of modeled ignition delay compared to experimental ignition delay from the results of Davidson et al.⁷⁸ for selected first-generation RMG-generated mechanisms

However, this mechanism series relied heavily on the $\text{cyc-C}_5\text{H}_4 \leftrightarrow \text{HCCCH} + \text{HCCH}$ reaction and associated kinetic parameter estimates; when this reaction was removed or when *ab initio* parameter estimates (obtained by Jorge Aguilera-Iparraguirre) were used, the ability of the model to reproduce the Davidson *et al.* data suffered substantially. Furthermore, mechanisms in this series suffered from unrealistic kinetic parameter estimates in some cases, which could exceed a reasonable collision limit value for reverse rate coefficients, leading to very stiff chemistry and preventing straightforward flame simulations with the model. To work around these issues, another round of mechanism generation was performed using the latest RMG source code and refinements to RMG's reaction libraries based on the findings during refinement of the 1st generation mechanism. This next round is described in the next section.

6.2.2 2nd generation mechanisms: avoiding suspicious cyclic C₅ decomposition routes

The first generation series of mechanisms from v0.10 to v0.19 was based on manual refinement of results of a single round of automated reaction mechanism generation with RMG performed in Fall 2009. In order to address the various shortcomings of this earlier mechanism series and improve the accuracy and comprehensiveness of the mechanism generation process, new rounds of automated reaction mechanism generation were begun. In these new rounds, experiences from the first generation, findings from quantum chemistry to improve the reaction chemistry and parameter values used by RMG in the mechanism generation process. Also, the QMTP system was set up on a Linux cluster with greater memory capacity to facilitate the generation of larger mechanisms; a personal laptop had been used for the first generation RMG mechanism. Finally, the latest updates to the RMG source code and core database were incorporated. Source code modifications included a fix to produce more reasonable kinetic parameter estimates that will not exceed the collision limit for the reverse reaction; as mentioned previously, in the first generation mechanism series, some reverse rate coefficients, computed according to the principles of thermodynamic consistence, were unrealistically high, particularly at low temperatures, leading to numerical difficulties in attempts to perform flame simulations. Database modifications included the addition of a new reaction family, “1,2-Birad_to_alkene”.

The development, implementation, and use of the “1,2-Birad_to_alkene” family was aimed at more accurately modeling the behavior of biradicals, which were common in our earlier mechanism generation efforts and are expected to be important for the JP-10 system; in particular, this reaction family aims to capture the “relaxation” of hydrocarbons with adjacent biradical sites to alkenes. The relaxation is treated as a first-order decay process, with the rate coefficient assumed to be $k = 1/\tau$, where:

$$\log_{10} \tau(\text{s}) = -8.0 + 0.2m + 0.3n \quad \text{Eq. 3}$$

with m = number of alkyl substituents and n = number of aryl/vinyl substituents; this correlation, as well as the justification for treatment of alkene triplets as 1,2-biradicals, comes from work by Caldwell and coworkers in the literature.⁹⁹ Each non-hydrogen substituent provides a slow-down effect. To extrapolate to other groups not explicitly considered in the cited literature, rough assignments to “ m -slowdown” or “ n -slowdown” were made based on the argument in the literature^{99a} that the slowdown effect is mainly related to number of hydrogens and mass of

substituents, rather than electronic stabilization/polar effects. Thus, “*Cs*” and “*Os*” RMG atom types were assigned to the “*m*-slowdown” group whereas “*Cd*”, “*Cr*”, “*Cb*”, “*CO*” atom types were assigned to the “*n*-slowdown” group, while “*H*” provides no slowdown. Note, however, that this will not correctly account for the large mass of extended groups like large alkyl chains; still, the effect is relatively small, and the resulting estimate should still be within an order or magnitude or so of what we would obtain if *m*-slowdown and *n*-slowdown groups had been assigned differently).

Many modifications to the custom JP-10 reaction and thermodynamics databases were also incorporated into the newest mechanism generation attempts. These modifications include the incorporation of the refined chemistry that had been manually applied in the process of developing v0.19 from the raw RMG base mechanism. Also, additional quantum calculations for cyc-C₅ chemistry were performed and the results were incorporated into the custom JP-10 reaction database. The reactions investigated and the associated rate parameters obtained via CBS-QB3 calculations and CanTherm are summarized in Figure 22, Figure 23, and Table 18. Intramolecular hydrogen migration and beta-scission ring opening pathways on the C₅H₅ potential energy surface (*R1-R5*) were investigated by the author and a pathway for breakup of the C₅ ring on the C₅H₇ potential energy surface (*R6-R8*) were studied by Jorge Aguilera-Iparraguirre.

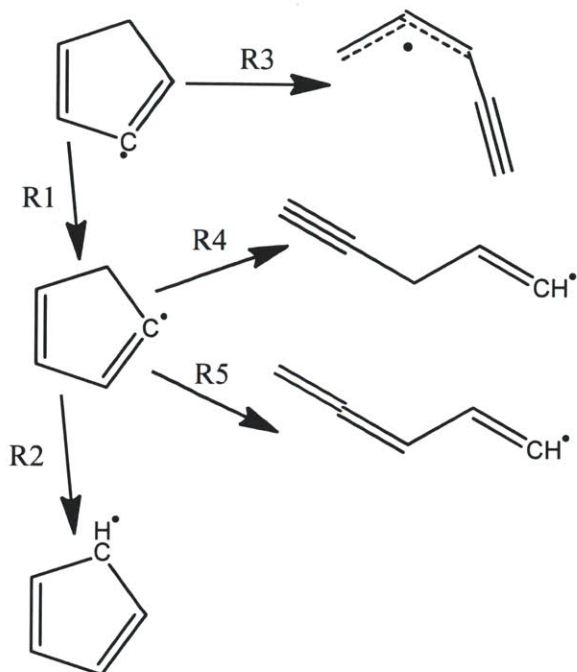


Figure 22. Intramolecular hydrogen migration and beta-scission ring opening pathways on the C_5H_5 potential energy surface investigated with CBS-QB3 quantum calculations

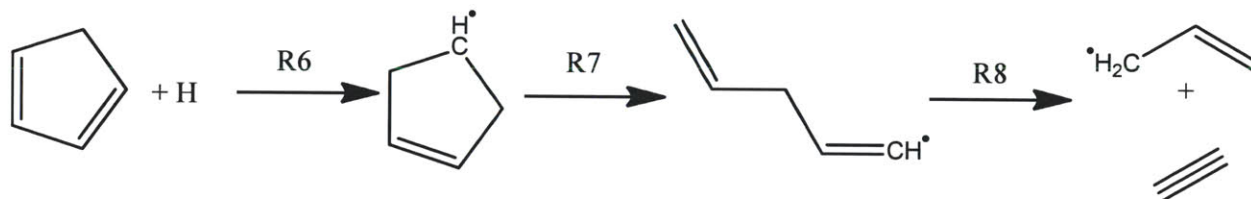


Figure 23. Breakup of the cyc- C_5 ring on the C_5H_7 potential energy surface investigated with CBS-QB3 quantum calculations

Table 18. Reaction rate parameters for cyclic- C_5 chemistry based on CBS-QB3 quantum calculations

Reaction	A ($1/s$ or $cm^3/mol\cdot s$)	n	E (kcal/mol)
R1	2.52×10^7	1.87	55.0
R2	5.82×10^6	1.88	24.67
R3	8.78×10^{11}	0.77	33.44
R4	9.66×10^{11}	0.65	42.43
R5	6.70×10^{11}	0.70	41.82
R6	1.108×10^9	1.55	2.28
R7	8.68×10^{12}	0.48	43.11
R8	4.72×10^{11}	0.81	24.35

One round of mechanism generation was performed using essentially the same conditions used to develop the first generation of RMG mechanisms (a slight change was made for the pressure, which was modified slightly from 1 bar to 1 atm). To avoid job suspension limitations, this job was run with a walltime limit of 5 days. With subsequent post-processing (analogous to that used in the generation of mechanism v0.10; *e.g.* manual addition of CH chemistry) this was named v0.20. After the job terminated due to the walltime limitation, mechanism generation was started again, now with many more pre-computed PM3 calculations available, allowing mechanism generation to generate a larger model before hitting the 5 day wall-time limit. With postprocessing, this was named v0.21. Details on the generated mechanisms and the mechanism generation process are shown in Table 19 and Table 20, below.

Table 19. Second generation raw RMG-generated mechanism size comparison

Mechanism	Core Species	Core Reactions	Edge Species	Edge Reactions
0.10 (raw)	316	7,704	25,352	1,264,644
0.20 (raw)	343	7,583	54,766	1,653,616
0.21 (raw)	369	4,829	56,364	2,159,290

Table 20. Second generation mechanism development statistics

Mechanism	Maximum JP-10 conversion considered during mechanism generation	Maximum reaction time considered during mechanism generation	Aromatic species included?	cyc-C ₅ H ₄ included?
0.10	98.5%	11 μ s	no	Yes
0.20	100.0%	364.5 μ s	phenyl (C ₆ H ₅)	No
0.21	100.0%	133.6 μ s	benzyl (C ₇ H ₇)	No

Simulated ignition delay results (based on the strongest CH peak) compared to the Davidson *et al.* shock tube data are shown in Figure 2, below.

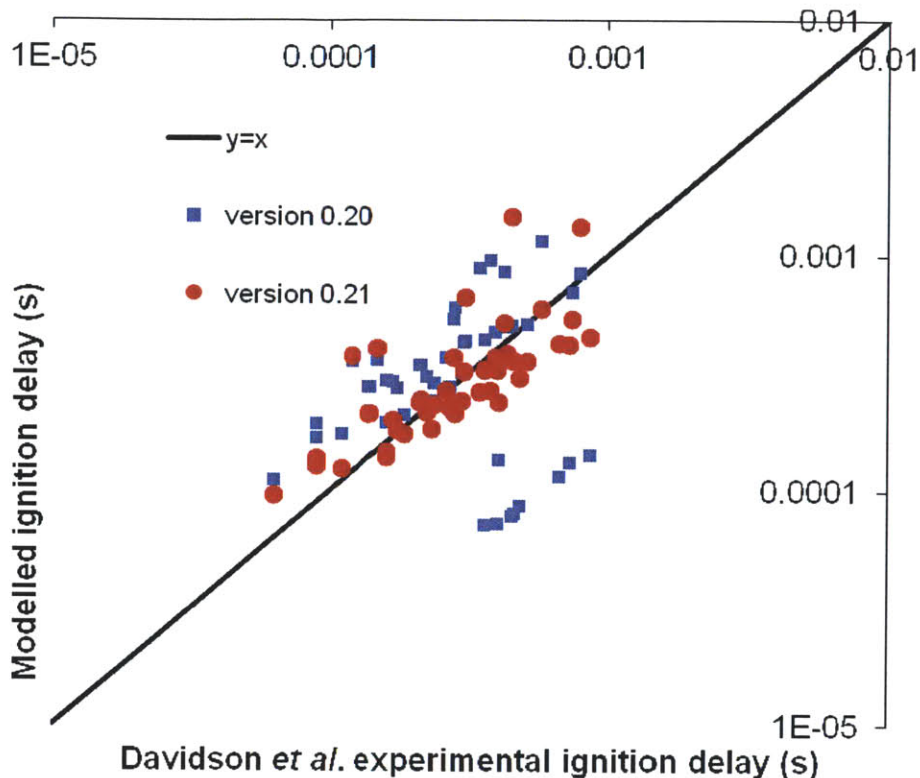


Figure 24. Parity plot comparing results of second-generation mechanism versions to Davidson et al. experiments

As shown in Table 20, mechanisms v0.20 and v0.21 include aromatic radical species, which had not appeared in first-generation mechanisms. This was promising, given emerging experimental evidence from Aerodyne, as well as published research, pointing to the significant fraction of aromatic products arising from JP-10 decomposition. Closer investigation of mechanism v0.20 suggested that phenyl radical was accumulating in the products. In addition to including aromatic radicals, the new mechanisms were noteworthy in that they were able to produce reasonable ignition delay predictions without including the speculative C_5H_4 decomposition kinetics that had been used in the first generation series of mechanisms.

In parallel with automated reaction mechanism generation efforts with v0.21, the effects of manually perturbing the v0.20 mechanism, particularly the small-molecule chemistry, were explored. Changes are summarized in Figure 25 below, and ignition delay simulation results are shown in Figure 26.

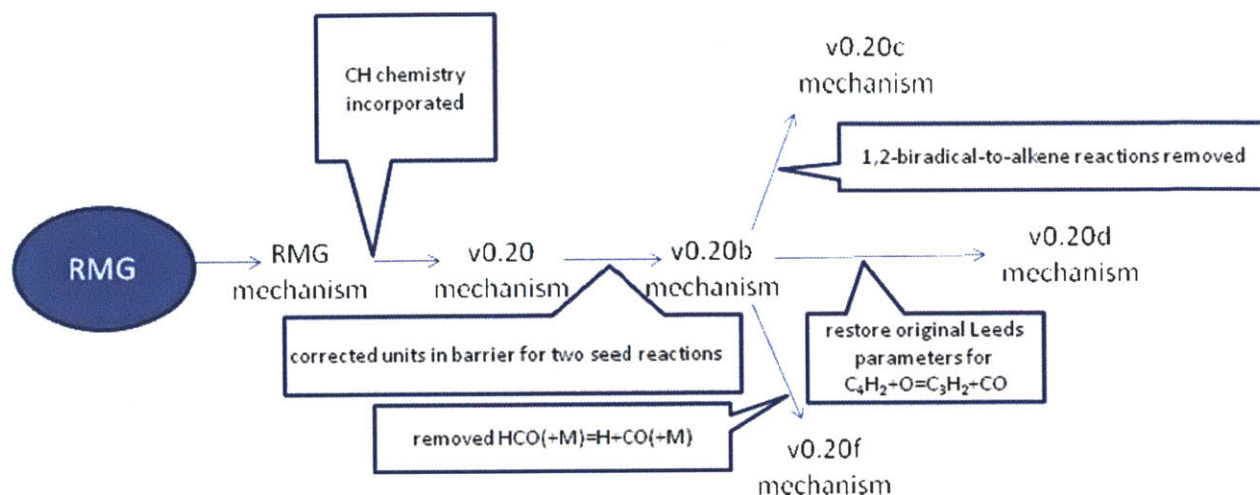


Figure 25. Summary of manual perturbations to mechanism v0.20.

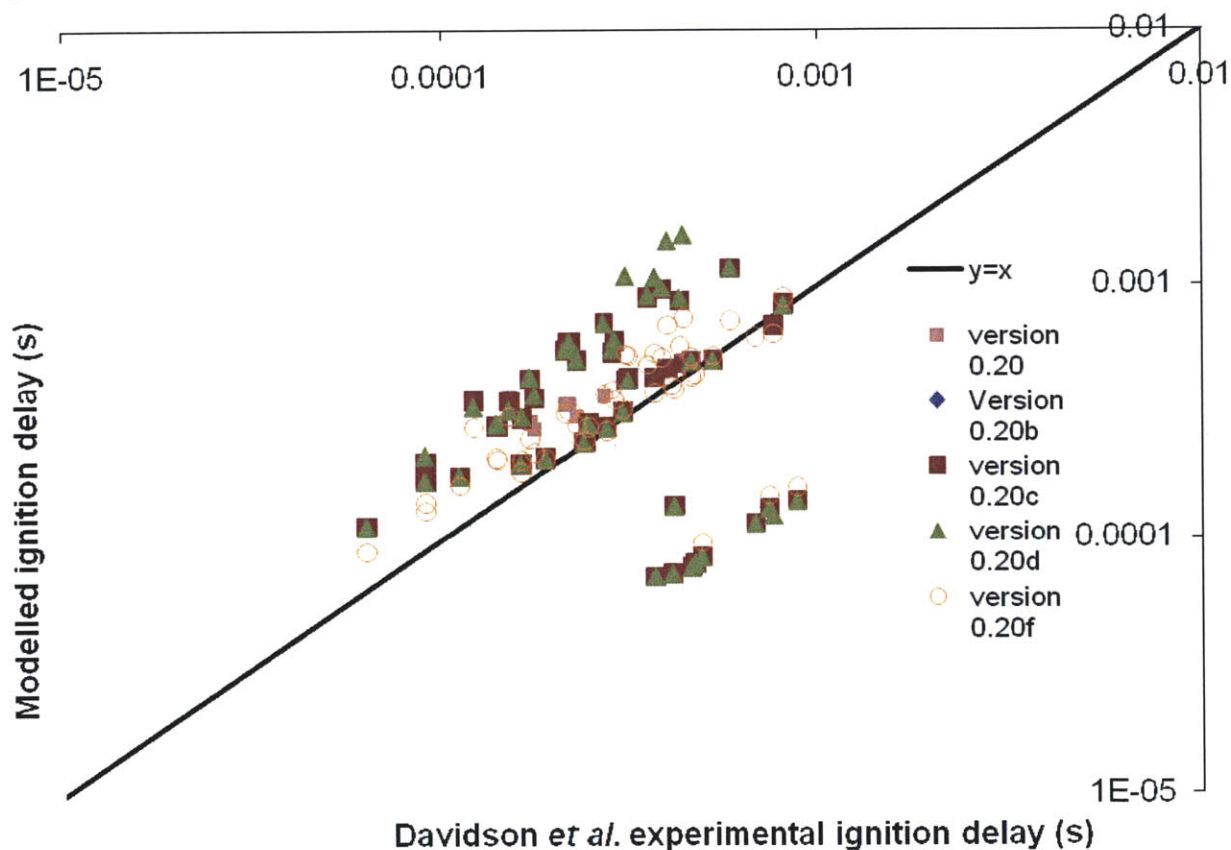


Figure 26. Parity plot comparing results of perturbed v0.20 to Davidson et al. experiments.

Relative to the reference v0.20b, the most noteworthy effect appears to be due to the modification going to v0.20f, which seems to noticeably improve agreement for most of the datapoints. As noted in Figure 25, this corresponds to the removal of the HCO decomposition reaction: $HCO (+M) \leftrightarrow H + CO (+M)$. Upon closer investigation of the parameters for this

reaction, we found that an alternative rate coefficient from GRI-Mech was significantly different. A comparison of the high-pressure limit for this reaction is shown in Figure 5, below, showing that the GRI-Mech rate coefficient is nearly an order of magnitude lower than the Leeds rate coefficient at sufficiently high temperatures.

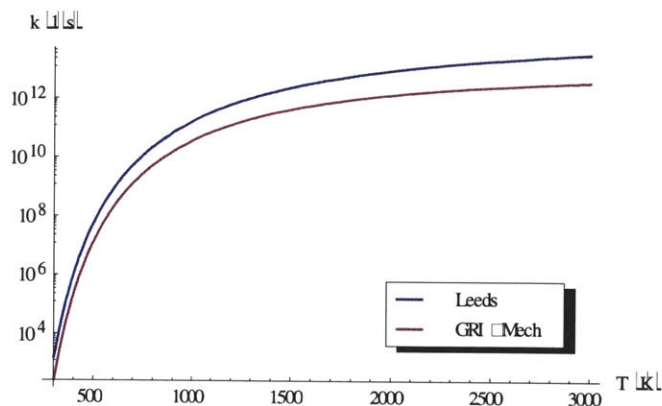


Figure 27. Comparison of high-pressure rate coefficients for $\text{HCO (+M)} \leftrightarrow \text{H + CO (+M)}$ from Leeds and GRI-Mech

As we expect the GRI-Mech rate to be more reliable for this reaction, we decided to switch to using GRI-Mech parameters for this reaction in future mechanism development efforts. The effect of this change should be intermediate between leaving the reaction unperturbed (0.20b) and eliminating it completely (v0.20f)

These second-generation mechanisms demonstrated several noteworthy improvements over previous mechanism generation efforts in the first-generation series, including:

- reasonable ignition delay predictions with mechanisms that do not include cyc-C₃H₄ and associated *ad hoc* decomposition reaction
- more physically-reasonable reaction rates at low temperature
- automatically incorporated “seeds” of aromatic chemistry (i.e. inclusion of benzyl and phenyl radicals)
- consideration of longer reaction times and higher JP-10 conversions during mechanism generation

The new mechanisms also had some limitations. In particular, simulated ignition delay predictions do not match the data of Davidson *et al.* as closely. Also, despite the promising appearance of “seeds” of aromatic chemistry in these mechanism development efforts, the aromatic chemistry were still not adequately captured (*e.g.* results exhibited accumulation of

aromatic radicals and most key aromatic species are still not included). Analysis of these limitations was used to guide further refinements to the mechanism generation process.

6.2.3 3rd generation mechanism: capturing multiple reaction conditions, pressure-dependence, and aromatic chemistry

The next round of mechanism generation efforts focused on performing mechanism at multiple reaction conditions, incorporating pressure-dependence more comprehensively, and modeling aromatic chemistry more completely.

There were substantial updates to the reaction libraries used by RMG for this round of mechanism generation, focused mainly on capturing aromatic chemistry from the literature. A description of the libraries that have been used in these mechanism generation efforts follows. The “*JP10rxnlib*” and “*JP10seedmech*” were developed by the author, while the others were developed by Nick Vandewiele, a visitor to the group from Ghent University.

- “*JP10rxnlib*” and “*JP10seedmech*”: The main JP-10 reaction library, “*JP10rxnlib*”, which had previously included chemistry from multiple sources, including Herbinet et al., Bacskay, and our own calculations, was updated to include new aromatic and cyclic C₅ chemistry. The reactions and kinetics consider chemistry related to compounds including benzene, phenol, cyclopentadiene, benzyne, styrene, phenylacetylene, toluene, fulvene, and methylcyclopentadiene, from a compilation (from multiple literature references) by Sharma *et al.*⁸ and from work by Sharma and Green⁹¹. Some of the newly-incorporated pressure-dependent chemistry was included in a separate seed mechanism, “*JP10seedmech*” as it couldn’t be included in the main reaction library due to a current RMG functionality limitation related to PLOG format kinetic parameters.
- “*Sabbe*”: The “*Sabbe*” library was a newly created library with chemistry for cyclopentene, cyclohexadiene, and cyclopentadiene taken from Sabbe *et al.*¹⁰⁰
- “*chae*”: The “*chae*” library was a newly created library with decalin chemistry taken from work by Chae and Violi.¹⁰¹
- “*fascella*”: The “*fascella*” library was a newly created library incorporating cyclopentadienyl and indene chemistry as reported by Fascella *et al.*¹⁰²

- “*kislovB*”: The “*kislovB*” library was a newly created library with kinetic pathways and parameters involving benzene, toluene, and indene chemistry as reported by Kislov and Mebel.¹⁰³
- “*kislovCPD*”: The “*kislovCPD*” library was a newly created library with kinetic pathways and parameters involving cyclopentadienyl, cyclopentadiene, and indene chemistry as reported by Kislov and Mebel.¹⁰⁴
- “*murakami*”: The “*murakami*” library was created to include the global reaction pathway cyclopentadienyl + cyclopentadienyl \leftrightarrow naphthalene + 2H with kinetic parameters from Murakami *et al.*¹⁰⁵
- “*wang*”: The “*wang*” library was a newly created library with kinetic parameters and reactions from Wang *et al.*¹⁰⁶ The library includes chemistry for cyclopentadiene and cyclopentadienyl.

At this stage, connectivity checking had been implemented in the QMTP system, but was set to “check” due to its experimental nature at the time. Pressure dependence calculation features of RMG were also turned on (in the form of Modified Strong Collision calculations, which had been implemented in RMG by Josh Allen); with the exception of a few reactions from the literature, the previous mechanism generations used the high-pressure limit for reaction coefficients, which is only a function of temperature; pressure-dependent effects were expected to be particularly important in this JP-10 system for modeling the chemistry of smaller species (C₅ and lower), especially at higher temperatures and lower pressures.

Also, this third generation mechanism generation effort made use of functionality implemented in RMG by the author to generate models valid at multiple conditions. This feature will consider a number of separate reaction systems simultaneously at each iteration. The reaction systems are each simulated until either a edge flux tolerance is exceeded or the target time/conversion is exceeded. As the edge species that will be brought into the core can be different for each reaction system, depending on the conditions, the core model can grow by more than one species at each iteration of the rate-based model generation algorithm. In this case, three temperatures (1250 K, 1500 K, and 2000 K), two pressures (1 atm and 10 atm), and three sets of initial concentrations (roughly corresponding to equivalence ratios of 0.5, 1.0 and 2.0 in air) were combinatorially combined to produce a total of 18 reaction systems that were considered during mechanism generation. Earlier test runs of this system with pressure-

dependent mechanism generation suggested that tolerances on the order of 0.15 used previously resulted in extensive (possibly overly so) and very large pressure-dependent networks, which became extremely computationally demanding. Therefore, a slightly looser mechanism generation tolerance (0.20) was used as a practical consideration to mitigate this issue, allowing more timely exploration of the relevant chemistry by RMG. An intermediate version of the RMG results was extracted after approximately five weeks of walltime.

Postprocessing and mechanism refinement is summarized in Figure 28. Initial postprocessing was similar to that used in generation of earlier mechanisms (*e.g.* addition of CH chemistry), and the resulting mechanism (“v0.30alpha”^{**}) exhibited significant stiffness; further postprocessing was performed to produce a more reliable mechanism. First, the stiffness issues were addressed by fixing units errors in PLOG expressions that had been part of new “JP10seedmech” library additions; one error was the result of a conversion error in the Arrhenius prefactor introduced during transcription from the literature and another was the result of an error in the RMG source code related to the activation barrier units. The resulting mechanism (“v0.30beta”^{††}) still exhibited poor ignition delay prediction when tested at the Davidson *et al.* conditions. Global sensitivity analysis on ignition delay predictions for this model was performed by Luwi Oluwole using the RS-HDMR (Random Sampling-High Dimensional Model Representation) method¹⁰⁷ at the Davidson *et al.* conditions with starting temperature 1527K at 1.15 atm; the eleven reactions determined through this analysis to be responsible for greater than 0.5% of the variation in ignition delay are summarized in Table 21. (It is noted that a glossary is provided in Appendix II, containing structures for species with ambiguous names in this table or in similar tables elsewhere in this chapter.)

Table 21. Results of RS-HDMR ignition delay sensitivity analysis of the v0.30beta mechanism at Davidson et al. conditions with starting temperature 1527K at 1.15 atm

Reaction	Fraction of variation in ignition delay
$O_2+H=OH+O$	67.7%
$C_5H_5_A+O_2(+m)=C_5H_5O+O(+m)$	6.7%

^{**} RMG_jp10_v0_30_chem_inchi.inp

^{††} RMG_jp10_v0_30_chem_inchi_fixedPLOG2.inp

cyclopentadienyl+H(+m)=cyclopentadiene(+m)	5.2%
H+ CH ₂ CHCH ₂ (+m)= CH ₂ CHCH ₃ (+m)	1.9%
CH ₃ +OH=CH ₂ (s)+H ₂ O	0.9%
C ₂ H ₂ +O=HCCO+H	0.9%
C ₂ H ₃ +CH ₃ = CH ₂ CHCH ₂ +H	0.9%
H ₂ O+H=H ₂ +OH	0.8%
HCO+m=H+CO+m	0.8%
CH ₃ +O=CH ₂ O+H	0.7%
C ₂ H ₂ +CH ₃ (+m)= CH ₂ CHCH ₂ (+m)	0.6%

It should be noted that by this point, we had also switched to using alternative ignition delay metrics based on the maximum dT/dt or on the maximum product of the mole fractions $y_{C_2H}y_{O_2}$ (as a proxy for excited CH). In the case of the RS-HDMR analysis above, the maximum in dT/dt was used to identify ignition.

Closer investigation of these reactions identified several opportunities for model refinement. A more reliable parameter set for the top reaction was found to have been published by Hong *et al.*¹⁰⁸ for the top reaction $O_2 + H \leftrightarrow OH + O$ and their results (with a rate coefficient roughly 10% lower at 1500 K) were incorporated into the model. Additionally, the kinetic parameters for $CH_3 + OH \leftrightarrow CH_2(S) + H_2O$, the fifth reaction in the above table, were refined based on work by Jasper *et al.*¹⁰⁹; their variant of Troe form for the rate coefficient of this reaction was re-fitted to Chebyshev form using RMG-Py for the case of Ar bath gas, and these refined kinetic parameters were incorporated into the model. Finally, related to the second reaction in the above table, the thermodynamic parameters for the two vinylic cyclic C₅H₅ radicals in the model were refined based on CBS-QB3 calculations; one of these changes involved a correction of about 16 kcal/mol in standard heat of formation at 298.15 K, and this was the case related to unreliable hydrogen bond increment values discussed in Chapter 4. Simulations (of a version of v0.30 that had been merged with Nick Vandiewiele's pyrolysis model^{‡‡}) by Luwi Oluwole showed that these thermochemistry refinements, in particular, produced a significant improvement in ignition delay predictions when compared to the

^{‡‡}; v0.40g (or v0.41; CombustionPyrolysis_v0_30_v1_00_gas_fixedPLOG2mod3 .inp) vs. v0.40f (CombustionPyrolysis_v0_30_v1_00_gas_fixedPLOG2mod2.inp)

Davidson *et al.* dataset. It should be noted however, that, at this point, the pressure-dependent network calculations were still based on the original thermochemistry values. The result of all these changes represents our “v0.30” model.

In parallel with these efforts, Nick Vandewiele, visitor from Ghent University, worked to generate a comprehensive, detailed JP-10 pyrolysis mechanism. By turning off pressure-dependence calculations and a number of reaction families specific to oxygen, he was able to speedily generate a very large pyrolysis mechanism (here termed “Pyrolysis v1.00”) validated against his flow reactor data; the model was also generated without the use of the QMTP functionality. The Pyrolysis v1.00 model was merged with our v0.30 model using the Mechanism Merge utility of the Reaction Workbench of CHEMKIN-MFC to produce a merged v0.41 model. In this merge, the combustion mechanism was treated as master and the pyrolysis mechanism was treated as donor such that the master kinetic and thermodynamic parameter values are used in cases of conflict. A summary of the number of species and reactions associated with the mechanism merge appears in Table 22. Model development is summarized in Figure 28.

Table 22. Summary statistics for third-generation mechanism merge (master and donor values do not sum to merged value due to mechanism overlap)

Mechanism	Number of species	Number of reactions
v0.30 (master)	263	9502
Pyrolysis v1.00 (donor)	838	11122
v0.41 (merged mechanism)	1018	19909

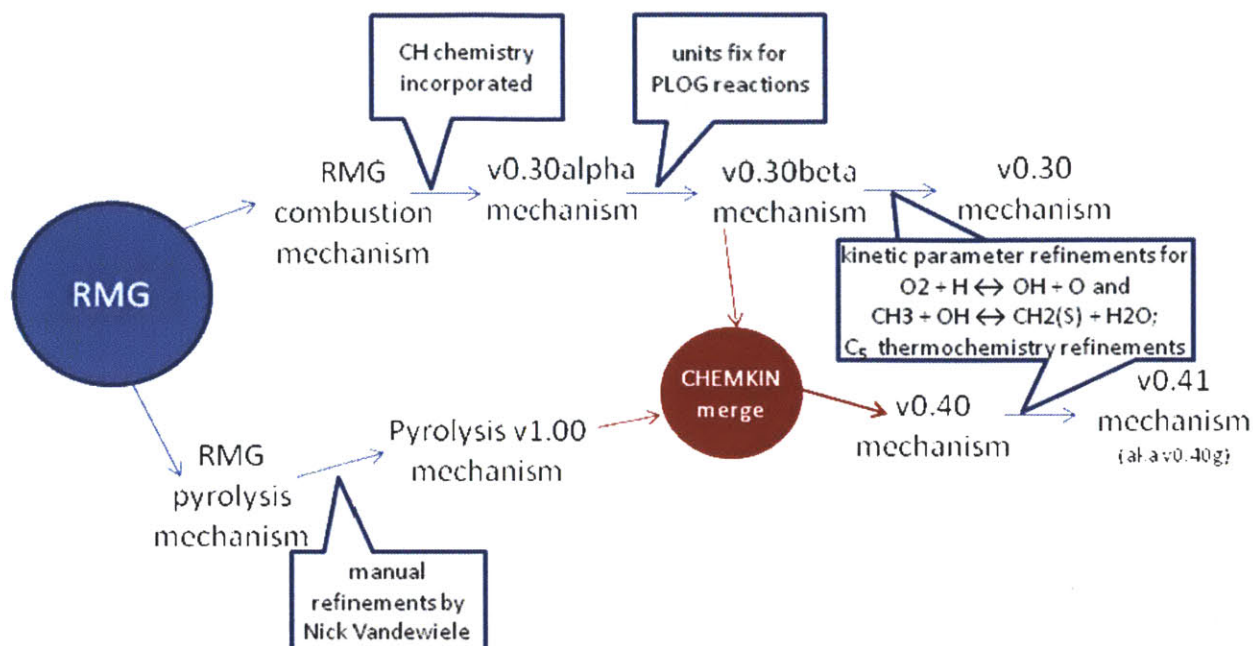


Figure 28. Summary of development of third-generation of RMG mechanisms (some intermediate versions and test versions omitted for clarity and brevity)

The v0.41 model was then compared with shock tube pyrolysis speciation data from Aerodyne Research, Inc. A significant fraction of the most predominant species from experiment (as identified by GC/MS) also appeared in the merged mechanism. Many of the species profiles agreed fairly well with the experimental results; the largest discrepancies involved cyclopentene, which was significantly overpredicted by the model, and propyne, which was significantly underpredicted by the model. To gain further insight into these discrepancies, sensitivity analysis for final cyclopentene and propyne mole fraction was performed at the 1290 K shock tube condition. The results for the most sensitive reactions are summarized in Table 23 and Table 24.

Table 23. Cyclopentene (C₅H₈) sensitivity analysis results for v0.41 mechanism at shock tube pyrolysis conditions of 1290 K

Reaction	Sensitivity coefficient $\left(\frac{d \ln y_{\text{cyclopentene}}}{d \ln A_j} \right)$
cyclopentene+H=C ₅ H _{7_A} +H ₂	-0.168
JP-10+H=JP10R8+H ₂	0.160
JP-10+H=JP10R6+H ₂	0.156
JP-10(+M)=C ₁₀ H _{16_A} (+M)	-0.152
JP-10(+M)=C ₁₀ H _{16_B} (+M)	-0.088

$C_5H_7_B + CH_4 = \text{cyclopentene} + CH_3$	-0.085
$\text{cyclopentene} + H = C_5H_7_C + H_2$	-0.085
$BR2 = C_{10}H_{16_A}$	-0.084
$H + CH_2CHCH_2(+M) = CH_2CHCH_3(+M)$	-0.084
$JP-10(+M) = C_{10}H_{16_C}(+M)$	-0.081

Table 24. Propyne (C_3H_4) sensitivity analysis results for v0.41 mechanism at shock tube pyrolysis conditions of 1290 K

Reaction	Sensitivity coefficient $\left(\frac{d \ln y_{\text{propyne}}}{d \ln A_j} \right)$
$JP-10(+M) = MA110(+M)$	0.254
$H + CH_2CHCH_2(+M) = CH_2CHCH_3(+M)$	-0.234
$\text{toluene} + H_2CCCH = \text{benzyl} + H_2CCCH_2$	0.213
$C_5H_4CH_3 + H_2CCCH = \text{fulvene} + H_3CCCH$	0.194
$C_7H_8_A = \text{toluene}$	-0.181
$H + H_2CCCH = H_3CCCH$	0.178
$H_2CCHCHCCH_2 = H_3CCCH + C_2H_2$	0.109
$C_7H_8_B + CH_3 = \text{benzyl} + CH_4$	0.099
$C_7H_8_C = C_7H_8_D =$	0.092
$H + H_2CCHCH_2 = H_2 + H_2CCCH_2$	0.085

The cyclopentene sensitivity analysis results, particularly second through fifth top reactions, are suggestive of an interesting tradeoff; the signs of the sensitivity coefficients suggest that bimolecular pathways leading to $C_{10}H_{15}$ radicals tend to promote formation of cyclopentene, while unimolecular pathways involving ring-opening of JP-10 on the $C_{10}H_{16}$ potential energy surface tend to inhibit cyclopentene formation. More generally, these sensitivity results point to the important role of the initial decomposition reactions, along with their associated kinetic parameters, in determining product distribution; JP-10 and initial decomposition products appear quite high on the ranked lists for both cyclopentene and propyne, despite the fact that these species are several elementary steps removed from the parent JP-10 molecule.

6.2.4 4th generation mechanisms: refinement of chemistry based on experimental comparisons and creation of a new comprehensive (pyrolysis+combustion) mechanism

Several refinements to the JP-10 system kinetic and thermodynamic libraries were made, mainly based on experience with third generation mechanism development and attempts to refine C₅ chemistry (based on findings with the merged model, v0.41). Also, with the connectivity checking feature of the QMTP system now well tested, connectivity checking was set to “confirm”. A new library (“*Robinson*”) had been developed by Nick Vandewiele for modeling indene formation based on the kinetic parameters for $C_5H_5 + C_5H_5 \leftrightarrow$ indenyl + CH₃ used by Robinson and Lindstedt¹¹⁰; this library was incorporated as a reaction library used during mechanism generation for these newest mechanism generation efforts to supplement the libraries discussed in the previous section. Refinements to the source code included work to improve aromaticity perception with the assistance of Nick Vandewiele.

As with the third-generation combustion mechanisms, the fourth generation combustion mechanisms were generated with pressure dependence turned on and at multiple conditions (as described previously). Several approaches were used for setting the mechanism enlargement thresholds and target criterion. In one approach, a mechanism was generated using a goal JP-10 conversion of 95% and a mechanism enlargement threshold of 0.20. This run (termed “0.50a_S1”; “S1”= Stage 1) completed (for all 18 reaction systems), resulting in a raw mechanism with 154 species and 2740 reactions. The resulting set of 154 species was subsequently turned into a bare seed mechanism (no reactions) that was used in a new mechanism generation run with a goal JP-10 conversion of 99.9% and a looser mechanism enlargement threshold of 0.50. This run (termed “0.50az”) completed (for all 18 reaction systems) resulting in a raw mechanism with 165 species and 3188 reactions.

Subsequent post-processing is summarized in Figure 29.

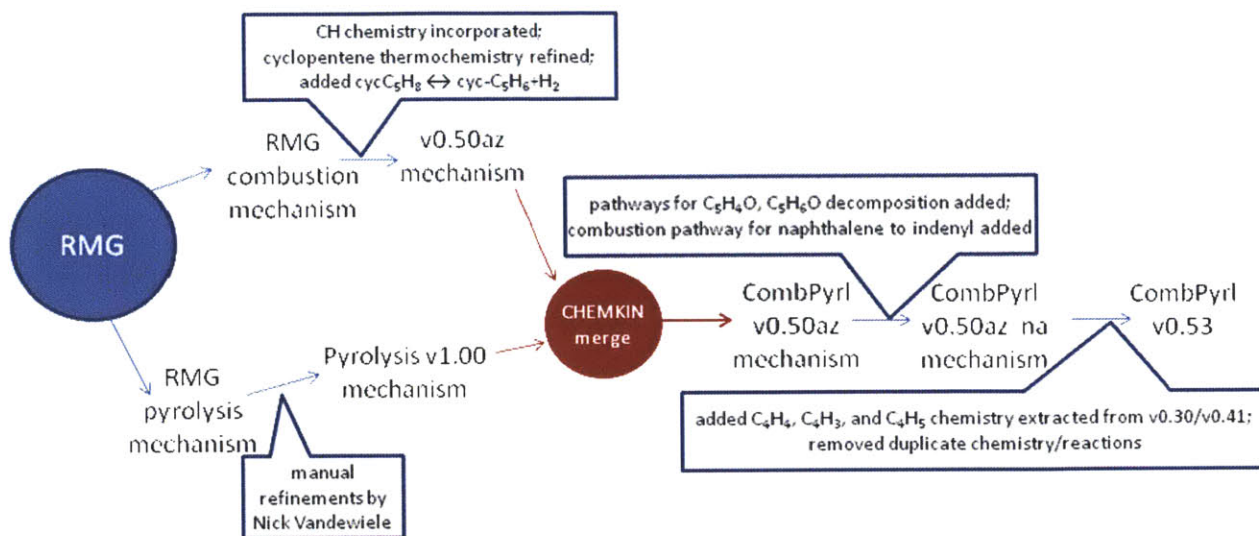


Figure 29. Summary of development of fourth-generation of RMG mechanisms (some intermediate versions and test versions omitted for clarity and brevity); in this case, unlike previous generations, the RMG combustion mechanism is the product of a two-stage approach, wherein a set of species generated using an earlier mechanism generation run was used as a seed for a new mechanism generation run

As in previous generations, initial post-processing included incorporation of CH chemistry. Additionally, in an attempt to address the apparent overprediction of cyclopentene identified in v0.41, a reaction, $\text{cyc-C}_5\text{H}_8 \leftrightarrow \text{cyc-C}_5\text{H}_6 + \text{H}_2$ was incorporated with forward kinetic parameters from experimental work by Lewis *et al.*;¹¹¹ the reaction was treated reversibly here to maintain thermodynamic consistency. Also, since an *ad hoc* ring correction for cyclopentene thermochemistry was available, the thermochemistry block for cyclopentene based on PM3 was replaced with “exact” group-additivity-based values; this had the effect of raising the Gibbs free energy of cyclopentene by about 4 kcal/mol at 1500 K.

The Pyrolysis v1.00 model was merged with our v0.50az model using the Mechanism Merge utility of the Reaction Workbench of CHEMKIN-MFC to produce a merged “CombPyr1 v0.50az” model. As in the merge described previously, the combustion mechanism was treated as master and the pyrolysis mechanism was treated as donor such that the master kinetic and thermodynamic parameter values are used in cases of conflict. A summary of the number of species and reactions associated with the mechanism merge appears in Table 25.

Table 25. Summary statistics for a fourth-generation mechanism merge (master and donor values do not sum to merged value due to mechanism overlap)

Mechanism	Number of species	Number of reactions
v0.50az (master)	168	3203
Pyrolysis v1.00 (donor)	838	11122
CombPyrl v0.50az (merged mechanism)	930	13633

Further testing of the CombPyrl v0.50az mechanism in collaboration with Nathan Yee suggested that the model exhibited accumulation of significant amounts of naphthalene, 2,4-cyclopentadien-1-one (C_5H_4O), and 2-cyclopenten-1-one (C_5H_6O) under approximately stoichiometric combustion conditions. Therefore, working with Nathan Yee, additional pathways were researched and added to the reaction model to form the “CombPyrl v0.50az_na” mechanism; this process also involved the addition of four new species. In particular, oxygen-mediated pathways from naphthalene to indenyl were incorporated, using kinetic parameters from Mati *et al.*,¹¹² a three-parameter fit from the kinetics.nist.gov database, and parameter estimates based on work by Frank *et al.*¹¹³ For 2,4-cyclopentadien-1-one, the reaction $C_5H_4O \leftrightarrow 2 C_2H_2 + CO$ was added, with kinetic parameters from Emdee *et al.*¹¹⁴ Finally, decomposition of 2-cyclopentene-1-one was modeled by adding a rapid equilibration reaction between 2-cyclopenten-1-one and 3-cyclopenten-1-one (effectively assuming that the two isomers are in rapid equilibrium under the conditions of interest); then, 3-cyclopenten-1-one depletion was modeled using a reaction decomposing this species into 1,3-butadiene and carbon monoxide with kinetic parameters obtained by a three-parameter fit to literature parameters for the reaction in the kinetics.nist.gov database.

A comparison of simulations results using CombPyrl v0.50az_na with shock tube pyrolysis data from Aerodyne exhibited a significant underprediction of 1-buten-3-yne (C_4H_4) and, to a lesser extent, 1,3-butadiyne (C_4H_2); these large discrepancies had not been observed in simulations using the v0.41 model. Therefore, C_4H_4 , C_4H_3 , and C_4H_5 chemistry was extracted from v0.30/v0.41 and merged into the CombPyrl 0.50az_na model. In conjunction with removal of duplicate reactions (one due to a CHEMKIN merge bug and three others that were effective duplicates arising as an artifact from merging pressure-dependent and non-pressure-dependent mechanisms), these updates defined the CombPyrl v0.53 model.

6.2.4.1 Ignition delay

The CombPyr1 v0.53 mechanism represents the most advanced JP-10 model considered here. Luwi Oluwole performed simulations with the mechanism to model the behavior of JP-10 ignition at conditions considered experimentally by Davidson *et al.*,⁷⁸ Colket and Spadaccini,^{87b} and Mikolaitis *et al.*^{87a} A comparison of simulation results with experimental data is shown in Figure 30.

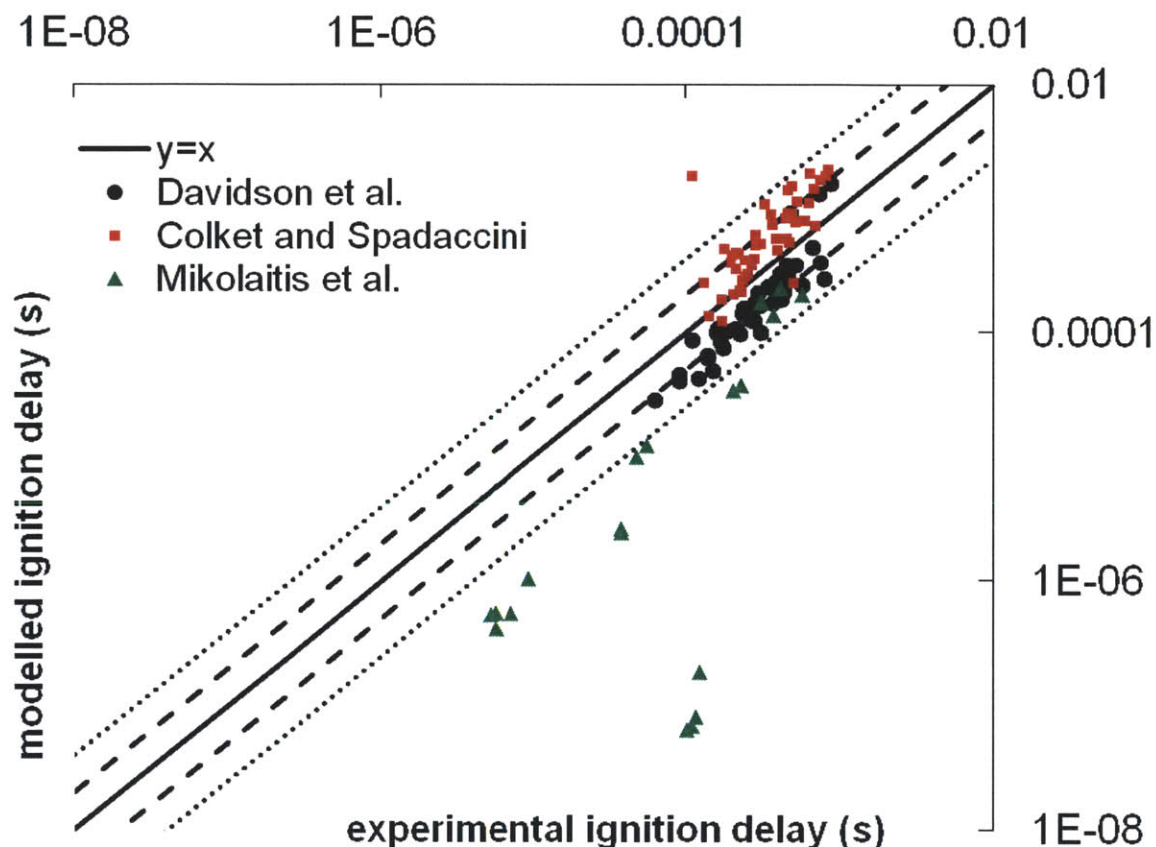


Figure 30. Parity plot of simulated ignition delay (as determined by the highest peak in the product of the mole fractions for C_2H and O_2) using CombPyr1 0.53 model versus experimental ignition delay results from Davidson *et al.*,⁷⁸ Colket and Spadaccini,^{87b} and Mikolaitis *et al.*^{87a}; the inner dotted lines correspond to a discrepancy of a factor of two and the outer dotted lines correspond to a discrepancy of a factor of four. The results show that at the majority of experimental conditions, the modeled ignition delay is within a factor of four of the experimental value, and often within a factor of two. The exceptions are a single condition in the Colket and Spadaccini dataset, where the modeled ignition delay is noticeably longer than the experimental value, and the conditions for the Mikolaitis *et al.* dataset at higher temperature, where the modeled ignition delay is significantly faster than the experimental value; this is particularly true at the four highest temperatures (with

only one CH peak from experiment). (It should be noted that Mikolaitis *et al.* observed two CH emission peaks at intermediate temperatures, and in these cases, the shorter ignition time is plotted in the figure above). The Mikolaitis dataset is also at higher pressures (above 10 atm) than the other datasets so this could be a factor in the apparent diminished performance of the model at these conditions. Considering smaller discrepancies, it is interesting to note that the points lying just above the parity line in the plot, including three Davidson points and most of the Colket and Spadaccini dataset tend to be associated with elevated pressures (generally above 4 atm) and reduced temperatures (below 1400 K).

Global sensitivity analysis on ignition delay predictions for this model was performed by Luwi Oluwole using the RS-HDMR method at the Davidson *et al.* conditions with starting temperature 1527K at 1.15 atm (with assistance from Nathan Yee); the eleven reactions determined through this analysis to be responsible for greater than 0.5% of the variation in ignition delay are summarized in Table 26.

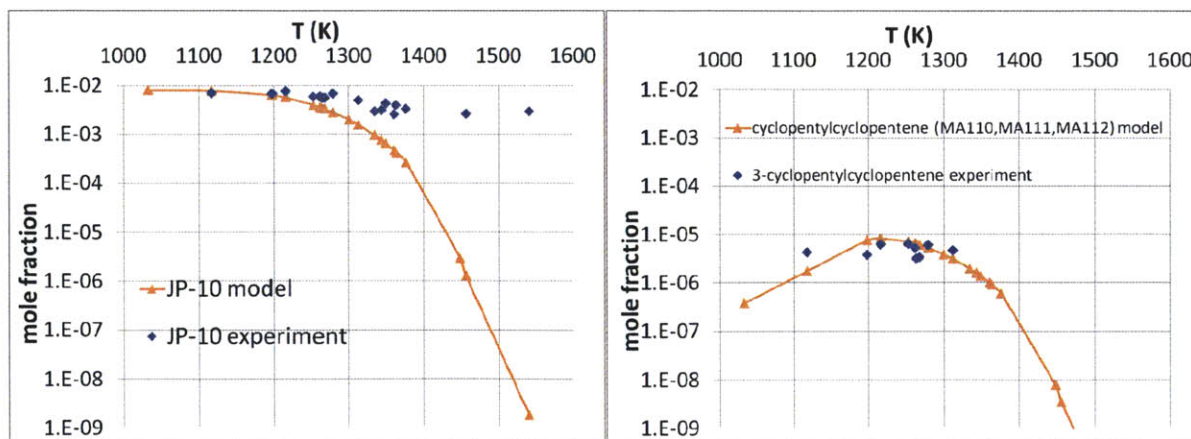
Table 26. Results of RS-HDMR ignition delay sensitivity analysis of the CombPyr1 v0.53 mechanism at Davidson et al. conditions with starting temperature 1527K at 1.15 atm

Reaction	Fraction of variation in ignition delay
$O_2+H=OH+O$	69.3%
$H_2CCCHO+O(+M)=H_2CCCH+O_2(+M)$	5.7%
$C_2H_2+O=HCCO+H$	3.8%
$H+H_2CCCH=H_3CCCH$	1.5%
$C_2H_2+H_2CCCH=HCCHCHCCH_2$	1.4%
cyclopentadienyl+H(+M)=cyclopentadiene(+M)	1.4%
$H_2O+H=H_2+OH$	1.4%
$O_2+HCCO=2CO+OH$	1.3%
$H+CH_3(+M)=CH_4(+M)$	0.7%
JP-10(+M)=C ₁₀ H ₁₆ _A(+M)	0.7%
$HCO+M=H+CO+M$	0.5%

The RS-HDMR sensitivity analysis results again demonstrate the important role of rate parameters for $O_2 + H \leftrightarrow OH + O$; in this case, however, the refinements based on work by Hong *et al.*¹⁰⁸ have been used in the model. The sensitivity analysis also points to the importance of kinetic parameters for propargyl and ketylenyl chemistry. Additionally, reactions involving hydrogen addition to cyclopentadienyl and chemically-activated ring opening of JP-10 also appear on the list.

6.2.4.2 Pyrolysis speciation

Although the model was not specifically designed to be applied to pyrolysis conditions, comparison against experimental pyrolysis data can also provide indications of model strengths and limitations. Shock tube data was acquired by Aerodyne Research, Inc., particularly David Lewis and Robin Edwards, at a number of shock temperatures. The mixture was modeled as an isobaric, adiabatic batch reactor at 7 atm starting with 0.8 mole percent JP-10 in argon with 0.5 millisecond residence time; initial temperature was determined based on shock speed measurements determined at the time of the experiment; the products were collected and later analyzed with two GC/MS instruments, one with a column designed to separate the heavier (C_5+) components and one with a column designed to separate lighter components. The results for the identified peaks were then compared with the final composition of the model simulation; a comparison of experimental results with simulated results for modeled species with non-zero concentration is shown in Figure 31, Figure 32, Figure 33, and Figure 34.



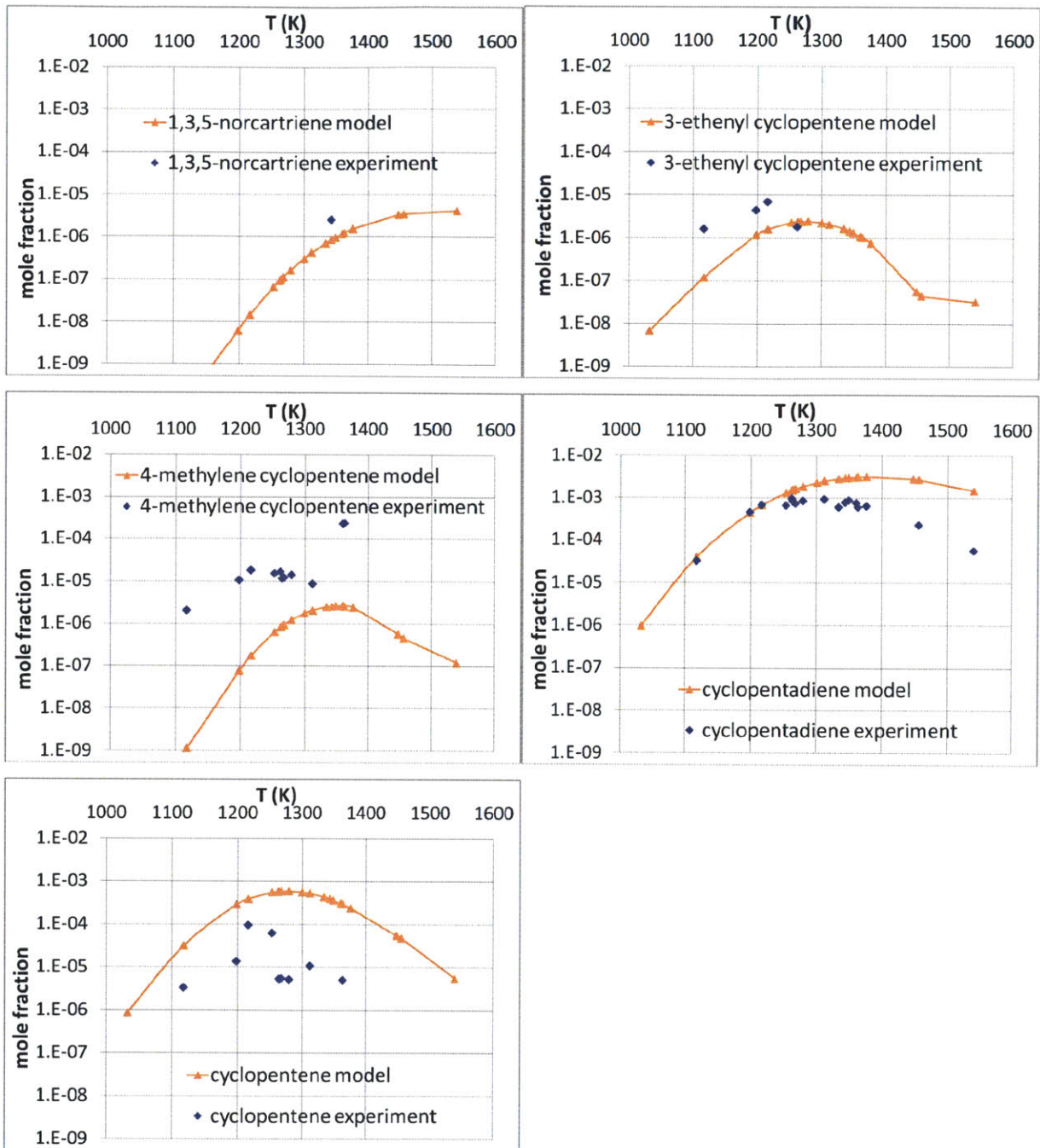


Figure 31. Comparison of experimental shock tube pyrolysis results with simulations based on CombPyrl v0.53 for cyclic species.

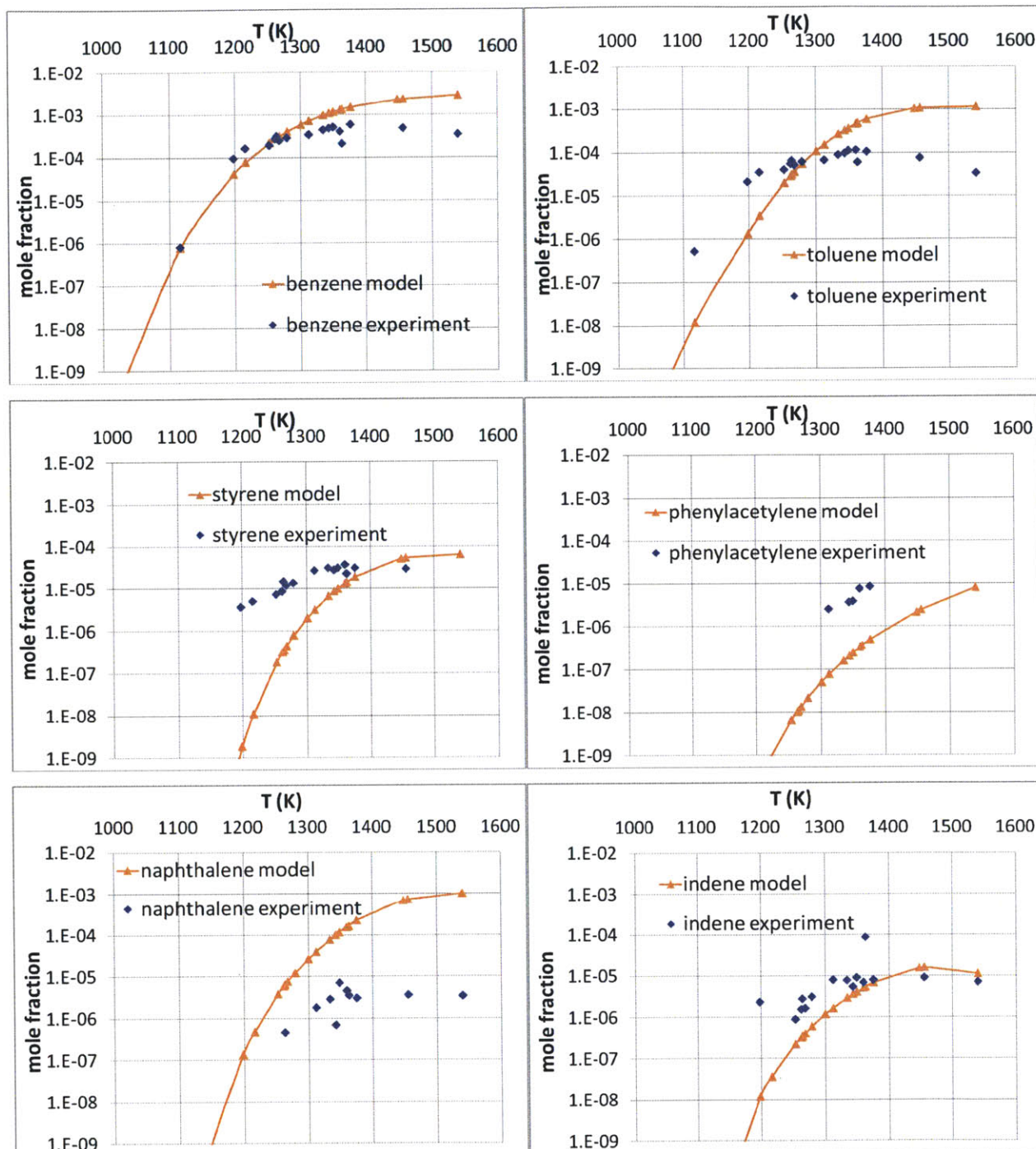


Figure 32. Comparison of experimental shock tube pyrolysis results with simulations based on CombPyrl v0.53 for aromatic species.

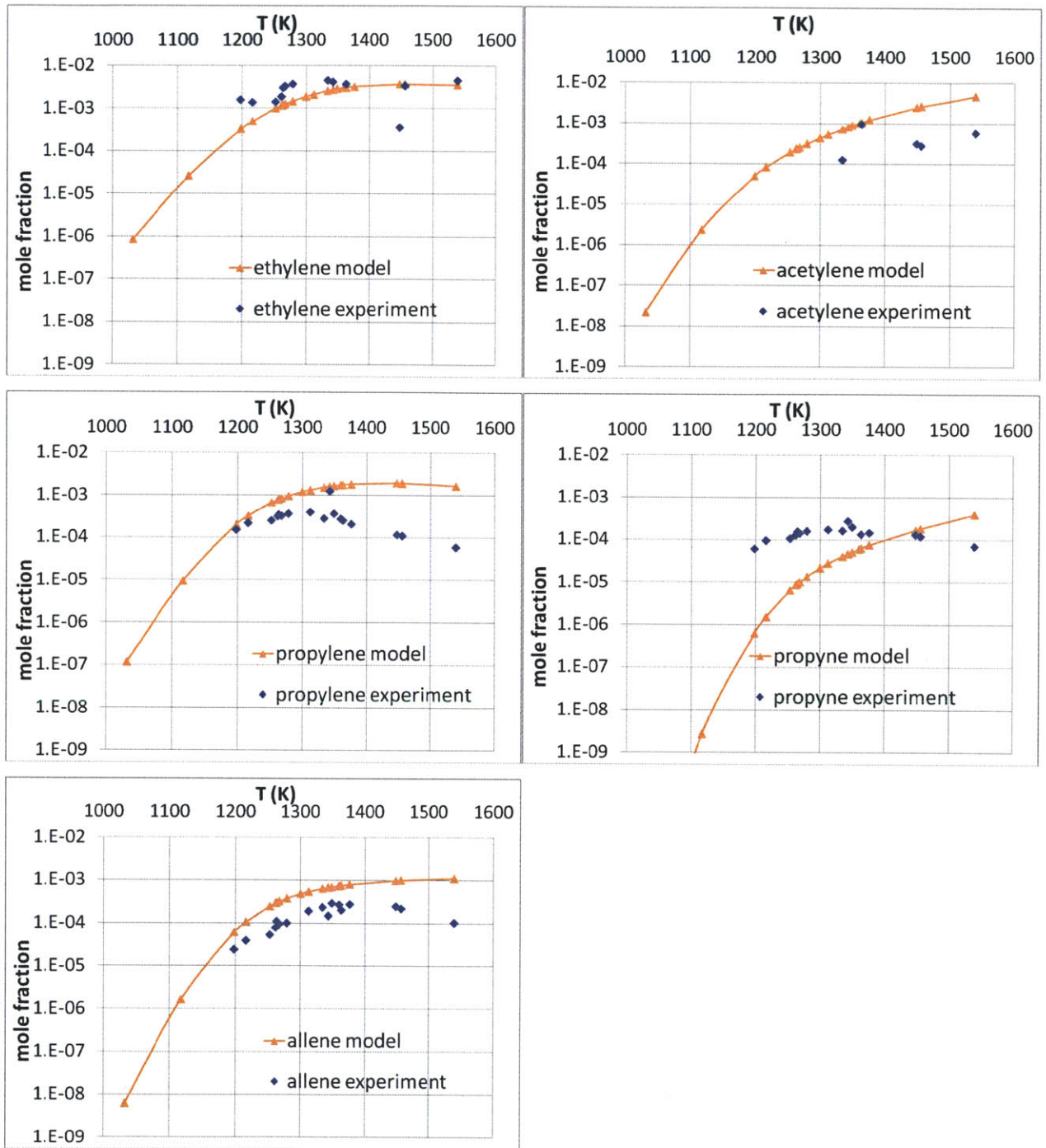


Figure 33. Comparison of experimental shock tube pyrolysis results with simulations based on CombPyr1 v0.53 for C_2-C_3 species.

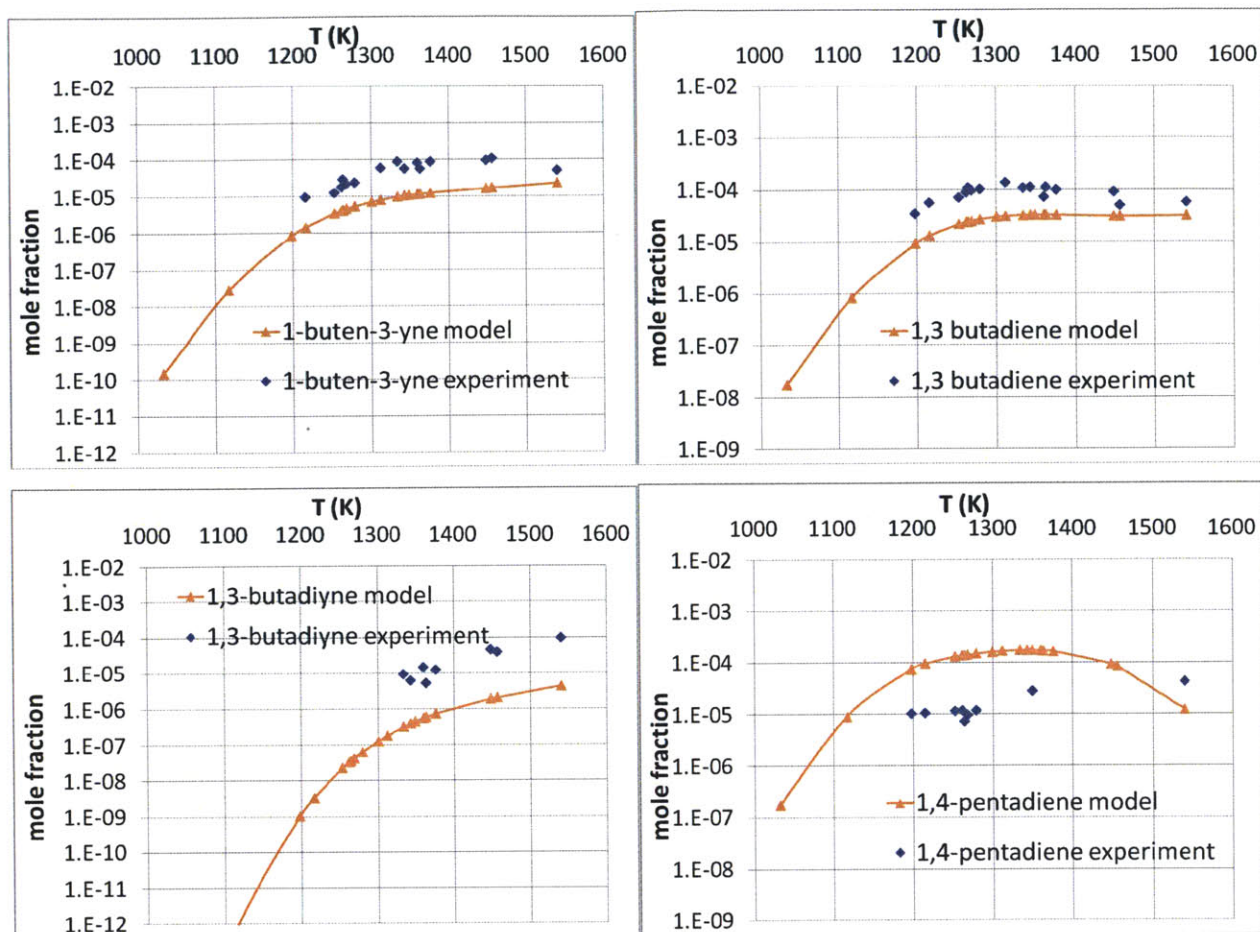


Figure 34. Comparison of experimental shock tube pyrolysis results with simulations based on CombPyrl v0.53 for C₄–C₅ species.

The comparison plots for cyclic species in Figure 31 show significantly higher experimental JP-10 concentration than predicted by the model at higher shock temperatures; this discrepancy has been attributed by the David Lewis (designer of the experimental shock tube apparatus) to a region of relatively cool, unreacted mixture near the walls of the shock tube. The second panel in this figure shows modelled results for cyclopentylcyclopentenes (three modelled species lumped together) with a species nominally identified via mass spectrum as 3-cyclopentylcyclopentene; the comparison is quite favorable, with the model capturing the presence of these species at intermediate and low temperatures and agreeing essentially quantitatively within the scatter of the experimental results. The model also captures the qualitative trend in cyclopentadiene results with temperature, with the highest concentrations at intermediate temperatures. Cyclopentene is apparently overpredicted by the model (though the results here represent an improvement over earlier model versions which exhibited more

substantial overprediction). On the other hand, 4-ethylene cyclopentene is apparently underpredicted by the model. Experimental results for 1,3,5-norcaradiene and 3-ethenyl cyclopentene are more limited, but appear to coincide with simulated results within roughly an order of magnitude.

Comparison plots for aromatic species in Figure 32 show that model and experiment tend to agree on a general trend to higher aromatic content at higher temperatures. The most noteworthy discrepancies include an apparent overprediction of naphthalene and toluene at high temperatures. On the other hand, indene appears to be captured well by the model.

Comparison plots for C₂-C₃ species are shown in Figure 33; the results include ethylene, which appears to be captured quite well by the model, with quantitative agreement (given the scatter in experimental results). Comparison plots for C₄-C₅ species are shown in Figure 34. The model seems to tend to underpredict the C₄ species, but captures the temperature trends for these species well.

Overall, comparison is fairly good considering the uncertainties in both experimental results and in model parameters. The model appears to capture many of the qualitative trends from experiment, and predicted mole fractions are often within an order of magnitude of experimental results.

To obtain insight into the key chemistry, sensitivity analysis was performed, with assistance from Nathan Yee, for select species at either 1198 K or 1540 K. The results for the five most sensitive reactions are presented in Table 27 through Table 37.

Table 27. Cyclopentene (C₅H₈) sensitivity analysis results for CombPyrl v0.53 mechanism at shock tube pyrolysis conditions of 1198 K

Reaction	Sensitivity coefficient $\left(\frac{d \ln y_{\text{cyclopentene}}}{d \ln A_j} \right)$
JP-10+H=JP10R8+H ₂	0.503
JP-10+H=JP10R5+H ₂	-0.237
JP-10(+M)=MA110(+M)	0.206
C ₁₀ H ₁₅ _A=C ₁₀ H ₁₅ _B	-0.183
JP-10=BR3	0.150

Table 28. Propyne (C₃H₄) sensitivity analysis results for CombPyrl v0.53 mechanism at shock tube pyrolysis conditions of 1198 K

Reaction	Sensitivity coefficient $\left(\frac{d \ln y_{\text{propyne}}}{d \ln A_j} \right)$
JP-10(+M)=MA110(+M)	0.879
CH ₂ CHCHCH ₃ +H ₂ CCCH=CH ₂ CHCHCH ₂ +H ₂ CCCH ₂	0.650
CH ₂ CHCH ₂ +H ₂ CCCH=2 H ₂ CCCH ₂	-0.273
CH ₂ CHCH ₂ +CH ₂ CHCHCH ₂ = H ₂ CCCH ₂ + CH ₂ CHCH ₂ CHCH ₂	0.216
2 CH ₂ CHCH ₂ = CH ₂ CHCH ₃ + H ₂ CCCH ₂	0.203

Table 29. Styrene sensitivity analysis results for CombPyrl v0.53 mechanism at shock tube pyrolysis conditions of 1198 K

Reaction	Sensitivity coefficient $\left(\frac{d \ln y_{\text{styrene}}}{d \ln A_j} \right)$
JP-10(+M)=MA110 (+M)	1.418
C ₈ H ₈ _A+H=styrene+H	0.926
CH ₂ CHCHCH ₃ +H ₂ CCCH=CH ₂ CHCHCH ₂ +H ₂ CCCH ₂	0.725
cyclopentadienyl+H ₂ CCCH= C ₈ H ₈ _B	0.724
C ₈ H ₈ _B=C ₈ H ₈ _A	0.695

Table 30. Benzene sensitivity analysis results for CombPyrl v0.53 mechanism at shock tube pyrolysis conditions of 1540 K

Reaction	Sensitivity coefficient $\left(\frac{d \ln y_{\text{benzene}}}{d \ln A_j} \right)$
JP-10(+M)= C ₁₀ H ₁₆ _C (+M)	0.186
JP-10(+M)=MA110(+M)	-0.086
JP-10+H=JP10R6+H ₂	-0.080
C ₇ H ₁₀ + CH ₂ CHCH ₂ =C ₁₀ H ₁₅ _A	0.078
C ₅ H ₇ _A+ CH ₂ CHCHCH ₂ = C ₁₀ H ₁₅ _A	-0.077

Table 31. Toluene sensitivity analysis results for CombPyr1 v0.53 mechanism at shock tube pyrolysis conditions of 1540 K

Reaction	Sensitivity coefficient $\left(\frac{d \ln y_{\text{toluene}}}{d \ln A_j} \right)$
JP-10(+M)=C ₁₀ H ₁₆ _A(+M)	0.809
JP-10=BR3	-0.207
C ₉ H ₁₃ +CH ₃ (+M)= C ₁₀ H ₁₆ _A (+M)	-0.169
C ₁₀ H ₁₆ _E= C ₁₀ H ₁₆ _A	0.121
JP-10(+M)=MA110(+M)	-0.097

Table 32. Cyclopentadiene (C₅H₆) sensitivity analysis results for CombPyr1 v0.53 mechanism at shock tube pyrolysis conditions of 1540 K

Reaction	Sensitivity coefficient $\left(\frac{d \ln y_{\text{cyclopentadiene}}}{d \ln A_j} \right)$
cyclopentadiene+H=cyclopentadienyl+H ₂	-0.362
JP-10(+M)= C ₁₀ H ₁₆ _A (+M)	-0.330
2 cyclopentadienyl=naphthalene+2 H	-0.249
JP-10+H=JP10R6+H ₂	0.135
JP-10(+M)= C ₁₀ H ₁₆ _C (+M)	-0.105

Table 33. Naphthalene sensitivity analysis results for CombPyr1 v0.53 mechanism at shock tube pyrolysis conditions of 1540 K

Reaction	Sensitivity coefficient $\left(\frac{d \ln y_{\text{naphthalene}}}{d \ln A_j} \right)$
2 cyclopentadienyl=naphthalene+2 H	0.397
cyclopentadiene+H=cyclopentadienyl+H ₂	0.249
C ₂ H ₂ +H ₂ CCCH=HCCHCHCCH ₂	-0.147
cyclopentadiene+H(+M)=C ₅ H ₇ _C(+M)	0.142
JP-10(+M)= C ₁₀ H ₁₆ _A (+M)	-0.129

Table 34. Allene (H₂CCCH₂) sensitivity analysis results for CombPyr1 v0.53 mechanism at shock tube pyrolysis conditions of 1540 K

Reaction	Sensitivity coefficient $\left(\frac{d \ln y_{\text{allene}}}{d \ln A_j} \right)$
JP-10(+M)= C ₁₀ H ₁₆ _A (+M)	-0.333
C ₂ H ₃ +CH ₃ = CH ₂ CHCH ₂ +H	-0.223
CH ₂ CCH ₂ +H(+M)= CH ₂ CHCH ₂ (+M)	0.116
JP-10(+M)=MA110(+M)	0.114
CH ₂ CHCH ₂ +H ₂ CCCH=2 CH ₂ CCH ₂	0.111

Table 35. 1-buten-3-yne (HCCCHCH₂) sensitivity analysis results for CombPyr1 v0.53 mechanism at shock tube pyrolysis conditions of 1540 K

Reaction	Sensitivity coefficient $\left(\frac{d \ln y_{1\text{-buten-3-yne}}}{d \ln A_j} \right)$
HCCHCHCH ₂ +CH ₃ = CH ₂ CHCHCHCH ₃	0.368
C ₁₀ H ₁₅ _C(+M)=JP10R5(+M)	0.286
JP10R5(+M)=C ₁₀ H ₁₅ _B(+M)	-0.281
JP-10+H=JP10R5+H ₂	0.237
HCCCHCH ₂ +H(+M)= HCCHCHCH ₂ (+M)	0.172

Table 36. 1,3-butadiyne (HCCCCH) sensitivity analysis results for CombPyr1 v0.53 mechanism at shock tube pyrolysis conditions of 1540 K

Reaction	Sensitivity coefficient $\left(\frac{d \ln y_{1,3\text{-butadiyne}}}{d \ln A_j} \right)$
C ₁₀ H ₁₅ _C (+M)=JP10R5(+M)	0.386
JP10R5(+M)=C ₁₀ H ₁₅ _B(+M)	-0.374
JP-10+H=JP10R5+H ₂	0.294
HCCCHCH+CH ₄ = HCCCHCH ₂ +CH ₃	0.290
H+ HCCCHCH ₂ =H ₂ +HCCCCH ₂	0.240

Table 37. Propylene (CH₂CHCH₃) sensitivity analysis results for CombPyr1 v0.53 mechanism at shock tube pyrolysis conditions of 1540 K

Reaction	Sensitivity coefficient $\left(\frac{d \ln y_{\text{propylene}}}{d \ln A_j} \right)$
JP-10(+M)= C ₁₀ H ₁₆ _A (+M)	-0.278
C ₂ H ₃ +CH ₃ = CH ₂ CHCH ₂ +H	-0.231
H+ CH ₂ CHCH ₃ =H ₂ + CH ₂ CHCH ₂	-0.174
JP-10(+M)= C ₁₀ H ₁₆ _C (+M)	0.160
H+ CH ₂ CHCH ₂ (+M)= CH ₂ CHCH ₃ (+M)	0.105

The three sensitivity analyses at low temperature point to the important role of the kinetic parameters for the unimolecular decomposition reaction, JP10 ↔ MA110 in determining product distribution. The sensitivity results also point to an interesting interplay between bimolecular channels in determining the concentration of cyclopentene; in particular, the signs of the sensitivity coefficients for the top two reactions suggest that abstraction of hydrogen from JP-10 to form JP10R8 tends to promote cyclopentene formation, whereas hydrogen abstraction at an alternative site to form JP10R5 tends to inhibit cyclopentene formation.

Sensitivity analyses at the high temperature conditions provide similar insight in the role of kinetic parameters in the model. For example, the sensitivity analyses for 1-buten-3-yne suggest an important role for the pathway JP-10 ↔ JP-10R5 ↔ C₁₀H₁₅_C in forming this species in the model. Looking at sensitivity coefficients for the reaction JP-10 ↔ C₁₀H₁₆_A suggests that this pathway tends to promote formation of toluene at the expense of cyclopentadiene, naphthalene, allene, and propylene. Similarly, results suggest that JP-10 ↔ MA110 tends to promote formation of allene and styrene at the expense of toluene and benzene; results suggest that JP-10 ↔ C₁₀H₁₆_C tends to promote formation of propylene and benzene at the expense of cyclopentadiene. These branching effects are evident at downstream steps, as well, with kinetic parameters for two competing decomposition pathways for C₁₀H₁₅_A affecting the amount of benzene formed; the sensitivity coefficients suggest that the pathway involving loss of three carbons tends to promote benzene formation whereas the pathway involving loss of five carbons tends to inhibit it. The sensitivity results also confirm the important role of the kinetic parameters

for the global reaction $2 \text{ cyclopentadienyl} \leftrightarrow \text{ naphthalene} + 2 \text{ H}$ from Murakami *et al.*¹⁰⁵ in determining naphthalene composition.

In addition to providing insight into the complex underlying chemistry of the model, these results point to the most promising routes to further mechanism refinement. As many of these sensitivity analyses correspond to species that are notably underpredicted or overpredicted at the conditions considered, the kinetic parameters for the most-highly ranked reactions can be more closely scrutinized and refined as necessary to either refine the current model or begin a new iteration of mechanism generation.

Generally speaking, the sensitivity analyses point to the important role of initial decomposition reactions in determining the product distribution; the results suggest that it is important to accurately capture both the absolute rates and the relative rates (branching ratios) among these initial decomposition reactions. Refinement of the associated kinetic parameters presents a clear opportunity for further model improvement. Efforts in this direction are presented in Chapter 7.

6.2.4.3 Flux analysis

As another approach to mechanism analysis, flux diagrams can be constructed and analyzed at conditions of interest. With the assistance of Nathan Yee and Josh Allen, flux diagrams were produced using a tool related to the “Generate Flux Diagram” utility on the rmg.mit.edu website; the CombPyr1 v0.53 mechanism was simulated with isothermal, isobaric conditions for a stoichiometric mixture of JP-10 in air at 1 atm and 1600 K; the tool creates diagrams with arrow thickness and direction corresponding to flux magnitude and direction, respectively, and with circle thickness around species indicating concentration on a logarithmic scale.

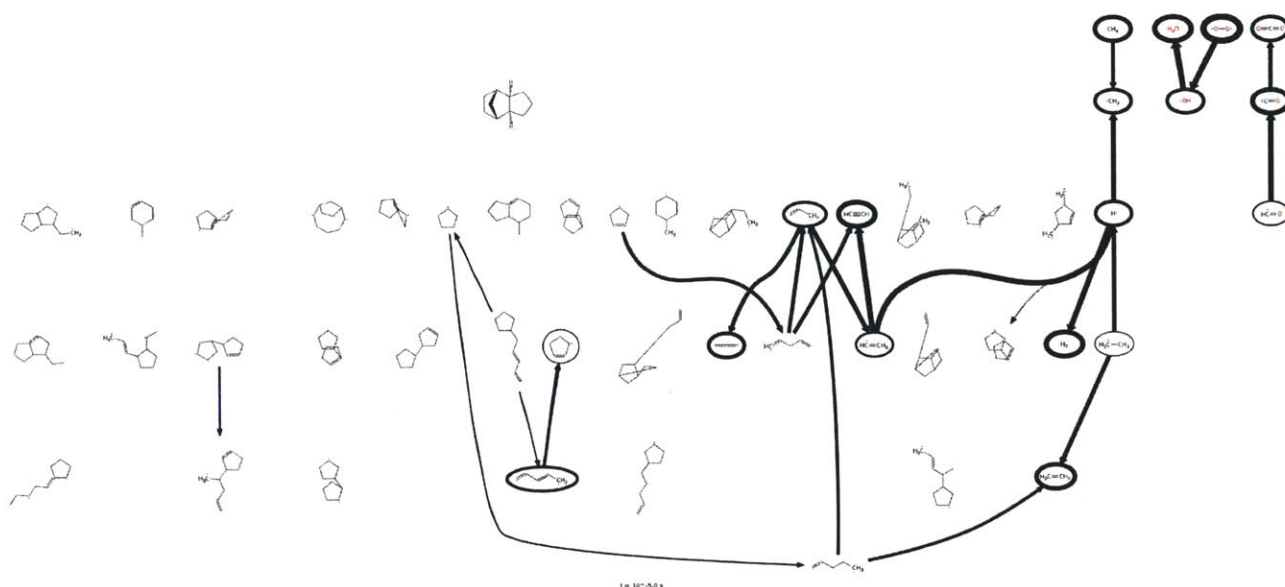


Figure 37. Flux diagram at $t=10^{-5}$ s, generated using a tool related to rmg.mit.edu’s “Generate Flux Diagram”. Diagram was constructed using the CombPyr1 v0.53 model for an isothermal, isobaric batch reactor with a stoichiometric mixture of JP-10 and air at 1 atm and 1600 K

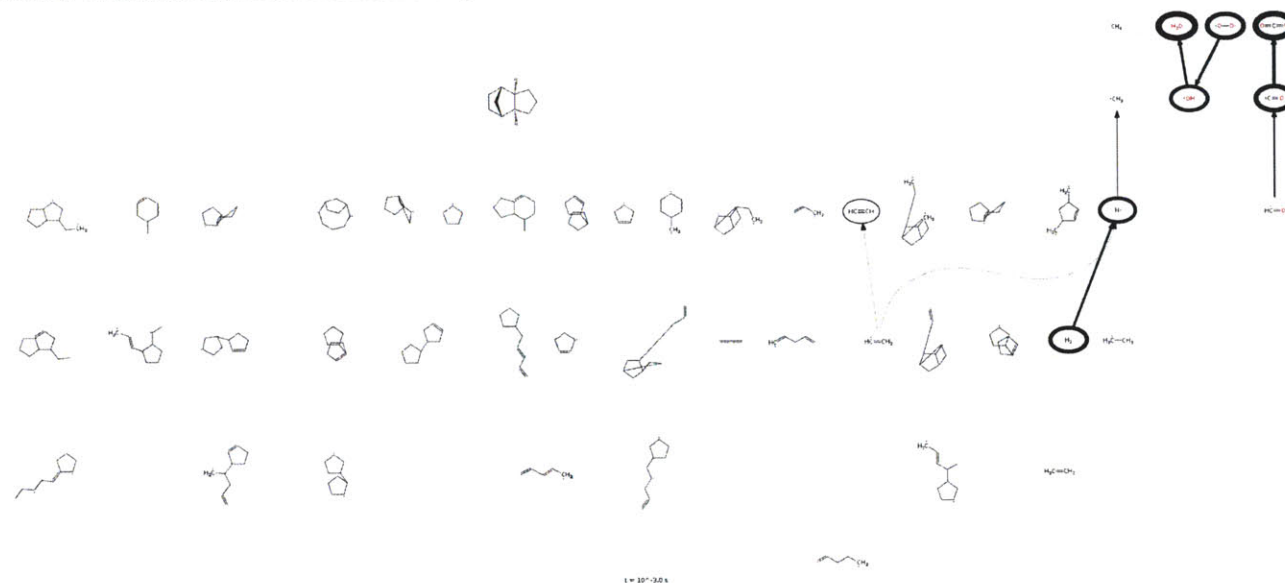


Figure 38. Flux diagram at $t=10^{-3}$ s, generated using a tool related to rmg.mit.edu’s “Generate Flux Diagram”. Diagram was constructed using the CombPyr1 v0.53 model for an isothermal, isobaric batch reactor with a stoichiometric mixture of JP-10 and air at 1 atm and 1600 K

Results at early time, corresponding to 10^{-16} s, are shown in Figure 35, and illustrate a number of initial decomposition steps to various C_{10} species.

Advancing to 10^{-7} s, Figure 36 shows many more active pathways. The flux analysis suggests that under these conditions, the main decomposition pathways appear to be unimolecular routes of the type considered by Herbinet *et al.* and discussed further in Chapter 7.

However, some bimolecular routes also appear in the diagram. Under intermediate- or high-temperature conditions (avoiding low-temperature peroxide chemistry), chemical intuition leads us to suspect the importance of the main (lower) pathway shown in Figure 39 as a key bimolecular route for the initial decomposition of JP-10.

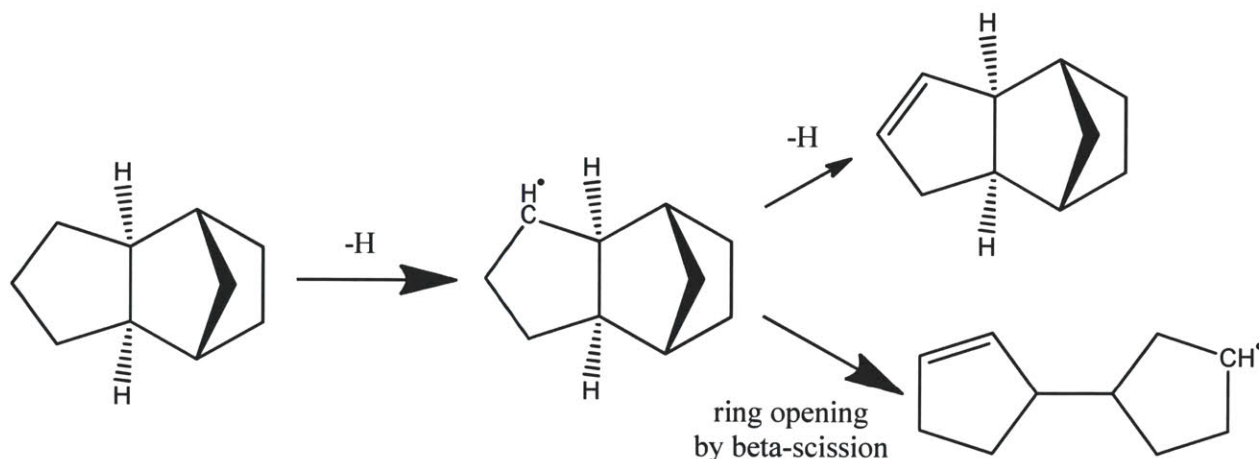


Figure 39. Selected bimolecular-mediated pathways for initial JP-10 decomposition; (the expected main pathway is depicted with large arrows)

In the scheme presented, a hydrogen is first abstracted from JP-10 at the site shown above; this site is preferred for energetic reasons as the resulting radical is the most stable among the six possible $C_{10}H_{15}$ radicals formed by hydrogen abstraction according to DFT¹¹⁵, our CBS-QB3 calculations, and a combined CBS-QB3/G3MP2B3 analysis¹¹⁶; the site is also favored by statistical considerations as there are four hydrogens (two of one variety and two of another) that can be abstracted to produce the radical.^{§§} This radical can lose another hydrogen to form a $C_{10}H_{14}$ alkene, which was observed as a minor product in flow reactor experiments by Nick Vandewiele. A more energetically favorable pathway appears to be unimolecular ring-opening by beta-scission; this pathway alleviates a significant source of ring-strain in the carbon framework of JP-10. The model captures this lower route, which is significant enough at the conditions considered here to appear toward the left of Figure 36. This route should play an even larger role at conditions with higher initial oxygen concentration or lower temperature. Indeed, a flux diagram (Figure 40) was generated for conditions suspected to be more conducive to bimolecular pathways (a temperature of 1300 K, twice the stoichiometric ratio of oxygen, and a

^{§§} There is one other “JP-10 radical” that can be formed by abstraction of four possible hydrogens, but this radical does not appear to be favored by energetic considerations.

pressure of 5 atm), which suggests a more critical role for this pathway (and similar bimolecular pathways).

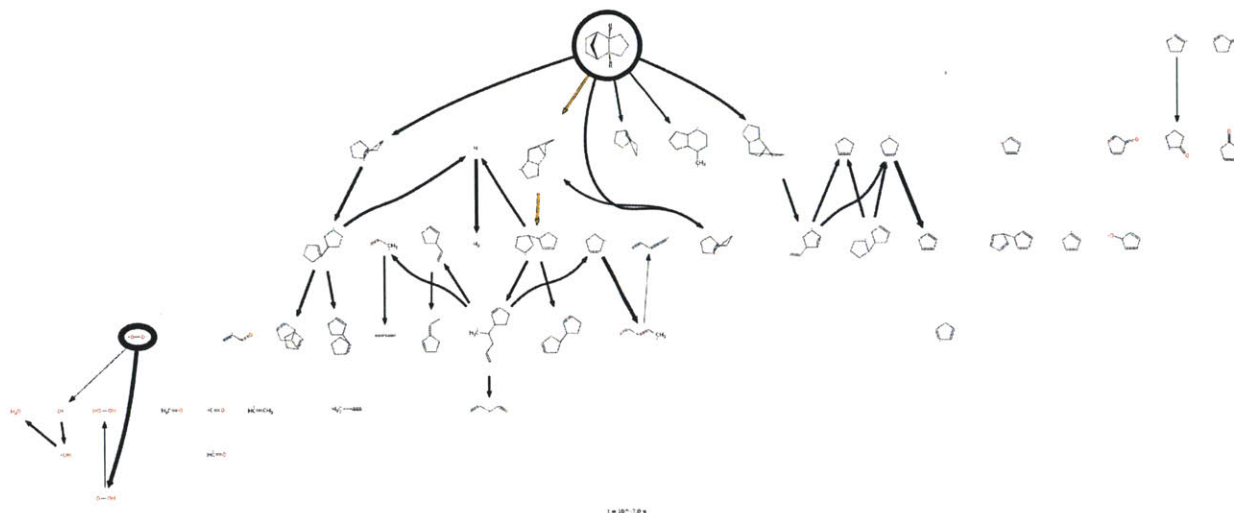


Figure 40. Flux diagram at $t=10^{-7}$ s, generated using a tool related to rmg.mit.edu's "Generate Flux Diagram". Diagram was constructed using the CombPyrl v0.53 model for an isothermal, isobaric batch reactor with JP-10 and twice the stoichiometric amount of oxygen, and nitrogen gas at 5 atm and 1300 K. The pathway highlighted in orange is the bimolecular pathway discussed in the text.

Advancing still further to 10^{-5} s at the originally considered conditions, Figure 37 shows that JP-10 is now essentially entirely gone, while a significant quantity of O_2 remains; this underscores the importance of pyrolysis chemistry in the initial decomposition of JP-10, even under oxidative conditions. Finally, advancing to 10^{-3} s, as shown in Figure 38, the diagram suggests significant oxygen depletion and the dominance of small molecule chemistry.

6.3 Model strengths, limitations, and next steps

In terms of comparing the model results with experimental results, the level of agreement is generally quite favorable, taking into consideration the uncertainties in both experimental results and in model parameters.

For ignition delay comparisons the level of agreement over a wide range of conditions is within a factor of four for most of the points, with a significant fraction of these lying within a factor of two.

The CombPyrl v0.53 model considers a number of important species that have been omitted from earlier JP-10 combustion models, including benzene, naphthalene, toluene, indene, and cyclopentylcyclopentenes. Even so, the pyrolysis speciation data also points to several

species with chemistry that is not considered by the model; this includes species such as xylene, 2-butene, 1,3,5-cycloheptatriene. One opportunity for further refinement in new generations of mechanism creation is the inclusion of chemistry for such species. It is possible that such refinements could improve predictions for other species as well; for example, it is possible that modeling xylene chemistry could account for the apparent overprediction of toluene at high temperatures by the CombPyr1 v0.53 model. The incorporation of appropriate chemistry from the literature through seed mechanisms or reaction libraries during the mechanism generation process could assist in the capturing of this chemistry. Additionally, changing mechanism generation parameters, particularly the tolerance for model enlargement, could allow automated consideration of some of these minor products. Of course, there are challenges involved, including the added computational expense of generating larger models; also, the resulting model would be more computationally expensive to simulate and analyze.

Another opportunity for refinement is conversion of the pyrolysis mechanism to a pressure dependent version. In addition to providing more reliable parameter estimates, this would also avoid effective duplicate reactions that can arise from merging pressure-dependent and non-pressure-dependent mechanisms. At present, the main obstacle is computational cost and the most straightforward remedy may simply be to allow for longer wall-time. Alternatively, this bottleneck could motivate future efforts to speed up the mechanism generation code.

It is also noted that the mechanism generation efforts described here have focused on high-temperature oxidation (1250 K and above); low-temperature oxidation would involve peroxide chemistry where uncertainties in parameter estimates can have a greater impact and the large number of additional species to consider creates added informatics burdens. While capturing this low-temperature oxidation chemistry would certainly be useful, the added challenges associated with this chemistry likely place an adequate low-temperature chemistry treatment out of reach of automated reaction mechanism generation efforts in the near term.

In terms of reaction kinetics and chemistry, it is the opinion of the author that the initial decomposition steps of JP-10 represent the biggest potential source of error and uncertainty in the model; analysis of the model has demonstrated the important role of these initial decomposition steps in determining product distribution from JP-10 decomposition. Efforts to investigate these initial decomposition processes are discussed in the following chapter, along with additional opportunities to better characterize this chemistry through further research. The

pathway to C₁₀H₁₆_C is a good example of opportunities for improvement in this area; it would seem that the kinetic parameter estimation scheme currently used by RMG would not take into account the apparently significant ring strain and steric effects involved in producing this molecule by intramolecular disproportionation in BR2, and it is possible that the rate coefficient for this reaction is significantly overestimated.

In terms of methodology, the approach of iterative refinement demonstrated here offers a general framework that can be continued to produce new generations of JP-10 combustion models. Key features of this approach include 1) identification of mechanism limitations from earlier generations (*e.g.* by sensitivity analysis and comparison of model simulations with experiment), 2) a targeted effort to address these limitations through updates to the RMG input files, software, and/or database, and 3) regeneration of mechanism with RMG, incorporating these updates.

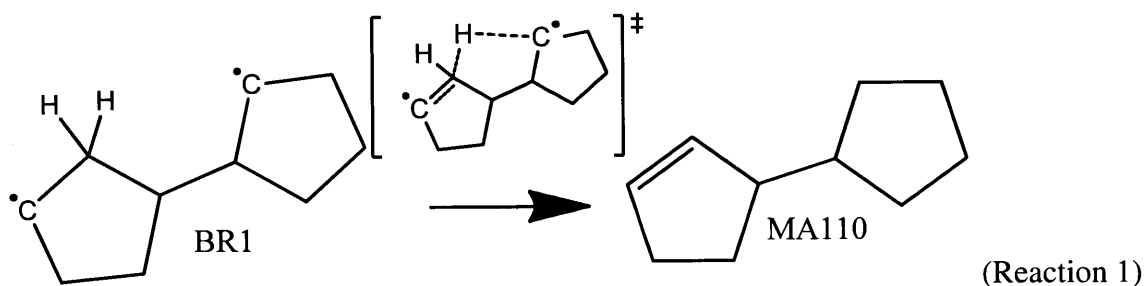
6.4 Conclusions

Several iterations of JP-10 combustion mechanism generation were performed using RMG, progressively incorporating more reliable chemistry and methods. The ability to generate a comprehensive, state-of-the-art model for JP-10 oxidation through utilization of the QMTP features of RMG demonstrates that the system described in previous chapters provide a practical and useful addition to the RMG utility, expanding the scope of mechanism generation without any substantial drawbacks.

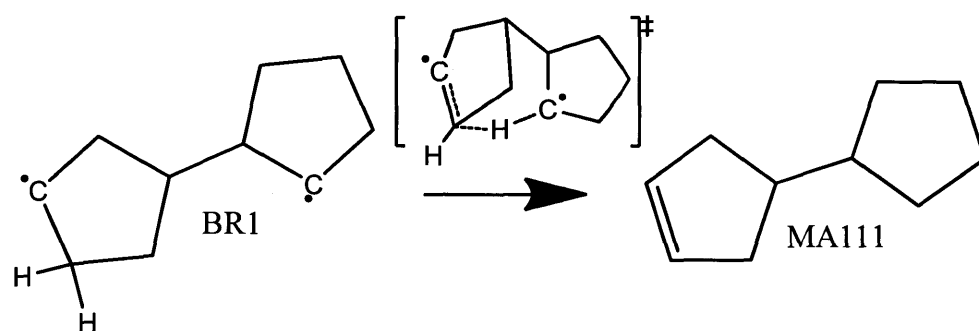
7 Chapter 7: Towards on-the-fly kinetic parameter estimation for improving intra-molecular disproportionation estimates^{***}

7.1 Introduction / motivation

Herbinet *et al.* investigated the suspected important role of biradicals in the initial phases of JP-10 decomposition.⁸⁴ In particular, they looked at the behavior of biradicals formed by carbon-carbon bond homolysis in JP-10. They discussed how these biradicals can participate in intramolecular disproportionation reactions (in which one radical site in the molecule abstracts a hydrogen from the carbon adjacent to a second radical site to form a carbon-carbon double bond) and identified the important role of these intramolecular disproportionation reactions in pyrolytic decomposition of JP-10. Two of the reactions considered by Herbinet *et al.* are Reaction 1 and Reaction 2, shown below. (The *BR/MA* notation of Herbinet *et al.* is followed.)



^{***} This chapter is based heavily on material published in collaboration with Piotr Piecuch and Jesse Lutz of Michigan State University in the International Journal of Chemical Kinetics. (Magoon, G. R.; Aguilera-Iparraguirre, J.; Green, W. H.; Lutz, J. J.; Piecuch, P.; Wong, H.-W.; Oluwole, O. O., Detailed chemical kinetic modeling of JP-10 (exo-tetrahydrodicyclopentadiene) high-temperature oxidation: Exploring the role of biradical species in initial decomposition steps. *Int J Chem Kinet* **2012**, *44* (3), 179-193.) Permission has been obtained from the publisher for reuse of this material in this thesis.



(Reaction 2)

Two pathways through Reaction 1 were identified by Herbinet *et al.* and the pathway shown, with a five-membered ring transition state, was proposed to have the lowest activation energy. Herbinet *et al.* proposed the scheme shown below for estimating the activation energy for these intramolecular disproportionation reactions,

$$E_a = E_{a,ref} - SE_{ref\ cycle} + SE_{sub-polycycle} \quad \text{Eq. 4}$$

In Equation 1, E_a is the activation energy and $E_{a,ref}$ is the activation energy for a reference n -alkdiyl intramolecular disproportionation with the transition state with the same ring size. The two SE terms refer to the ring strain energy for the reference cycle and the sub-polycycle characteristic of the actual transition state. Using this approach to estimate the barriers, Herbinet *et al.* estimated rate constant expressions for Reaction 1 (with five-membered ring transition state) and Reaction 2 (with [3.2.1] bicyclic ring transition state) as shown below.

$$k_1(T) = 1.9 \times 10^{13} \text{ s}^{-1} \left(\frac{T}{1000 \text{ K}} \right)^{1.0} e^{\frac{-7.75 \text{ kcal/mol}}{RT}} \quad \text{Eq. 5}$$

$$k_2(T) = 1.9 \times 10^{13} \text{ s}^{-1} \left(\frac{T}{1000 \text{ K}} \right)^{1.0} e^{\frac{-19.85 \text{ kcal/mol}}{RT}} \quad \text{Eq. 6}$$

However, the intramolecular disproportionation family in RMG has been implemented using a first-order approximation of the Herbinet *et al.* scheme due to challenges in implementing the full Herbinet *et al.* scheme within the current RMG framework:

$$E_a = E_{a,ref} \quad \text{Eq. 7}$$

As described previously, a preliminary detailed JP-10 combustion mechanism was developed using RMG.^{87g} This initial (v0.10) mechanism was refined to the v0.19 mechanism, considered here; refinements include the addition of important cyclic-C₅ chemistry, as well as updates to thermodynamic and kinetic parameters based on *ab initio* calculations or the published literature; the refined (v0.19) mechanism includes 320 species reacting through 7740 elementary steps. With the exception of four BR1 pathways (where the estimates of Herbinet *et al.* were used), kinetic parameter estimates in the mechanism series from v0.10 to v0.19 for this reaction class were estimated on-the-fly by RMG, which, as discussed previously, uses a first-order approximation of the estimation scheme employed by Herbinet *et al.* where only the first term in their estimation procedure is retained.

In collaboration with Luwi Oluwole of Aerodyne Research, Inc., sensitivity analysis was performed with software based on SENKIN¹¹⁷ for a constant-pressure adiabatic ignition simulation with a stoichiometric JP-10/air mixture starting at 1500 K and 1 atm, using the refined (v0.19) mechanism. Following the approach used by Davidson *et al.* in their experiments,⁷⁸ ignition was modeled in our calculations as the time to peak CH composition. Sensitivity coefficients for each reaction were computed using the expression below.

$$Z_i = \left| \frac{\partial y'_{CH}(\tau_0)}{\partial \ln k_i} \right| \quad \text{Eq. 8}$$

When these sensitivity metrics were ranked, it was found that kinetic parameters for multiple intramolecular disproportionation reactions (including Reaction 1) appear amongst the top 2.5% (top 200) of all reactions in this large mechanism.^{†††} This suggests that ignition delay can be sensitive to the kinetic parameter estimates for these intramolecular disproportionation reactions. The kinetic parameters for these intramolecular disproportionation reactions are also likely to be

^{†††} Note that y'_{CH} in this equation is the rate of change in the CH mass fraction and τ_0 refers to the ignition time for the unperturbed system. The partial derivative is evaluated holding the thermochemistry constant, *i.e.* both the forward and reverse rate coefficients of reaction i are adjusted by the same amount so the equilibrium constant is not changed by the perturbation.

quite important in other aspects of the model besides ignition delay, such as speciation. For example, sensitivity analysis on a more recent JP-10 mechanism versions (v0.41) has suggested that the cyclopentene concentration for shock tube pyrolysis conditions of 1290 K is quite sensitive to these intramolecular disproportionation reactions; in particular, two unimolecular reactions of JP-10 involving chemically-activated paths involving intra-molecular disproportionation as a component are ranked 4th and 5th in terms of sensitivity coefficient; in fact, the results suggest that cyclopentene formation may be significantly impacted by a competition between hydrogen abstraction from the JP-10 fuel precursor and unimolecular reactions to relieve strain in the JP-10 fuel precursor, with the former favoring formation of cyclopentene and the latter inhibiting it. Therefore, it will be important in future mechanism development to more accurately estimate activation energies for these reactions, particularly in view of relatively small amount of research that has been performed in this area.

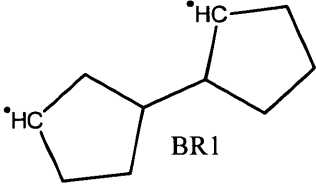
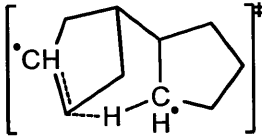
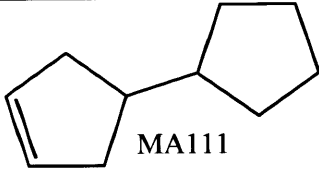
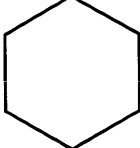
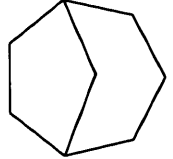
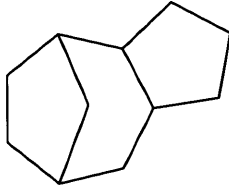
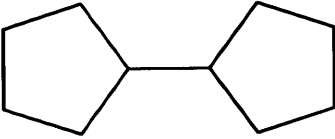
It was initially envisioned that the RMG framework could be adjusted to allow activation energy estimation procedures like the full approach of Herbinet *et al.* for intramolecular disproportionation to be implemented in RMG. Plans were constructed to design and implement a variation of the Herbinet *et al.* approach using the availability three-dimensional molecular geometries and on-the-fly calculations based on them made possible by separate efforts to improve RMG's thermochemical parameter estimates discussed previously. The proposed variant of the Herbinet *et al.* approach is summarized below.

$$E_a = E_{a,ref} - SE_{ref\ cycle} + SE_{polycycle} - SE_{saturated\ reactant} \quad \text{Eq. 9}$$

The first two terms on the right hand side are identical to the Herbinet *et al.* expression. However, to make the approach more amenable to implementation in RMG, the third term in the Herbinet *et al.* expression involving the strain energy of the sub-polycycle has been replaced here by the difference between the strain energy of the (full) polycycle associated with the transition state and the strain energy of the saturated reactant. This adjustment avoids the need for RMG to identify a sub-polycycle (which could be difficult to implement and potentially ambiguous even when analyzed by a human), while allowing potential contributions from side-rings and ligands to be taken into account. In this approach, RMG could determine the relevant polycycle by making the breaking and forming bonds in the transition state into full bonds and changing the transferred hydrogen into a saturated carbon. The reference cycle could be determined either by the ring size identified by RMG's current group matching approach or via

an approach based on SSSR (smallest set of smallest rings). The relevant species in the above expressions for Reaction 2 are shown in Table 38 below to illustrate by example.

Table 38. Species discussed in the text in the context of activation barrier estimation for the example case of Reaction 2

Reactant	 BR1
Transition state	
Product	 MA111
Ref. cycle	 cyclohexane
Sub-polycycle	 [3.2.1]cyclooctane
Polycycle	
Saturated reactant	 1,1'-bi(cyclopentane) (cyclopentylcyclopentane)

It was envisioned that strain energy could be approximated via heats of formation as shown below.

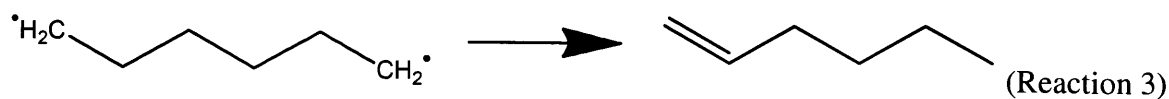
$$SE_i \approx \Delta H_{f,298,i}^\circ(\text{MM4 or PM3}) - \Delta H_{f,298,i}^\circ(\text{GA}) \quad \text{Eq. 10}$$

Care would need to be taken to ensure that the group-additivity-based estimate for heat of formation in the expression above does not include ring corrections. Under this plan, intramolecular disproportionation would become the first test case (a proof of principle) in using three-dimensional structure representations to improve on-the-fly kinetic parameter estimation during automated reaction mechanism generation, which might ultimately be extended to other reaction classes with similar steric and three-dimensional considerations, such as intramolecular hydrogen migration.

However, before implementing such an approach for intramolecular disproportionation, focus was placed on investigating this reaction class using high-level *ab initio* methods, given the limited amount of existing work in this area in the literature. In particular, as the Herbinet *et al.* JP-10 intramolecular disproportionation estimates are not directly based on experiment or *ab initio* electronic structure calculations, independent estimates based directly on *ab initio* electronic structure calculations are desirable. The present chapter presents an initial attempt in this direction, based on work performed in collaboration with Piotr Piecuch and Jesse J. Lutz of the Department of Chemistry at Michigan State University.

7.2 Methodology

Investigation of the C₁₀H₁₆ system focused on Reaction 1 proceeding through a five-membered ring transition state, as considered by Herbinet *et al.* In order to validate the approach used for the C₁₀H₁₆ system and obtain additional insights, two smaller, simpler systems were studied as well, namely, a C₆H₁₂ system, depicted in Reaction 3, and a C₃H₈ system, depicted in Reaction 4.





It should be noted that the C₃H₈ system (Reaction 4) is, unlike the others, a bimolecular reaction, and is the smallest possible alkyl disproportionation reaction.

Due to the biradical nature of the species and reactions of interest, special *ab initio* methodologies were required to accurately treat the systems considered here. Reactions 1, 3, and 4 were examined using the following three electronic structure methods : (i) the complete-active-space self-consistent-field (CASSCF) approach,¹¹⁸ which is a multi-reference approach that considers non-dynamical many-electron correlation effects, while neglecting dynamical electron correlation (ii) the second-order multi-configurational quasi-degenerate perturbation theory (MCQDPT),¹¹⁹ which is one of several different commonly-used approaches for applying second-order perturbation theory to CASSCF to account for dynamical electron correlation, and (iii) the completely renormalized (CR) coupled-cluster (CC) method abbreviated CR-CC(2,3),¹²⁰ which can provide an accurate, balanced description of both non-dynamical and dynamical correlation effects for chemical reaction pathways involving single bond breaking and biradicals while using a single-reference framework¹²⁰⁻¹²¹. Density functional theory (DFT) UB3LYP calculations¹²² were also performed for comparison.

CASSCF, MCQDPT, and CR-CC(2,3) calculations used the 6-311G(d,p) basis set,¹²³ whereas UB3LYP calculations used the CBSB7 basis set.²⁵ UB3LYP calculations were carried out using Gaussian09¹²⁴ and CASSCF, MCQDPT, and CR-CC(2,3) calculations were performed using GAMESS.¹²⁵ In particular, the CR-CC(2,3) calculations used the routines developed at Michigan State University^{120a, 126} that form part of the GAMESS distribution. MacMolPlt¹²⁷ was used for visualization. All calculations were performed on the singlet potential energy surfaces. CR-CC(2,3) calculations were performed using the frozen-core treatment wherein 1s carbon atom electrons are not considered for excitations.

As is typical for these sorts of analyses, each transition state was found as a saddle point on the potential energy surface of the indicated method, with one, and only one, imaginary frequency, and each minimum was identified as a stationary point with real frequencies only. Intrinsic reaction coordinate (IRC) calculations were performed with the same method as that used to determine transition states to confirm that each saddle point is connected to appropriate minima. The saddle points, minima, and IRCs obtained in CASSCF optimizations were used in

the subsequent single-point calculations at the higher MCQDPT and CR-CC(2,3) levels to examine the reliability of the CASSCF reaction profiles (this essentially constitutes a manual implementation of the established IRCMax method¹²⁸). The idea behind this approach is that the lower-level (CASSCF in this case) calculations should provide a reasonable reaction path, though the details of the energetics, particularly the location of the energy maximum along the reaction coordinate and the height of this barrier, will vary as higher-level methods are applied. Earlier applications and benchmark studies suggest that the MCQDPT and CR-CC(2,3) methods are capable of providing the results in the chemical (approximately 1 kcal/mol) accuracy range for the activation energies (see, for example, Refs. ^{118c, 121f, 129}) provided a large enough basis set is used. We did not use such large basis sets in this work to avoid excessive computational cost. Although the 6-311G(d,p) basis set used here has triple zeta plus polarization quality, it lacks higher angular momentum (*f* and higher) and diffuse functions, so, based on the earlier studies, such as those presented in Ref. ^{121f}, our collaborators at Michigan State University estimate that the errors resulting from these MCQDPT and CR-CC(2,3) calculations are on the order of about 2-3 kcal/mol, or so.

A four-electron, four-orbital active space (abbreviated “(4,4)”) was used in CASSCF and MCQDPT calculations, and the active orbitals were chosen to be those of chemical significance for the intramolecular disproportionation process. The active orbitals for a representative example, corresponding to one of the transition states for Reaction 3 designated in the next section as TS3A, are plotted in Figure 41. The figure shows that the molecular orbitals in the active space used in CASSCF(4,4) calculations correspond to the partially occupied molecular orbitals associated with breaking/forming bonds. The lowest energy molecular orbital is shown in the upper left and is associated with the C-H bond breakage that takes place during intramolecular disproportionation. The next lowest energy molecular orbital in the top right includes the *p* orbitals associated with the radical sites. The lower left panel depicts the corresponding anti-bonding orbital. Finally, the bottom right panel shows the highest energy orbital in the active space, which is the anti-bonding orbital corresponding to the lowest energy orbital in the active space. Such active space considerations necessitate extra care when performing the calculations; in particular, the converged orbitals for each calculation must be checked to confirm that they correspond to the desired active space; situations frequently arise

where this is not the case on the initial attempt, and the calculation procedure must be adjusted, for example, by intelligent interchange of orbitals.

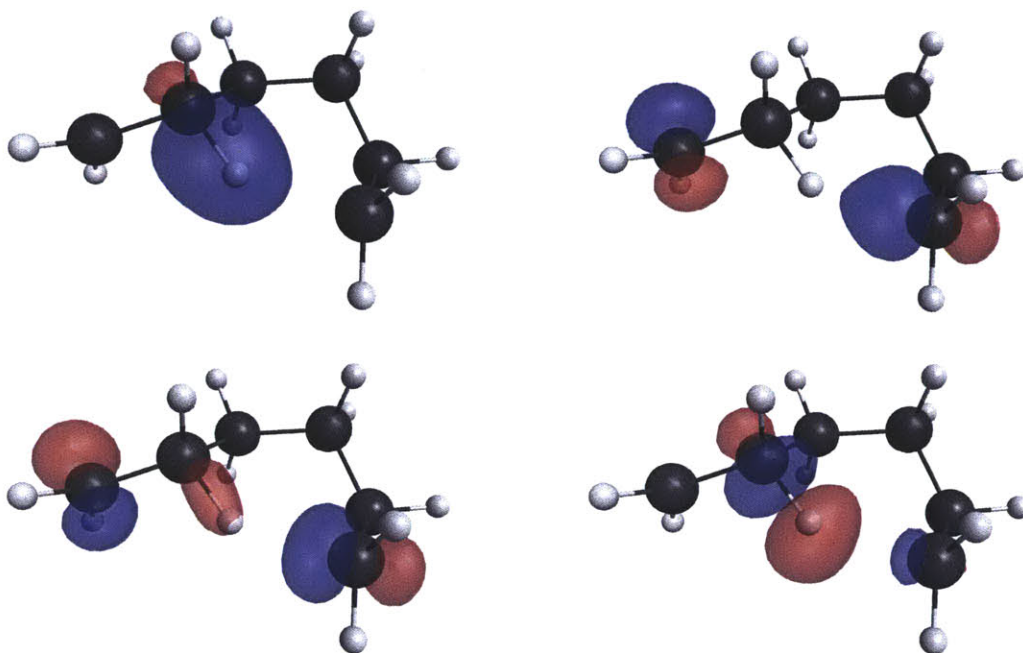


Figure 41. CASSCF(4,4) active orbitals for transition state TS3A of Reaction 3

7.3 Results and Discussion

7.3.1 Intramolecular disproportionation

7.3.1.1 C₃H₈ System

The first system considered, C₃H₈, corresponding to Reaction 4, involves methyl + ethyl disproportionation, and is the smallest possible alkyl disproportionation reaction. In addition to calculating single-point CR-CC(2,3) and MCQDPT energies along the CASSCF(4,4) IRC pathways, the C₃H₈ system is small enough to enable us to explore the CR-CC(2,3) and MCQDPT potential energy surfaces with numerical energy gradients. This system has recently been studied by Mousavipour and Homayoon¹³⁰ and by Zhu *et al.*¹³¹ The results of the CASSCF(4,4) calculations and the potential energy surface scans along the CASSCF(4,4)-optimized IRC leading to the methyl + ethyl reactants obtained with the higher level MCQDPT(4,4) and CR-CC(2,3) approaches are shown in Figure 3. The various single-point

energies calculated at the saddle-point and reactant geometries obtained with CASSCF(4,4) are shown in Table 39.

Table 39. Total electronic energies at the saddle-point (SP) and reactant (R) geometries obtained in the CASSCF(4,4)/6-311G(d,p) geometry optimizations, reported in Hartree,^a the corresponding apparent barrier heights (ABH, in kcal/mol; calculated in each case as the difference between the appropriate energies at the SP and R geometries found with CASSCF(4,4)), and the IRC barrier heights (IBH, in kcal/mol; calculated in each case as the energy difference between the highest-energy point along the CASSCF(4,4) IRC available, different for each method, and the corresponding energy at the CASSCF(4,4)-optimized R geometry), characterizing the CASSCF(4,4)/6-311G(d,p)-optimized reaction pathway representing Reaction 4.

CASSCF(4,4)			MCQDPT(4,4)//CASSCF(4,4)				CR-CC(2,3)//CASSCF(4,4)			
SP	R	ABH=IBH	SP	R	ABH	IBH	SP	R	ABH	IBH
-0.192740	-0.200490	4.9	-0.630532	-0.624321	-3.9	0	-0.682549	-0.680137	-1.5	1.0

^a Each electronic energy E is reported as $(E+118)$ Hartree.

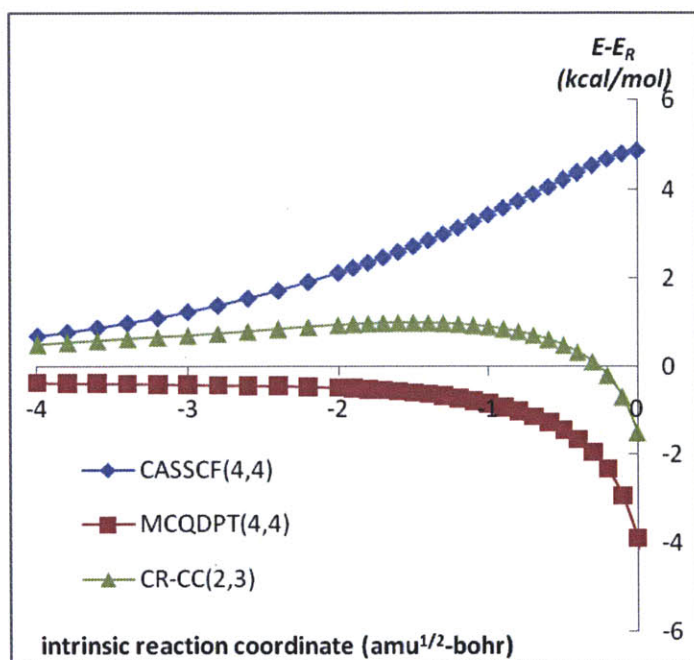


Figure 42. Energy profiles for Reaction 4 along the CASSCF(4,4) IRC, calculated at different levels of theory. In each case, the points correspond to selected structures on the relevant CASSCF(4,4)-optimized IRC, with the abscissa indicating the approximate intrinsic reaction coordinate relative to the saddle point. The energies E characterizing the profile are reported relative to the energy (E_R) of a non-interacting complex representing radical reactants identified by following the CASSCF(4,4) IRC.

As shown in Figure 42, the CASSCF(4,4) calculations show a clear barrier (of about 5 kcal/mol), while the CR-CC(2,3) single-point energy calculations along the CASSCF(4,4) IRC

show only a small barrier (under 1 kcal/mol), well within the intrinsic error of the method, shifted towards the reactant side. The analogous single-point MCQDPT calculations do not show any barrier along the CASSCF(4,4) path. Also, our own calculations with UB3LYP confirmed the findings of Zhu *et al.*, who observed the existence of a stable complex and a saddle point using that method.¹³¹ Thus, this system shows four qualitatively different potential energy surfaces as the method is varied between CASSCF(4,4), CR-CC(2,3), MCQDPT(4,4) and UB3LYP. Due to the small size of the system, we could afford to search for a saddle-point using numerical gradients of CR-CC(2,3) and MCQDPT(4,4); our efforts to find a saddle point using these two approaches were unsuccessful, suggesting that this reaction is barrierless at these higher levels of theory. This indicates a need to include both dynamical and non-dynamical electron correlation effects in order to obtain reliable reaction path energetics. The CR-CC(2,3) and MCQDPT methods describe both types of electron correlation effects in an accurate manner, eliminating the barrier on the CASSCF(4,4) potential energy surface, which includes relevant non-dynamical correlations through multi-reference considerations, but neglects dynamical correlation effects.

7.3.1.2 C₆H₁₂ System

The second system, which is associated with the cyclohexane ring opening and can be regarded as an intermediate between the previously discussed C₃H₈ case and JP-10, is the C₆H₁₂ system corresponding to Reaction 3. The C₆H₁₂ system is still sufficiently small to allow relatively rapid exploration of the MCQDPT(4,4) potential energy surface with numerical energy gradients. Three saddle points were found for Reaction 3 (hexamethylene↔1-hexene) using CASSCF(4,4). These saddle points, labeled as TS3A, TS3B, and TS3C, are shown in Figure 4.

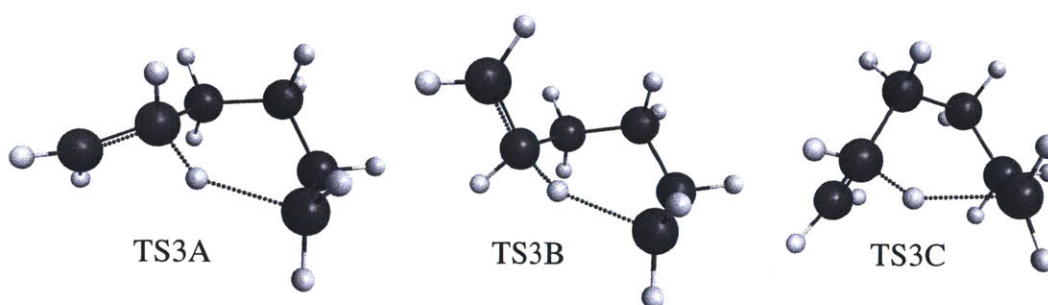


Figure 43. CASSCF(4,4) saddle points for Reaction 3

Similar saddle points were found with UB3LYP; two of these saddle points (associated with saddle points TS3A and TS3B) were considered by Sirjean *et al.*¹³² The results of the

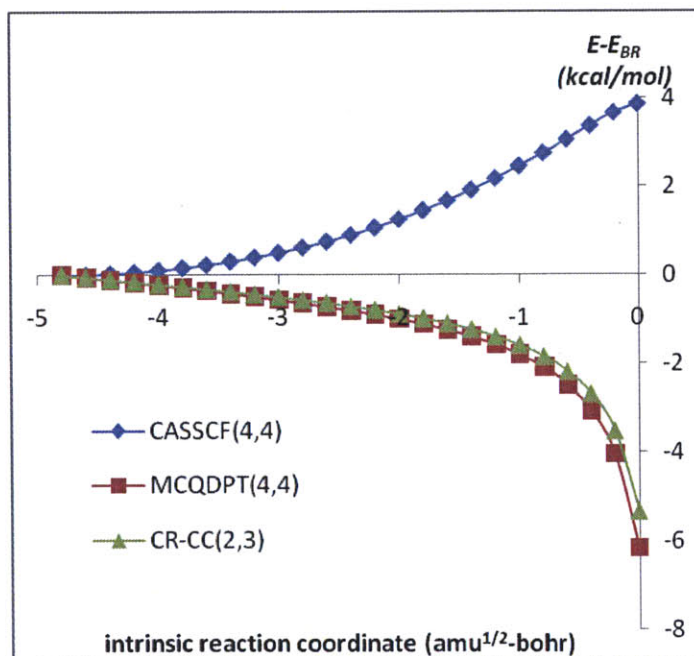
CASSCF(4,4) calculations and the potential energy surface scans along the CASSCF(4,4)-optimized IRCs leading from saddle point to the biradical reactants obtained with the higher-level MCQDPT(4,4) and CR-CC(2,3) approaches are shown in Figure 44. The various single-point energies calculated at the saddle-point and reactant geometries obtained with CASSCF(4,4) are shown in Table 40.

Table 40. Total electronic energies at the CASSCF(4,4)/6-311G(d,p)-optimized saddle points (SP) and biradical reactant minima (BR), reported in Hartree,^a the corresponding apparent barrier heights (ABH, in kcal/mol; calculated in each case as the difference between the appropriate energies at the SP and BR geometries found with CASSCF(4,4)), and the IRC barrier heights (IBH, in kcal/mol; calculated in each case as the energy difference between the highest-energy point along the CASSCF(4,4) IRC available, different for each method, and the corresponding energy at the CASSCF(4,4)-optimized BR geometry), characterizing the three CASSCF(4,4)/6-311G(d,p) optimized reaction pathways representing Reaction 3. N/C=not calculated in this work

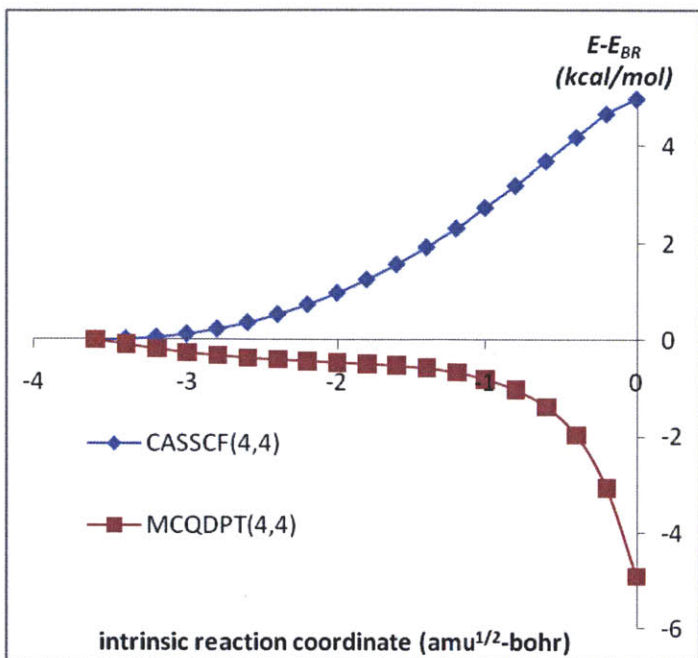
IRC	CASSCF(4,4)			MCQDPT(4,4)//CASSCF(4,4)				CR-CC(2,3)//CASSCF(4,4)			
	SP	BR	ABH=IBH	SP	BR	ABH	IBH	SP	BR	ABH	IBH
TS3A	-0.161615	-0.167792	3.9	-1.042388	-1.032588	-6.1	0	-1.133733	-1.125251	-5.3	0
TS3B	-0.158028	-0.165924	5.0	-1.038881	-1.031050	-4.9	0	-1.131023	-1.123573	-4.7	N/C
TS3C	-0.155247	-0.167922	8.0	-1.040458	-1.033069	-4.6	0.3	-1.131659	-1.125778	-3.7	0.4

^a Each electronic energy E is reported as $(E+234)$ Hartree.

(a) TS3A



(b) TS3B



(c) TS3C

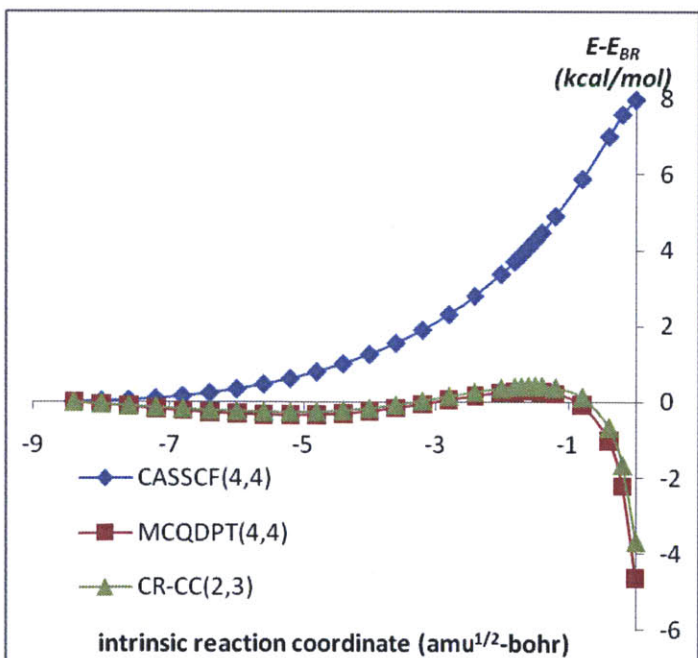


Figure 44. Energy profiles for Reaction 3 along the CASSCF(4,4) IRCs passing through the (a) TS3A, (b) TS3B, and (c) TS3C saddle points, calculated at different levels of theory. In each case, the points correspond to selected structures on the relevant CASSCF(4,4)-optimized IRC, with the abscissa indicating the approximate intrinsic reaction coordinate relative to the saddle point. The energies E characterizing each profile are reported relative to the corresponding IRC-connected biradical minimum (E_{BR}), which corresponds to the terminal point at the left.

The single-point energy calculations at the selected structures along the CASSCF(4,4) IRCs corresponding to the transition states TS3A, TS3B, and TS3C found with CASSCF(4,4) and UB3LYP suggest no barriers or negligible barriers at the MCQDPT(4,4) and CR-CC(2,3) theory levels. In fact, efforts to refine the CASSCF(4,4) saddle-point geometries to corresponding MCQDPT minima (with numerical gradients) were unsuccessful, suggesting that these are not valid saddle points at the MCQDPT level of theory. Although we were unable to perform similar saddle-point searches with the CR-CC(2,3) method, the similarity of the MCQDPT and CR-CC(2,3) results strongly suggests that the transition states found with CASSCF(4,4) aren't valid saddle points at the CR-CC(2,3) level either. This again demonstrates that lower-order methods, such as CASSCF, which neglects dynamical correlations, and UB3LYP, which has known difficulties in describing multi-reference phenomena and biradical energetics, cannot be used to obtain reliable reaction path energetics. Appropriate dynamical and non-dynamical correlation effects, as considered by MCQDPT and CR-CC(2,3) approaches, must be taken into account in order to obtain reasonable results. The absence of clear barriers in the reaction profiles obtained by performing single-point MCQDPT(4,4) and CR-CC(2,3) calculations along the CASSCF(4,4) IRC pathways suggests that the CBS-QB3 barrier (based on B3LYP geometries) of 3.85 kcal/mol used by Herbinet *et al.* (apparently based on the aforementioned study from Sirjean *et al.*) may be too high.

Our efforts to optimize the corresponding biradical reactant minima (found at either CASSCF(4,4) or UB3LYP levels) to a stable MCQDPT(4,4) minimum energy configuration were unsuccessful as well. However, explorations of the C₆H₁₂ system by Kiefer *et al.*¹³³ with CASPT2(2,2) produced a viable saddle point (more akin to the biradical reactants than the previously-discussed CASSCF(4,4) and UB3LYP TS3 geometries); this geometry was readily re-optimized in this work to a saddle point at the MCQDPT(4,4) level of theory in this study. The MCQDPT-optimized saddle point geometry is shown in Figure 45.

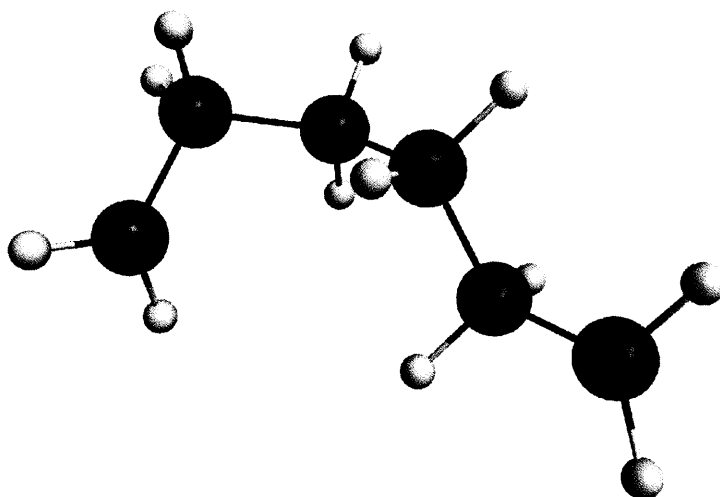


Figure 45. MCQDPT(4,4) saddle point for Reaction 3

The molecular geometry exhibits significant biradical character, with a substantial separation (about 3.26 Å) between the atoms that will become bonded in product; without reaction path following, this geometry might be mistaken as a saddle point only for rotation about a single bond. (The ZPE-uncorrected barrier was found to be 2.6 kcal/mol at the MCQDPT(4,4) level.) These results underscore the lack of reliability of CASSCF(4,4) and UB3LYP geometries for this type of reaction system.

7.3.1.3 C₁₀H₁₆ System

As in the case of the C₆H₁₂ system discussed in the previous subsection, three saddle points, labeled TS1A, TS1B, and TS1C, were found for Reaction 1 using CASSCF(4,4), as shown in Figure 46.

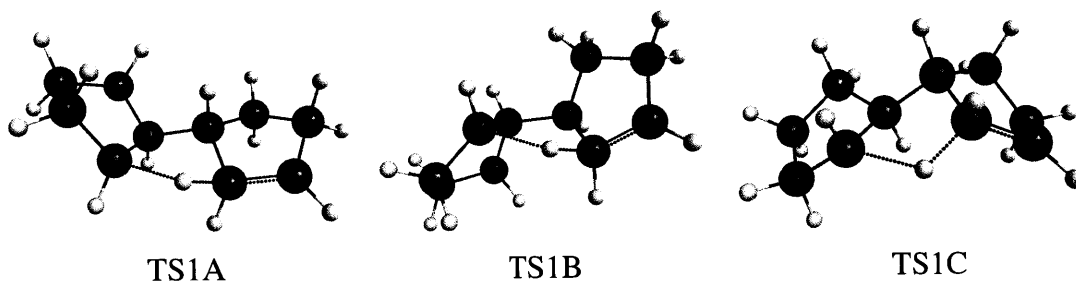


Figure 46. CASSCF(4,4) saddle points for Reaction 1

IRC calculations were performed to confirm that each saddle point connects the minima that correspond to the desired reactant and product species. Similar saddle-point geometries, with similar IRCs, were found using UB3LYP. Higher level MCQDPT and CR-CC(2,3) single-point calculations that describe the dynamical as well as non-dynamical correlation effects were performed at selected points along the CASSCF(4,4) IRC, and results are shown in Figure 47 and Table 41.

Table 41. Total electronic energies at the CASSCF(4,4)/6-311G(d,p)-optimized saddle points (SP) and biradical reactant minima (BR), reported in Hartree,^a the corresponding apparent barrier heights (ABH, in kcal/mol; calculated in each case as the difference between the appropriate energies at the SP and R geometries found with CASSCF(4,4)), and the IRC barrier heights (IBH, in kcal/mol; calculated in each case as the energy difference between the highest-energy point along the CASSCF(4,4) IRC available, different for each method, and the corresponding energy at the CASSCF(4,4)-optimized BR geometry), characterizing the three CASSCF(4,4)/6-311G(d,p) optimized reaction pathways representing Reaction 1.

IRC	CASSCF(4,4)			MCQDPT(4,4)//CASSCF(4,4)				CR-CC(2,3)//CASSCF(4,4)			
	SP	BR	ABH=IBH	SP	BR	ABH	IBH	SP	BR	ABH	IBH
TS1A	-0.963522	-1.015185	32.4	-2.434862	-2.453652	11.8	14.6	-2.566955	-2.592631	16.1	17.0
TS1B	-1.000906	-1.014091	8.3	-2.468157	-2.459969	-5.1	0.5	-2.601343	-2.594176	-4.5	0.4
TS1C	-0.980512	-1.012022	19.8	-2.448613	-2.452571	2.5	9.6	-2.581793	-2.591074	5.8	9.1

^a Each electronic energy E is reported as $(E+387)$ Hartree.

In analogy to the previously discussed C_3H_8 and C_6H_{12} systems, the potential energy surface scans along the CASSCF(4,4)-optimized IRCs obtained with the higher-level MCQDPT(4,4) and CR-CC(2,3) approaches suggest that the CASSCF calculations do not provide a reliable description of pathway energetics characterizing Reaction 1. This is particularly evident in the case of the reaction pathway corresponding to transition state TS1B, where CASSCF(4,4) predicts the existence of the clear activation barrier on the order of 8 kcal/mol; however, this barrier nearly disappears in the MCQDPT(4,4) and CR-CC(2,3) calculations, mimicking the behavior we have seen with all three transition states for the C_6H_{12} case. The results suggest that the neglect of dynamical correlations by the CASSCF(4,4) calculations produces an unphysical barrier in the potential energy surface in this case, whereas the inclusion of dynamical correlation by MCQDPT(4,4) and CR-CC(2,3) methods (in addition to non-dynamical correlation) does not produce such a barrier.

Work by collaborators at Michigan State University has found that the inclusion of more active orbitals and more active electrons in the CASSCF calculations lowers the apparent TS1B barrier, which, according to their analysis, reinforces the above observations. They found, for example, that the apparent TS1B barriers with CASSCF(6,6) and CASSCF(8,8) calculations are 6.6 and 5.3 kcal/mol, as opposed to about 8 kcal/mol obtained in the CASSCF(4,4) calculations.

For the remaining two pathways, which correspond to transition states TS1A and TS1C, the barriers obtained in the CASSCF(4,4) calculations survive the inclusion of dynamical correlation effects with MCQDPT(4,4) and CR-CC(2,3) methods, but the barriers are significantly reduced.

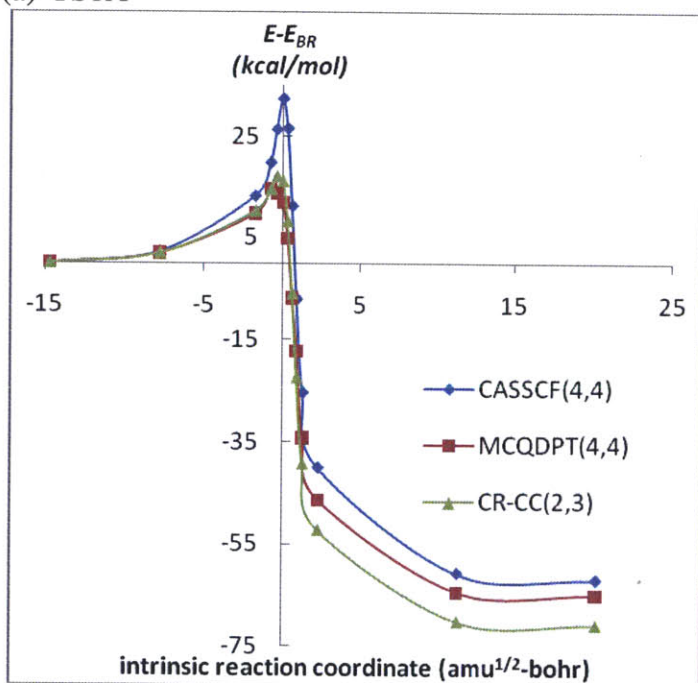
We were unable to perform saddle-point optimizations using MCQDPT and CR-CC(2,3) approaches due to prohibitive costs of such calculations without analytic gradients. However, the shift in the maximum in the each of the energy profiles in Figures 7 (a) and 7 (c) suggests that the inclusion of dynamical correlation effects via MCQDPT or CR-CC(2,3) may tend to shift the saddle-points TS1A and TS1C toward the corresponding biradical reactant minima.

Although there are some differences between barrier heights resulting from the MCQDPT and CR-CC(2,3) calculations for the pathways involving the TS1A and TS1C transition states, collaborators at Michigan State University found that these differences become smaller as the active space employed in the MCQDPT calculations is expanded. The apparent barrier height corresponding to the transition state TS1A obtained via MCQDPT(8,8) calculations, for example, was found to be 13.1 kcal/mol, closer to the CR-CC(2,3) result of 16.1 kcal/mol than the MCQDPT(4,4) result of 11.8 kcal/mol. Likewise, a 6.1 kcal/mol apparent barrier height for TS1C was found using MCQDPT(8,8), significantly closer to the CR-CC(2,3) result of 5.8 kcal/mol than the MCQDPT(4,4) result of 2.5 kcal/mol. Noticing the similarity of results obtained with the MCQDPT and CR-CC(2,3) approaches for this reaction, the Michigan State University collaborators interpreted this as providing further confirmation that the CASSCF approach is unreliable for the energetics of Reaction 1.

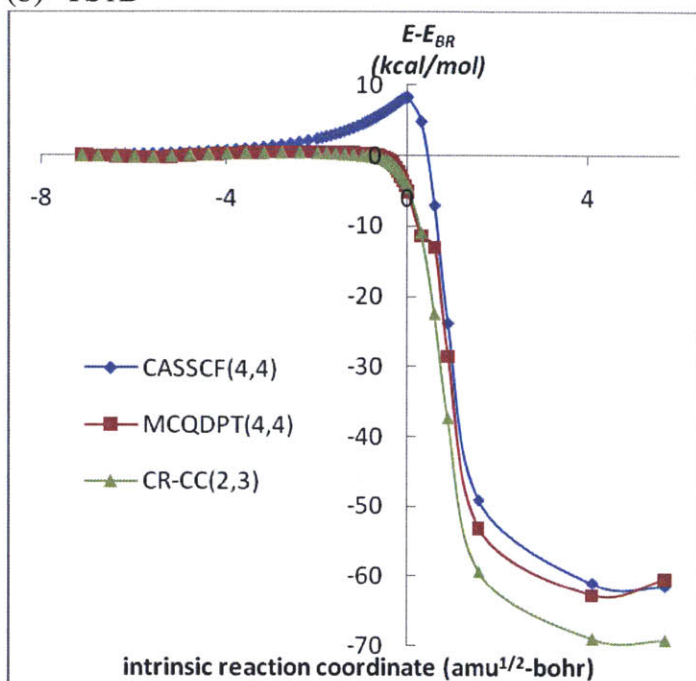
It should be noted that the above results do not consider zero-point energy (ZPE) effects, but these considerations should not have a material impact on the conclusions presented here. CASSCF(4,4) calculations for each of the saddle-points suggest that ZPE effects should lower the barriers by approximately 1.0 kcal/mol. It should also be noted that the biradical that each of the three saddle points connects to is a different CASSCF(4,4) minimum; the biradical minimum associated with TS1A has lowest energy at CASSCF(4,4) level, whereas the biradical minimum

associated with TS1B has the lowest energy at MCQDPT(4,4) and CR-CC(2,3) levels. Calculations using a CASSCF(4,4) geometry for JP-10 also enabled estimation of 0 K bond energy associated with carbon-carbon bond breaking to form the BR1 biradical associated with TS1A, producing values of 59, 75, and 81 kcal/mol for CASSCF(4,4), MCQDPT(4,4)//CASSCF(4,4), and CR-CC(2,3)//CASSCF(4,4), respectively (all with CASSCF(4,4) ZPE); these values are lower than the 0 K bond dissociation energy in ethane of about 88 kcal/mol.¹³⁴

(a) TS1A



(b) TS1B



(c) TS1C

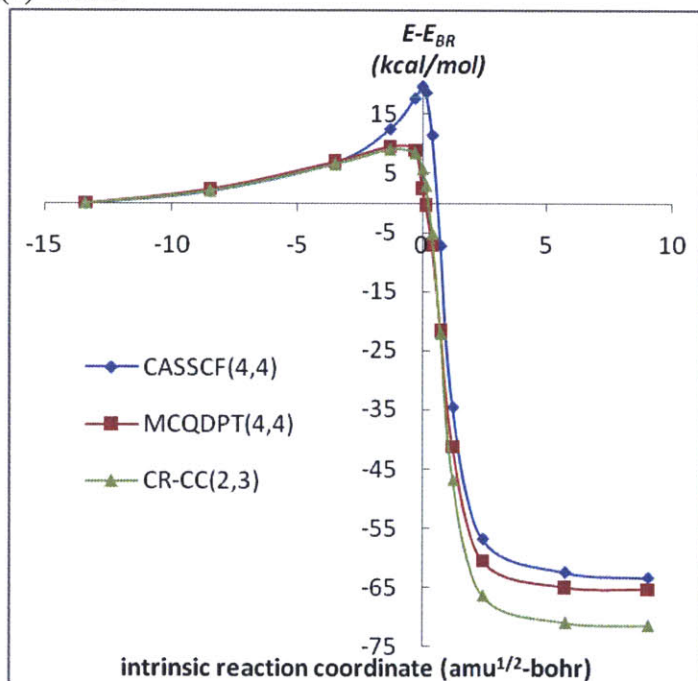


Figure 47. Energy profiles for Reaction 1 along the CASSCF(4,4) IRCs passing through the (a) TS1A, (b) TS1B, and (c) TS1C saddle points, calculated at different levels of theory. In each case, the points correspond to selected structures on the relevant CASSCF(4,4)-optimized IRC, with the abscissa indicating the approximate intrinsic reaction coordinate relative to the saddle point. The energies E characterizing each profile are reported relative to the corresponding IRC-connected biradical minimum (E_{BR}), which corresponds to the terminal point at the left.

7.3.1.4 Discussion / Kinetic modeling of the C₁₀H₁₆ system

Overall, the above three systems illustrate interesting method-dependent phenomena in the potential energy surfaces of disproportionation reactions, with a strong indication toward the need to use higher-level quantum chemistry approaches, such as MCQDPT or CR-CC(2,3), that account for appropriate dynamical and non-dynamical correlation effects. In the case of the disproportionation reactions examined here, results clearly show that the inclusion of dynamical electron correlation on top of non-dynamical correlation reduces or eliminates the reaction barriers along paths determined using lower-order (CASSCF) calculations.

To ascertain the influence of lower intramolecular disproportionation barriers on combustion behavior of JP-10, the activation energies of all intramolecular disproportionation reactions in the mechanism were dropped by 7.75 kcal/mol (the activation energy for Reaction 1 from the Herbinet approach) to a minimum of 0.0 kcal/mol. The resulting mechanism was labeled v0.19b, and ignition simulations were performed using the isobaric simulation feature in SENKIN with ignition determined by the point with highest CH mole fraction.^{117a} Ignition delay results for the unperturbed v0.19 and the perturbed v0.19b mechanisms were compared to experimental values of Davidson *et al.*,⁷⁸ as shown in Figure 48.

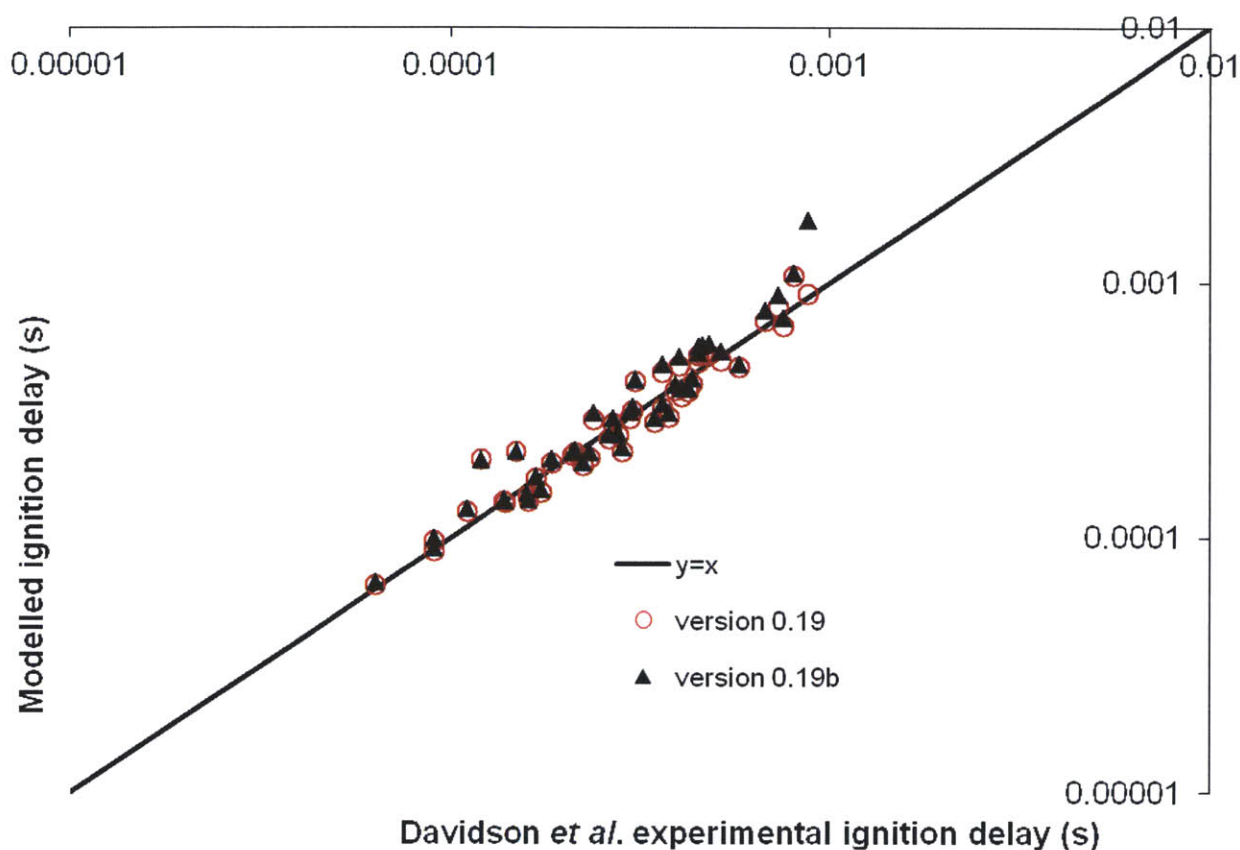


Figure 48. Parity plot comparing ignition delay from unperturbed and perturbed mechanisms (v0.19 and v0.19b, respectively) to experimental results of Davidson et al.⁷⁸

The results suggest that the activation energy perturbation has its biggest effect on the conditions with the longest ignition delay. In particular, for such conditions, the activation energy reduction appears to further delay ignition. One possible explanation for this is that the faster termination reaction of the biradical is depleting the (bi)radical pool, hence delaying the onset of ignition. The importance of these reactions at conditions with longer ignition delay is supported by sensitivity analysis, which shows, for example, that Reaction 1 increases from an importance rank of 171 out of 7740 reactions (again, computed based on the sensitivity metric defined by Equation 5, discussed previously) at conditions corresponding to the point at the far left (1671 K, 1.05 atm, 0.20 mole % JP-10, 2.99 mole %, O₂; remainder Ar) to a rank of 130 at the conditions corresponding to the point at the far right (1352 K, 8.69 atm, 0.20 mole % JP-10, 2.78 mole %, O₂; remainder Ar).

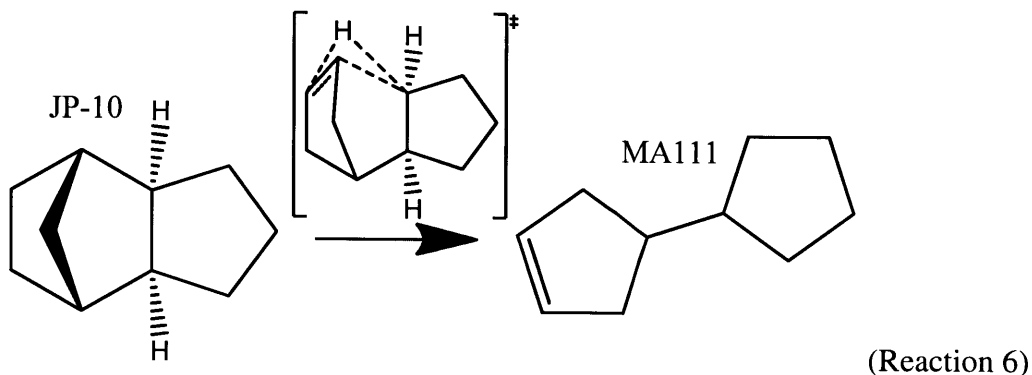
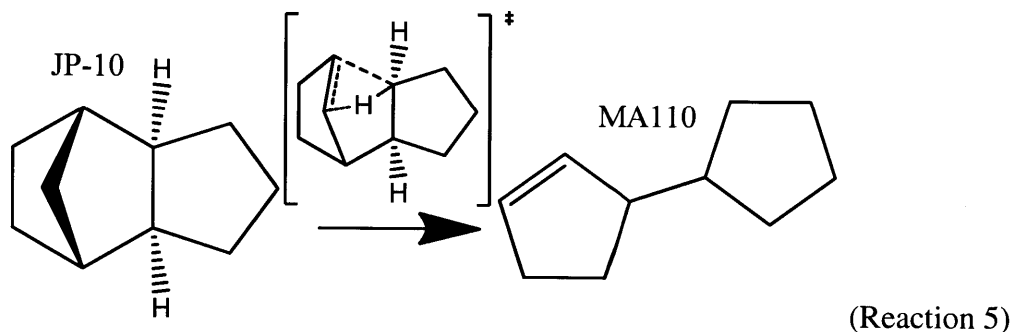
7.3.2 Ring-opening pathways with concerted hydrogen transfer

7.3.2.1 Background

The possibility of concerted reactions leading directly from ring to alkene, bypassing biradical intermediate, was also considered. There appears to be limited discussion of these pathways in the literature. An investigation of C_4H_8 by Ventura *et al.*¹³⁵ found that tetramethylene was not a stable molecule (*i.e.* there was no minimum on the potential energy surface) when using the MR-AQCC approach,¹³⁶ which is a method related to multireference configuration interaction (MR-CI), taking both dynamical and non-dynamical correlation effects into account. This would suggest that ring-opening of cyclobutane leads directly to 1-butene without a biradical intermediate. Later, Kiefer *et al.* found a concerted pathway from cyclohexane to 1-hexene using CASPT2(2,2) with relatively low barrier, but concluded that the kinetics were not competitive with the biradical pathway due to a low Arrhenius pre-factor.¹³³

7.3.2.2 DFT calculations

Two saddle points, labeled as TS5 and TS6, corresponding to a concerted JP-10 \leftrightarrow MA110 reaction, have been found at the UB3LYP level. The corresponding reactions are shown below.



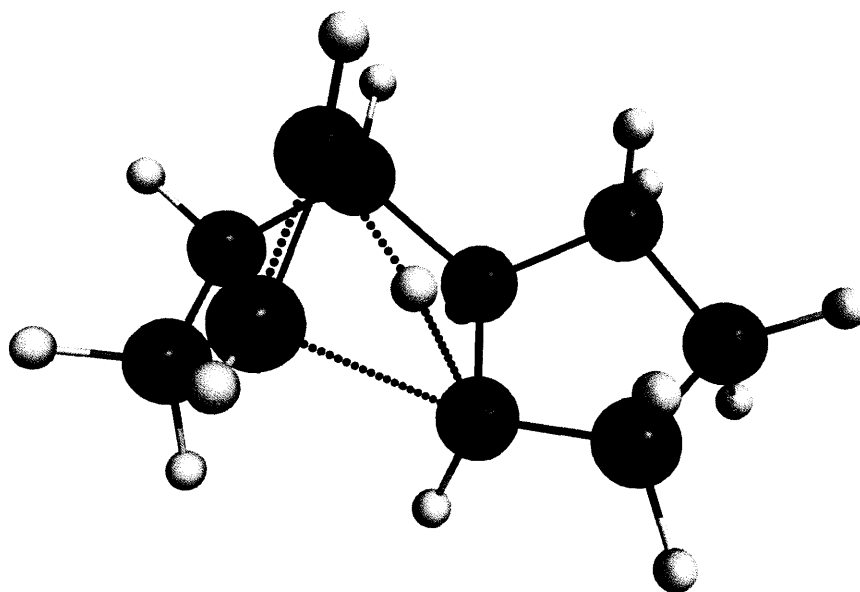
Our CBS-QB3 calculations,²⁵ performed with Gaussian09, suggest a high barrier of over 110 kcal/mol for TS5 and a lower barrier of about 80 kcal/mol for TS6 (see Table 4). The latter is comparable to the barrier of 77 kcal/mol estimated by Herbinet *et al.* for the ring-opening of JP-10 to form BR1, which provides a rough indication that the concerted pathway could be competitive with the biradical pathway for JP-10 decomposition. The existence of the TS5 and TS6 transition states has been reexamined using the CASSCF(4,4) approach. We were unable to locate a saddle point similar to TS5 on the CASSCF(4,4) potential energy surface, but calculations affirmed the existence of the saddle-point structure for TS6 on the CASSCF(4,4) potential energy surface. The lack of success in locating a CASSCF(4,4) saddle point similar to TS5 could suggest that it is spurious, but it is also noted that this method has been shown to be unreliable for different (though somewhat similar) systems elsewhere in this work.

Table 42. CBS-QB3 absolute and relative energetics for Reactions 5 and 6, with T1 diagnostic¹³⁷ results for the CCSD(T) step of the CBS-QB3 calculation for each saddle-point

Species	CBS-QB3 (0 K) energy (Hartree)	Barrier Height (kcal/mol)	T1 diagnostic
JP-10	-389.881322	-	-
TS5	-389.702457	112.2	0.0127
TS6	-389.754219	79.8	0.0173

Table 42 reports the CBS-QB3 results (electronic energy + zero point energy) for the JP-10 reactant and TS5 and TS6 saddle points on the UB3LYP surface. The T1 diagnostic values, both being less than 0.02, suggest that multi-reference considerations are not critical for this pathway. (It is noted, however, that the utility of the T1 diagnostic has been called into question.) The UB3LYP geometries of TS5 and TS6 are shown in Figure 49.

(a) TS5



(b) TS6

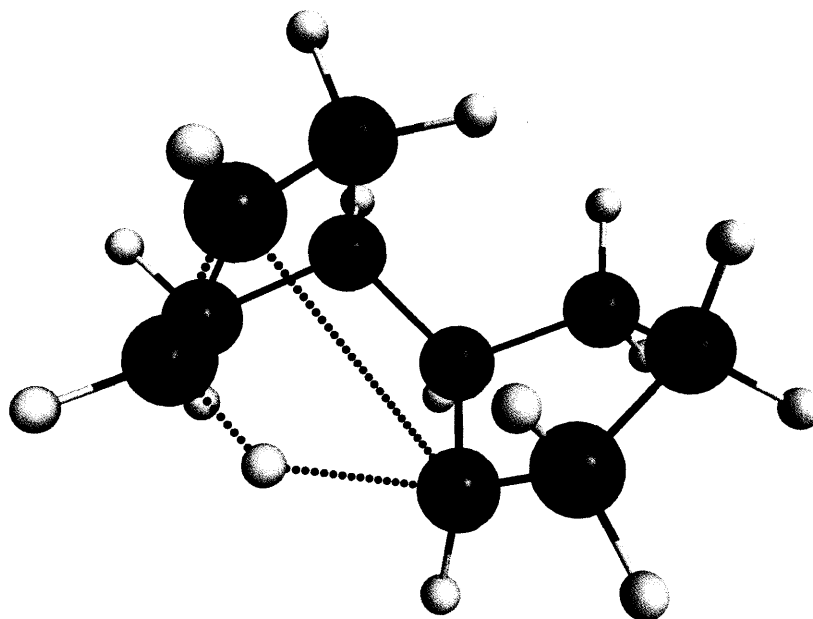


Figure 49. UB3LYP saddle-point geometries for Reactions 5 and 6: (a) TS5 and (b) TS6

Using CanTherm,⁷² a modified Arrhenius fit was obtained for each of these reactions based on CBS-QB3 results and transition state theory. This led to the following rate parameter estimates for Reactions 5 and 6:

$$k_5(T) = 2.54 \times 10^{11} \text{ s}^{-1} \left(\frac{T}{1000 \text{ K}} \right)^{6.88} e^{\frac{-99.31 \text{ kcal/mol}}{RT}} \quad \text{Eq. 11}$$

$$k_6(T) = 9.1 \times 10^{14} \text{ s}^{-1} \left(\frac{T}{1000 \text{ K}} \right)^{1.23} e^{\frac{-80.34 \text{ kcal/mol}}{RT}} \quad \text{Eq. 12}$$

In both cases, the chiral nature of the saddle-point (*i.e.* the fact that it is not superimposable on its mirror image) indicates that there are two distinct, but isoenergetic pathways for each of these reactions; consequently, the Arrhenius prefactors from CanTherm have been multiplied by a factor of two to account for this reaction path degeneracy in the expressions above.

It is possible that additional concerted JP-10 decomposition pathways exist that were not identified in this study.

7.3.2.3 Kinetics implications

An attempt was made to assess the importance of the Reaction 6 concerted pathway relative to the biradical pathways considered by Herbinet *et al.* For the biradical pathways, kinetic parameters for the JP-10 \leftrightarrow BR1 reaction, Reaction 1 (both pathways), and Reaction 2 were taken from Herbinet *et al.* Thermochemistry for JP-10 is based on the Herbinet *et al.* values and thermochemistry for BR1 is based on the values from Herbinet *et al.* estimated using quantum methods. Two approaches were used to calculate an effective rate constant for the two-step biradical pathway. One approach was to assume rapid equilibrium between BR1 and JP-10. The second approach was to apply the pseudo-steady-state approximation to BR1. A comparison of these effective rates with the Reaction 6 concerted pathway rate (computed as described previously) is shown in Figure 50.

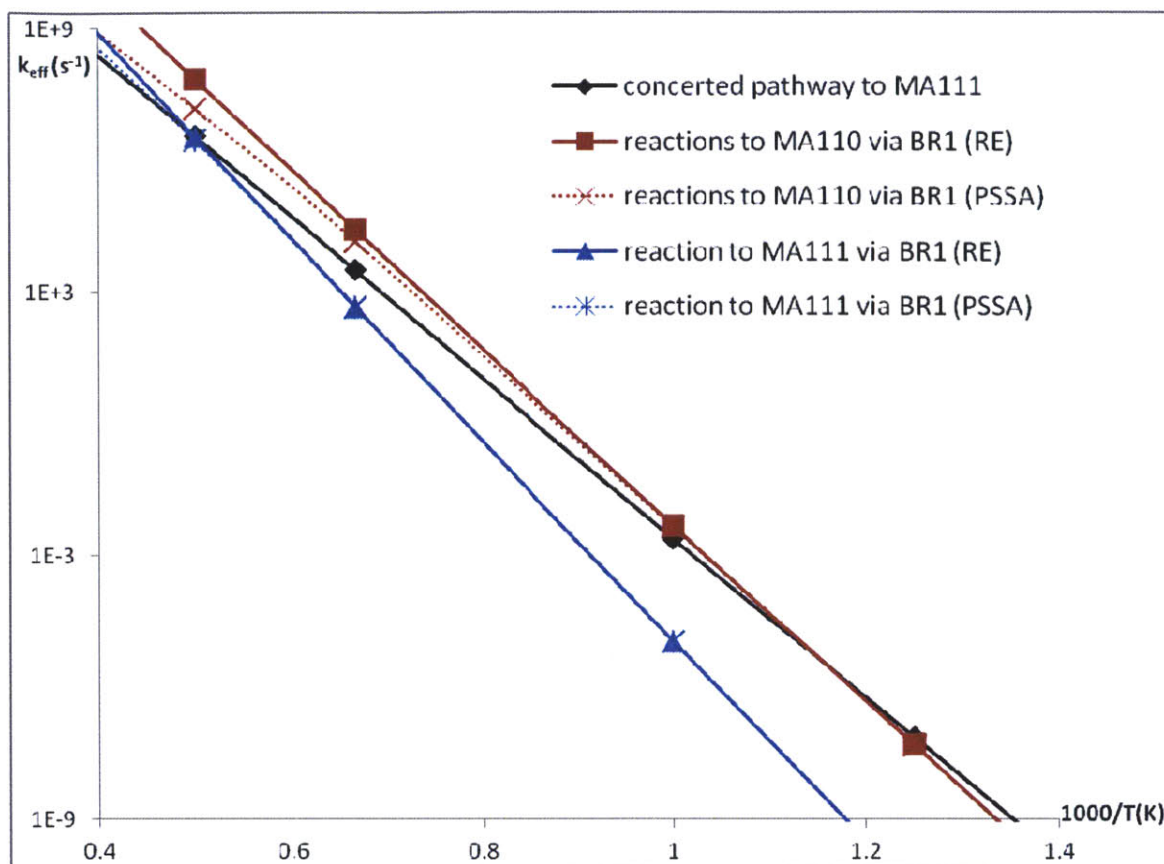


Figure 50. Effective rates for initial decomposition of JP-10 through various pathways (RE=rapid equilibrium assumption; PSSA=pseudo-steady-state approximation)

The comparison suggests that the concerted pathway is competitive at very low temperatures, but the biradical pathway to MA110 dominates at higher temperatures of interest to combustion. Even so, the concerted pathway to MA111 appears to be faster than the corresponding biradical pathway for temperatures less than 2000 K. The lower magnitude slope for the concerted reaction in the Arrhenius plots shown in Figure 10 suggests that the activation energy for the concerted pathway is lower than the effective activation energy for the intramolecular disproportionation process (taking recombination and disproportionation of the biradical into consideration).

It should be noted that this comparison is particularly sensitive to both the Herbinet *et al.* kinetic parameters for the BR1 intramolecular disproportionation reactions (which, as previously discussed, may be inaccurate) and the thermochemistry of JP-10 and BR1. For example, if the group additivity estimates for BR1 thermochemistry from Herbinet *et al.* are used in place of the BR1 quantum chemistry values from Herbinet *et al.*, the concerted pathway appears to be noticeably less competitive, apparently due largely to the significant entropy discrepancy

between the two estimates. This suggests that accurately capturing biradical thermochemistry will be an important consideration in future efforts to refine JP-10 decomposition models.

It should also be noted that it has recently been proposed that ring opening via carbenes may be competitive with biradical routes,¹³⁸ but the issue is far from settled. These carbene pathways have not been considered here.

7.4 Conclusions

Multireference (CASSCF and MCQDPT) and completely renormalized coupled-cluster (CR-CC(2,3)) calculations have been applied to study the initial stages of JP-10 decomposition. Results of these studies demonstrate interesting method-dependent phenomena for the potential energy surfaces of disproportionation reactions and provide evidence that the barriers to these reactions may be much lower than previously thought in the case of intramolecular disproportionation in a key JP-10 decomposition pathway. The new calculations provide evidence that some disproportionation steps may be barrierless. Results suggest that the CASSCF(4,4) approach's neglect of dynamical electron correlation produces unphysical saddle-point structures and unreliable reaction path energetics for intramolecular disproportionation processes, while these artifacts may be eliminated through the inclusion of appropriate dynamical and non-dynamical many-electron correlation effects via use of MCQDPT or CR-CC(2,3) methods. Modeling results suggest that the effect of reducing activation energies for the intramolecular disproportionation pathways would be to extend the time to ignition for conditions with the longest ignition delay. Our quantum chemical calculations have also identified previously unexplored pathways for JP-10 decomposition involving concerted ring-opening pathways that bypass biradical formation. Preliminary analysis suggests that these additional pathways are likely not quite competitive with the biradical pathways that are characterized (according to high-level MCQDPT and CR-CC(2,3) calculations) by low-energy barriers or lack of activation barriers, though they are not necessarily insignificant. Overall, our findings demonstrate that the chemistry of ring-opening and associated pathways is not yet completely understood and further study is needed. In particular, in order to obtain reliable estimates, effort should focus on incorporating the results of higher level *ab initio* quantum chemistry calculations that accurately accounting for dynamical and non-dynamical correlation effects.

It appears that this study calls into question not only the Herbinet *et al.* kinetic parameters for the JP-10 system considered here, but also the kinetic modeling of smaller cycloalkane systems that have been studied with DFT methods previously. As such, it is judged that further study of the intramolecular disproportionation reaction class is warranted before effort is placed on implementing intramolecular disproportionation kinetic parameter estimation routines in RMG that account for three-dimensional geometry effects. (However, other reaction classes, such as intramolecular hydrogen migration, may be more suitable to applying such three-dimensional geometry-based kinetic parameter estimation approaches, without the need for substantial additional high-level theoretical investigation; such opportunities are discussed further in the final chapter.)

The clearest route to further explore these reactions in a computationally efficient manner appears to be through use of CASPT2(4,4) calculations with MOLPRO. The CASPT2 method is a multi-reference approach incorporating second-order perturbation theory, and as such, it incorporates both dynamical and non-dynamical electron correlation effects. It is similar to the MCQDPT method in GAMESS; however, the MOLPRO implementation of CASPT2 includes analytical gradients, which greatly speed optimization to saddle-points and minima, as well as the computation of Hessians in order to compute frequencies, partition functions, zero-point energies, etc.

Ultimately, this work sows the seeds for future improvements to RMG rate parameter estimation. In particular, evidence has been presented to suggest that application of methods incorporating both dynamical and non-dynamical correlation can potentially improve upon existing intramolecular disproportionation rate parameter estimates. Following such refinement, the next step would be to improve upon the rate parameter estimation scheme used by RMG. This might be done using an approach like that originally envisioned, similar to the approach of Herbinet *et al.* or, depending on the findings of the application of more accurate *ab initio* methods, some variant of this approach may be warranted. Finally, the concerted reactions identified here involving ring opening with simultaneous hydrogen transfer may ultimately be found to be suitable for inclusion as a new and separate reaction class within RMG's reaction templates, particularly if further investigation demonstrates that these reactions are more widespread than the cases presented here and if they are found to be important relative to competing pathways.

8 Chapter 8: Opportunities for further investigation and conclusions

8.1 Introduction

In addition to providing concrete contributions in areas such as JP-10 combustion modeling and improvements to RMG functionality and accuracy, the work undertaken in this thesis has generated a number of ideas for further avenues of exploration. Several of these opportunities have been presented, as appropriate, in the preceding chapters. Here, the discussion is expanded to introduce an opportunity for improving the reaction mechanism generation algorithm, and further discuss opportunities to better account for conformational flexibility and to improve kinetic parameter estimates using explicit three-dimensional geometry-based approaches. The chapter concludes with a summary of the contributions of this thesis and some general conclusions.

8.2 Improvement to reaction mechanism generation algorithm

The main focus of this thesis is improvement to parameter estimation and expanding the scope of applications for mechanism generation. Nevertheless, the underlying mechanism generation algorithm is an important aspect to consider, and an idea for improving upon the existing approach is proposed here.

The algorithm currently implemented in RMG determines model validity using a criterion based on a user-specified tolerance and a characteristic flux, R_{char} :

$$\frac{dC_j}{dt} < tol \cdot R_{char} \quad \forall j \in \text{edge species}$$
$$\text{with } R_{char} = \left[\sum_i \left(\frac{dC_i}{dt} \right)^2 \right]^{\frac{1}{2}} \quad i \in \text{core species}$$

Eq. 13

This criterion works well under many circumstances, and has been found to produce reasonable mechanisms. However, a potential issue arises when trying to generate a model that is valid over long time scales or for high conversions. In particular, the system will tend to approach equilibrium, such that the net rate of change of the core species concentrations tends to zero, correspondingly causing the characteristic flux, R_{char} , as defined above, to tend to zero. The flux

to edge species, however, only considers unidirectional flux as there is no “reverse” reaction to consider for reactions to edge species. Thus, the model validity criterion becomes difficult to satisfy as equilibrium is approached, and the algorithm becomes more likely to enlarge the model. This appears to be an artifact related to the definition of the characteristic flux, rather than a phenomenon that should be captured by the model building algorithm. The potential consequence is that mechanisms are needlessly expanded, diverting computational resources from important chemistry to explore relatively unimportant processes. To work around this issue, a change in definition of the characteristic flux is proposed, wherein both forward and reverse contributions are considered separately such that the opposing sign of these terms does not cancel, as it would if the net flux were used:

$$R_{char,proposed} = \left[\sum_i \left(\left. \frac{dC_i}{dt} \right|_{forward} \right)^2 + \left(\left. \frac{dC_i}{dt} \right|_{reverse} \right)^2 \right]^{\frac{1}{2}} \quad i \in \text{core species} \quad \text{Eq. 14}$$

In contrast to the original definition, this characteristic flux does not approach zero as equilibrium is approached, and appears to offer a potentially more appropriate metric for mechanism generation. The net effect of this definition change of R_{char} should be a shift to a focus on earlier time scales when species concentrations are changing rapidly and away from later time scales near steady-state. It is anticipated that this could allow more efficient generation of chemical kinetic models, by allowing the use of lower tolerances (given fixed computational resources) and/or producing smaller mechanisms that still capture all of the important chemistry.

8.3 Better approaches for accounting for conformational flexibility

8.3.1 Models to account for conformational flexibility

As discussed previously, the typical treatment by the QMTP system is to assume harmonic behavior for all internal modes; we have implemented experimental rotor scan methods with the MM4 force field that account for conformational flexibility at a higher level (separable hindered rotor treatment), but even this approach may not achieve a sufficient level of accuracy for certain molecules/applications, and some examples of this appear in the literature.⁷⁹ There is opportunity for improvement of this treatment, particularly when using the relatively computationally inexpensive MM4 approach.

One possible approach is to use “conformational averaging” or a “mixture of conformers model”.¹³⁹ This is the approach commonly taken with Allinger’s MM*n* force fields.²³⁻²⁴ The first step in such an approach would be to enumerate most (ideally all) low-lying local minima on the molecule’s potential energy surface. Then, corrections are applied, using the minimum energy conformation as a reference. For example, corrections to heat of formation include a “POP” correction which incorporates a weighted average of the conformer energies; the weighing is done with respect to population, with populations determined by Gibbs free energy computed with the harmonic oscillator approximation. Entropy is adjusted by considering the entropy of mixing associated with a mixture of conformers with populations weighted by the Gibbs free energy, again computed with the harmonic oscillator approximation. Although Allinger’s literature is not clear on the recommended heat capacity treatment with this approach, one could extrapolate from the heat of formation and entropy treatments to suggest that it also would involve population-weighted average of harmonic-oscillator-based conformer heat capacities; alternatively, it could be computed based on $S(T)$. One advantage of this approach is the fact that it implicitly accounts for non-separable multi-dimensional torsional considerations; additionally, it can account for certain non-rotor anharmonic motions, such as ring inversion. However, the theoretical foundations for this approach are not very solid, and, for example, in a simple averaging approach, the heat capacity wouldn’t obey proper high-temperature limits for molecules with rotors, as this approach would effectively include $+R$ for each rotor mode, rather than the appropriate $+R/2$ contribution.

There are also much more sophisticated approaches that have a strong physical foundation based on path-integral calculations.¹⁴⁰ Unfortunately, these approaches can be extremely computationally demanding, even for relatively small molecules. Alternative approaches based on various approximations and simplifications are discussed elsewhere.¹⁴¹

The most promising approach that appears to offer a favorable balance between accuracy and speed for our application is the recently introduced multi-structural method of Truhlar and coworkers⁸⁰; this method is reportedly able to account for conformational flexibility via both rotation about bonds as well as other modes such as ring inversion, and obeys desired statistical mechanical limits, including high temperature heat capacity. Implementation of this approach within the QMTP system for use with MM4 calculations would seem to be the next logical step for improving treatment of conformational flexibility.

8.3.2 Enumeration of conformers

A key aspect of such an implementation would be accurate and robust enumeration of all the minima on the molecule's potential energy surface, without double counting equivalent conformations. Many approaches have been developed to explore a molecule's potential energy surface with the goal of identifying low-lying local minima; a complete discussion is beyond the scope of this chapter, but potentially-promising methods include distance geometry,⁴⁵ stochastic proximity embedding,^{46-48, 142} the Monte Carlo method of Saunders,¹⁴³ and systematic searches over torsion angles. The Saunders approach is, in fact, built in to the MM4 program; however, internal tests suggest that this implementation can have some difficulties identifying essentially equivalent conformations as being equivalent. Also, this type of approach might be expected to be relatively slow at identifying new conformers compared to more directed approaches, such as systematic searches over torsion angles.

8.3.2.1 Conformational equivalence testing

Regardless of the chosen approach for exploring a molecule's potential energy surface, a key aspect of the conformer enumeration process is the identification of conformational equivalence to avoid double counting the same conformer. Toward this end, I have developed and implemented a novel conformational equivalence algorithm, designed for this particular application, in the MoleCoor utility.⁶⁶ This algorithm is described in further detail below.

8.3.2.1.1 Background

First, the definition of conformational equivalence used here should be made clear. Examples of equivalent and distinct conformers are shown in Figure 51.

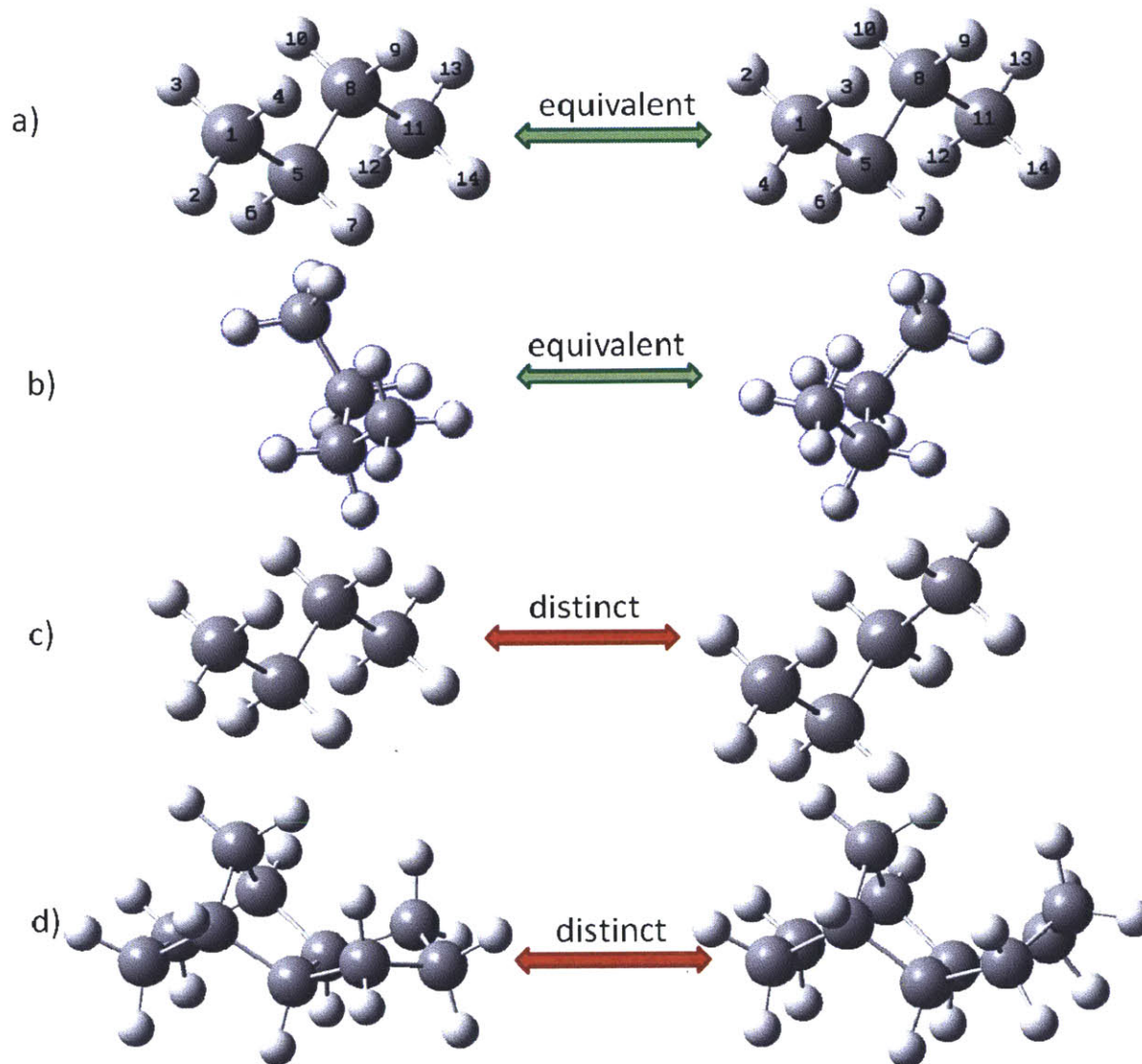


Figure 51. Examples of conformational equivalence as defined in text

In the figure, panel (a) shows two configurations of *n*-butane that differ with respect to atom numbering but are still considered equivalent; panel (b) shows that two mirror image conformations of gauche *n*-butane are considered equivalent here; panel (c) shows that the gauche and anti conformations of *n*-butane are distinct; finally, panel (d) shows that two conformations of JP-10 arising from ring flipping are also distinct. It should be emphasized that for our purposes, we consider mirror image conformers to be identical (e.g. panel (b) above), as although the two forms are topologically distinct and non-superposable, they are iso-energetic, and in practice, the two iso-energetic enantiomers may be accounted for by a simple statistical chirality correction (*i.e.* multiplying the partition function by two).

A number of approaches have been developed for testing conformational equivalence for various applications. MM4(2008), for example, determines conformational equivalence based upon structures having nearly the same energy and moment of inertia;⁷⁶ however, because these are necessary, but not sufficient, conditions for conformational equivalence, the approach is susceptible to false positive results in rare cases. Additionally, as alluded to previously, the implementation of conformational equivalence testing in MM4(2008) does not appear to be robust for our application and seems to produce many false negative results; in particular, the tolerance threshold for conformational equivalence is not adjustable, and the approach will unexpectedly find dozens of “different” conformations for *n*-butane. Another common approach is to use a molecular alignment procedure to minimize the root-mean-squared-deviation (RMSD) between the atomic coordinates; alignment may be performed using the well-known Kabsch algorithm.¹⁴⁴ The resulting RMSD of the optimally aligned structures then provides a measure of the (mis)alignment between two conformers which can be used, in conjunction with an appropriate tolerance threshold, to test for conformational equivalence.

In these approaches, it should be noted that the Kabsch algorithm requires a mapping between corresponding atoms in the two conformers to be compared. In general, the number of potential mappings is quite significant (for a molecule with *N* atoms, all the same type, there will be *N!* possible atom mappings), making an exhaustive search over all possible mappings computationally intractable. The program *TORMAT* uses connectivity information to perform graph matching and narrow down the number of possible atom mappings.¹⁴⁵ However, connectivity is often ambiguous, and such information can be unreliable or unavailable (e.g. transition states or molecules with hydrogen bonds). More recently, Bond developed an approach that utilizes the general assignment algorithm of Munkres¹⁴⁶ to obtain appropriate atom mappings and does not rely on *a priori* connectivity or atomic correspondence information;¹⁴⁷ however, in determining the best atom mapping, it relies on a global optimization approach, which is not, in general, guaranteed to converge to the global optimum. Additionally, the use of RMSD as an error metric can, itself, be seen as a limitation of these methods. Limitations of the RMSD metric have been noted elsewhere,¹⁴⁷⁻¹⁴⁸ and Bond, for example, citing Collins *et al.*, notes that “the r.m.s. deviation for the least-squares overlay may not always be the most revealing measure of molecular similarity, since it is an average measure and may conceal cases

where large parts of molecules are closely comparable but a small number of significant discrepancies exist”.

The approach described here and implemented in MoleCoor avoids these limitations: it relies instead on an alternative error metric based on maximum deviations between intraconformer atom-pair distances and is robust (guaranteed to produce an accurate result) without relying on connectivity information; it is well-suited to the application of conformational equivalence testing where we want to check whether two conformations are equivalent within some small tolerance.

8.3.2.1.2 The MoleCoor algorithm

As alluded to previously, the MoleCoor algorithm makes use of an error metric based on maximum deviations between intraconformer atom-pair distances. Let us consider distance matrices for each of the two conformers, labeled D and E , respectively. The distance matrices contain the Euclidean distances between atoms i and j , labeled, for example, D_{ij} . Assuming we have an atomic correspondence, mapping atoms in one conformer to the other, we will consider the two conformers to be equivalent if and only if:

$$\begin{aligned} \epsilon_{ij,a} &\leq \epsilon_a \forall i, j \text{ with} \\ \epsilon_{ij,a} &= |D_{ij} - E_{ij}| \end{aligned} \tag{Eq. 15}$$

In this expression, ϵ_a is an absolute tolerance specified by the user. Alternatively, the user can specify a relative tolerance, which will tolerate larger absolute deviations over large distances, while requiring smaller deviations over short ranges; this could be more appropriate for extended molecular systems, where small deviations in an angle can lead to large variations in positions of atoms connected through long arms from that angle. In this case, the requirement for conformational equivalence, given a user-specified relative tolerance, ϵ_r , is:

$$\begin{aligned} \epsilon_{ij,r} &\leq \epsilon_r \forall i, j \text{ with} \\ \epsilon_{ij,r} &= \frac{|D_{ij} - E_{ij}|}{\min(D_{ij}, E_{ij})} \end{aligned} \tag{Eq. 16}$$

The use of the minimum of D_{ij} and E_{ij} in the denominator ensures that the criterion is symmetric with respect to interchange of the two conformers. In the following discussion, the absolute case will be considered for the sake of conciseness and simplicity, but it is noted that the treatment when using relative tolerance readily follows by analogy.

The use of distance matrices in the comparison avoids the aforementioned averaging-related shortcomings of the RMSD metric. Also, it is a reasonable and intuitive metric in the context of energy calculations where energy is solely a function of the distance matrix. It should be noted that the concept of comparing distance matrices has been put forth previously;¹⁴³ however, the issue of establishing atomic correspondence for general cases that can arise, along with other details associated with the distance matrix comparison, do not seem to have been addressed.

So, given an atomic correspondence, the conformational equivalence test is relatively straightforward, and the conditions above may be readily evaluated to determine whether the criteria above are met; alternatively, if we are looking for a more quantitative answer, the quantities $\max(\epsilon_{ij,a})$ and $\max(\epsilon_{ij,r})$ are simple to compute. The challenge, which will be described next, is how to establish correct, consistent atomic correspondence between the two conformers.

As mentioned previously, if considered in a brute force manner, the problem suffers from combinatorial explosion in the number of possible atom mappings. The basic idea of the MoleCoor algorithm is to reduce the number of potential atom mappings considered by asking the Boolean question of conformational equivalence (“Is $\max(\epsilon_{ij,a}) \leq \epsilon_a$?”) rather than asking the quantitative question of how close two the two conformers are to each other (“What is $\max(\epsilon_{ij,a})$?”). We don’t directly obtain as much information from asking the Boolean question, but if the answer is “yes”, it turns out that we can obtain one or more viable atom mappings in the process of computing the answer, which in turn, can be used to quickly obtain the answer to the quantitative question, if desired.

The mechanism behind the approach for asking this Boolean question is demonstrated below for a simplistic case that omits the consideration of atom types.

$$\begin{array}{cc}
 \begin{array}{c} \text{Conformer A} \\ \left(\begin{array}{ccccc} 0 & D_{12} & D_{13} & D_{14} & D_{15} \\ & 0 & D_{23} & D_{24} & D_{25} \\ & & 0 & D_{34} & D_{35} \\ & & & 0 & D_{45} \\ & & & & 0 \end{array} \right) \end{array} &
 \begin{array}{c} \text{Conformer B} \\ \left(\begin{array}{ccccc} 0 & E_{1'2'} & E_{1'3'} & E_{1'4'} & E_{1'5'} \\ & 0 & E_{2'3'} & E_{2'4'} & E_{2'5'} \\ & & 0 & E_{3'4'} & E_{3'5'} \\ & & & 0 & E_{4'5'} \\ & & & & 0 \end{array} \right) \end{array} \\
 \\
 \begin{array}{c} \left(\begin{array}{ccccc} 0 & 1.00 & 3.14 & 2.00 & 1.41 \\ & 0 & 2.72 & 1.62 & 1.73 \\ & & 0 & 1.64 & 3.50 \\ & & & 0 & 4.00 \\ & & & & 0 \end{array} \right) &
 \begin{array}{c} \left(\begin{array}{ccccc} 0 & 1.41 & 1.73 & 3.50 & 4.00 \\ & 0 & 1.00 & 3.14 & 1.73 \\ & & 0 & 2.72 & 1.62 \\ & & & 0 & 1.64 \\ & & & & 0 \end{array} \right) \end{array}
 \end{array}
 \end{array}$$

$$\varepsilon_a = 0.01$$

Figure 52. Distance matrices for example illustrating MoleCoor algorithm mechanics; as the matrices are symmetric, the lower left hand portion has been omitted

Figure 52 illustrates the distance matrices that will be used for the example. In this case, we illustrate working with an absolute tolerance, $\varepsilon_a = 0.01$, though there is a mapping ($1 \leftrightarrow 2'$; $2 \leftrightarrow 3'$; $3 \leftrightarrow 4'$; $4 \leftrightarrow 5'$; $5 \leftrightarrow 1'$) with exactly equivalent intraconformer atom-pair distances (*i.e.* $\max(\varepsilon_{ij,a}) = 0$).

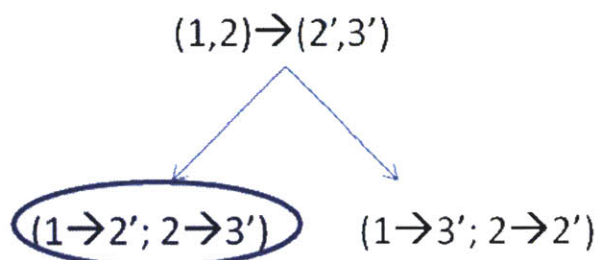
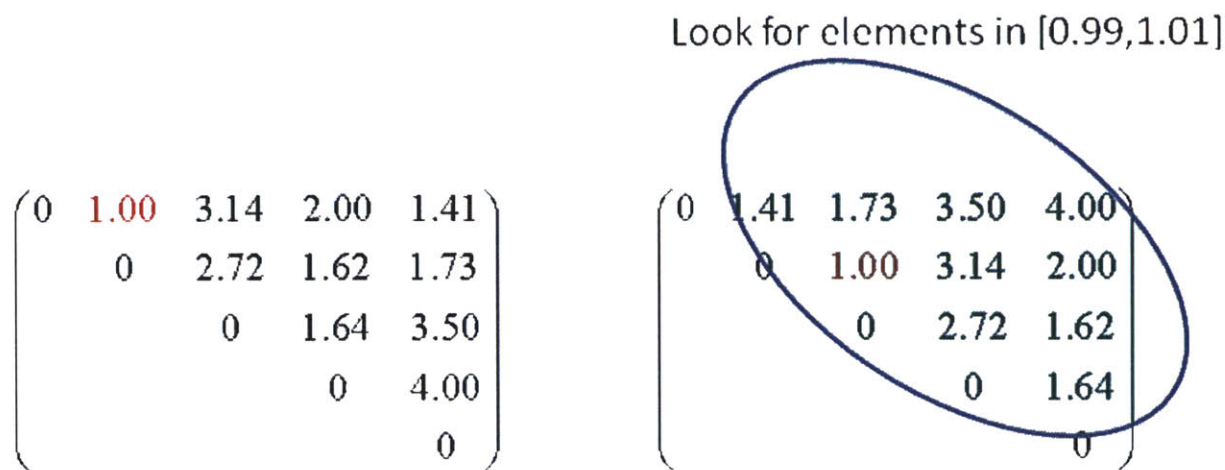


Figure 53. Iteration 0 for example illustrating MoleCooor algorithm mechanics

The 0th iteration of the algorithm is depicted in Figure 53. We start with an arbitrary element of the distance matrix for Conformer A, which includes a distance of 1.00. Based on the chosen ϵ_a , we look for elements of the Conformer B distance matrix in the range [0.99,1.01]. One such element is found, and two possible partial mappings are implied by this match: (1↔2'; 2↔3') or (1↔3'; 2↔2'). Each of these mappings is explored in the subsequent steps, though only the first will be considered here, as the second mapping leads to a “dead end” in the algorithm, and does not produce a viable atom mapping.

(1→2'; 2→3')

Part a:

$$\begin{pmatrix} 0 & 3.14 & 2.00 & 1.41 \\ & 0 & 2.72 & 1.62 & 1.73 \\ & & 0 & 1.64 & 3.50 \\ & & & 0 & 4.00 \\ & & & & 0 \end{pmatrix}$$

Look for elements involving 2' in [3.13,3.15]

$$\begin{pmatrix} 0 & 1.41 & 1.73 & 3.50 & 4.00 \\ & 0 & & 3.14 & 2.00 \\ & & 0 & 2.72 & 1.62 \\ & & & 0 & 1.64 \\ & & & & 0 \end{pmatrix}$$

Tentatively (1→2'; 2→3'; 3→4')

Part b:

$$\begin{pmatrix} 0 & & 2.00 & 1.41 \\ & 0 & 2.72 & 1.62 & 1.73 \\ & & 0 & 1.64 & 3.50 \\ & & & 0 & 4.00 \\ & & & & 0 \end{pmatrix}$$

Check that $E_{3,4}$ is in [2.71,2.73]

$$\begin{pmatrix} 0 & 1.41 & 1.73 & 3.50 & 4.00 \\ & 0 & & & 2.00 \\ & & 0 & 2.72 & 1.62 \\ & & & 0 & 1.64 \\ & & & & 0 \end{pmatrix}$$

Confirmed (1→2'; 2→3'; 3→4')

Figure 54. Iteration 1 for example illustrating MoleCoor algorithm mechanics

The 1st iteration of the algorithm with the first mapping is shown in Figure 54. We start with the (1,3) element of the Conformation A distance matrix, which, based on the proposed mapping, should correspond to an element involving 2' in the Conformation B distance matrix. Examining such matrix elements (again using the user-specified ϵ_a) produces one match, suggesting a mapping of (1↔2'; 2↔3'; 3↔4'). In the second portion of this iteration, we confirm this mapping by checking that all remaining distances involving the mapped atoms also match within the user-specified tolerance. A similar process repeats in subsequent iterations, as shown in Figure 55 and Figure 56.

(1→2'; 2→3'; 3→4')

Part a:

$$\begin{pmatrix} 0 & 2.00 & 1.41 \\ & 0 & 1.62 & 1.73 \\ & & 0 & 1.64 & 3.50 \\ & & & 0 & 4.00 \\ & & & & 0 \end{pmatrix}$$

Look for elements involving 2' in [1.99,2.01]

$$\begin{pmatrix} 0 & 1.41 & 1.73 & 3.50 & 4.00 \\ & 0 & & & 2.00 \\ & & 0 & & 1.62 \\ & & & 0 & 1.64 \\ & & & & 0 \end{pmatrix}$$

Tentatively (1→2'; 2→3'; 3→4'; 4→5')

Part b:

$$\begin{pmatrix} 0 & & & & 1.41 \\ & 0 & 1.62 & 1.73 & \\ & & 0 & 1.64 & 3.50 \\ & & & 0 & 4.00 \\ & & & & 0 \end{pmatrix}$$

Check that $E_{3,5}$ is in [1.61,1.63] and $E_{4,5}$ is in [1.63,1.65]

$$\begin{pmatrix} 0 & 1.41 & 1.73 & 3.50 & 4.00 \\ & 0 & & & 1.62 \\ & & 0 & & 1.64 \\ & & & 0 & \\ & & & & 0 \end{pmatrix}$$

Confirmed (1→2'; 2→3'; 3→4'; 4→5')

Figure 55. Iteration 2 for example illustrating MoleCoor algorithm mechanics

(1→2'; 2→3'; 3→4'; 4→5')

Part a:

$$\begin{pmatrix} 0 & & & & 1.41 \\ & 0 & & & 1.73 \\ & & 0 & & 3.50 \\ & & & 0 & 4.00 \\ & & & & 0 \end{pmatrix}$$

Look for elements involving 2' in [1.40,1.42]

$$\begin{pmatrix} 0 & 1.41 & 1.73 & 3.50 & 4.00 \\ & 0 & & & \\ & & 0 & & \\ & & & 0 & \\ & & & & 0 \end{pmatrix}$$

Tentatively (1→2'; 2→3'; 3→4'; 4→5'; 5→1')

Part b:

$$\begin{pmatrix} 0 & & & & & \\ & 0 & & & & 1.73 \\ & & 0 & & & 3.50 \\ & & & 0 & & 4.00 \\ & & & & 0 & \\ & & & & & 0 \end{pmatrix}$$

Check that $E_{1,3'}$ is in [1.72,1.74], $E_{1,4}$ in [3.49,3.51], and $E_{1,5}$ in [3.99,4.01]

$$\begin{pmatrix} 0 & & & & & \\ & 0 & & & & 1.73 \\ & & 0 & & & 3.50 \\ & & & 0 & & 4.00 \\ & & & & 0 & \\ & & & & & 0 \end{pmatrix}$$

Confirmed (1→2'; 2→3'; 3→4'; 4→5'; 5→1')

Figure 56. Iteration 3 for example illustrating MoleCoor algorithm mechanics

8.3.2.1.3 Results

Testing was performed on several cases of practical relevance to conformational equivalence testing. MoleCooor was used, with $\epsilon_a = 0.001 \text{ \AA}$, to compare the (independently optimized) mirror-image conformers shown in part (b) of Figure 51, and the expected “True” result was returned in under 2 milliseconds. The conformers shown in part (d) of Figure 51 were compared with $\epsilon_a = 0.10 \text{ \AA}$, and the expected “False” result was returned in under 7 milliseconds. In a third test, a three-dimensional molecular structure of buckminsterfullerene (C_{60}) was compared to itself, producing 120 unique viable atom mappings in about one second. The effect of the tolerance on the algorithm speed in the case of the conformers from part (d) of Figure 51 is shown in Figure 58.

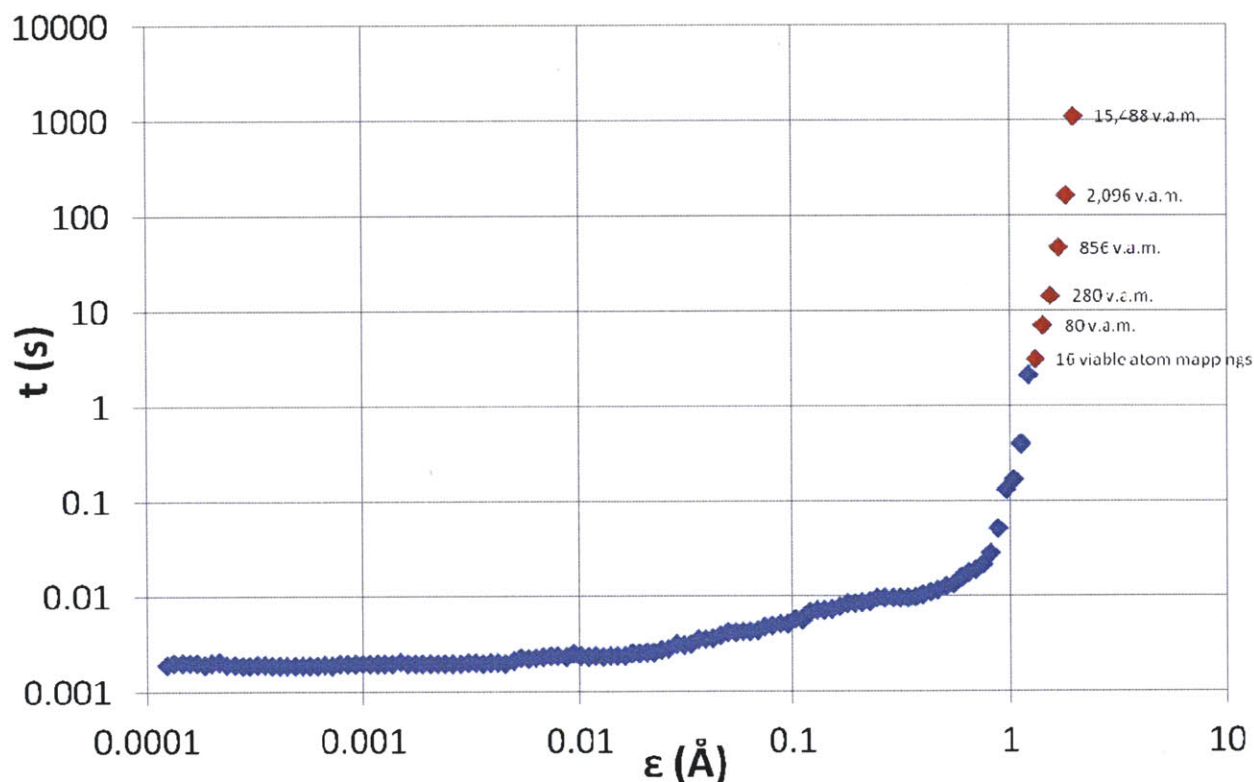


Figure 58. Effect of the absolute tolerance, ϵ_a , on MoleCooor algorithm speed for the conformers shown in part (d) of Figure 51

The graph demonstrates the dramatic effect of the tolerance on the speed of the algorithm. As the tolerance is loosened, the space of possible atom mappings is explored more broadly and deeply, resulting in longer time to return the result; in fact, the plot shows that if the tolerance is loosened excessively, viable atom mappings are identified. (For a given pair of conformers, the

computational cost of the algorithm monotonically increases with the tolerance, and deviations from monotonic increase in the figure are due to timing noise.) It is anticipated that when working with typical geometries from quantum mechanics programs, optimized with default tolerances using the same method and basis set, the tolerances can be chosen quite low (e.g. $\epsilon_a \sim 0.05 \text{ \AA}$ or $\epsilon_r \sim 0.01$).

The utility of the (dis)similarity metric used here was also evaluated, using the pair of distinct molecules ($C_{49}H_{94}O_6$) from the Cambridge Structural Database with refcode ABOPUC¹⁴⁹ considered by Bond¹⁴⁷. Using constraints of the type considered here (atom type constraints, but no connectivity constraint), Bond's approach produces a quite low RMSD of 1.2718 \AA , illustrating the aforementioned issues with averaging effects. In comparison, tests with the MoleCoor approach using the intraconformer atom-pair distance deviation metric suggest that $\max(\epsilon_{ij,a}) > 2.85 \text{ \AA}$; in particular, MoleCoor returns "False" in under a second with $\epsilon_a = 2.00 \text{ \AA}$ and was also found to produce a result of "False" with $\epsilon_a = 2.85 \text{ \AA}$; results with looser tolerances are not available due to excessive computational cost. This demonstrates how the alternative metric used here avoids the averaging issues associated with the use of the RMSD metric.

Additionally, testing was performed to assess best-case algorithm scaling with the number of atoms in the system. This best case scaling is possible when the user-specified tolerance is sufficiently small and there are $O(1)$ viable atom mappings. Results are shown in Figure 59. The slope of this log-log plot suggests that the actual scaling is fairly close to the expected theoretical scaling of $O(N^2)$ and are much better than hypothetical $O(N!)$ scaling.

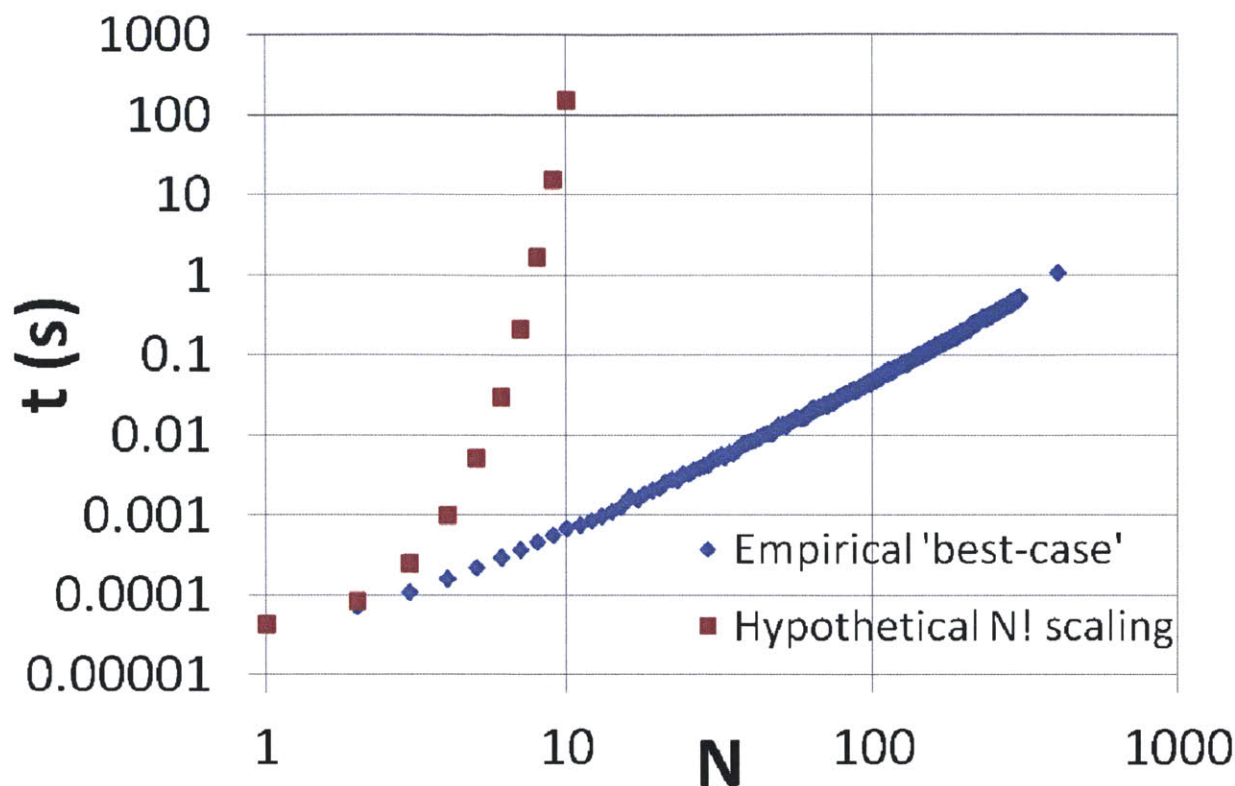


Figure 59. ‘Best-case’ scaling of MoleCoor algorithm

In the “worst case”, when the tolerance is sufficiently large that all possible atom mappings are explored, scaling will be worse than $O(N!)$.

8.3.2.1.4 Implementation

The MoleCoor algorithm is implemented in the publicly available *MoleCoor* utility hosted at <https://github.com/gmagoon/MoleCoor>. The utility is written in Python 2.x and also includes several unit tests. The *MoleCoor* utility accepts MM4, MOL and XYZ formats as input; additionally, the utility has file-conversion functionality and can produce MM4, MOL, and XYZ formats.

8.3.2.1.5 Summary

The MoleCoor algorithm introduced here appears to be well-suited to conformational equivalence testing of the type that would be desired when performing conformational sampling using a consistent force field or electronic structure method to identify unique conformations. More generally, the algorithm works well if the goal is just to robustly and quickly determine

whether two conformers are equivalent within some (small) user-specified tolerance; on the other hand, the algorithm is far from ideal for quantifying deviations for different conformations of large molecules with no atomic correspondence information provided. For application to automated thermodynamic parameter estimation based on three-dimensional-structure, the tool could be practically useful and determines conformational equivalence for reasonably sized molecules within a fraction of a second.

8.4 Improved kinetic parameter estimates based on explicit three-dimensional structures

Work with the JP-10 system brought attention to a number of reaction families in which kinetic parameter estimates can be significantly impacted by steric effects and three-dimensional geometry considerations; it is expected that the bulk of these effects (and the biggest opportunity for improvement) is associated with the intramolecular reaction families including: intramolecular hydrogen migration, intramolecular disproportionation, intramolecular radical addition to multiple bond (the reverse of ring opening by beta-scission), and biradical recombination (with its reverse, ring opening). In fact, it can be argued that these opportunities for improvement in kinetic parameter estimation are analogous to the limitations of group-additivity for thermodynamic parameter estimation; in particular, the groups currently used in kinetic parameter estimation for these families all involve a ring component, where the size of the ring can affect the estimated parameters.

It is envisaged that kinetic parameter estimation based on explicit three-dimensional structures could follow either of two different approaches: minimum-based and saddle-point-based. Additionally, three major challenges can be associated with these approaches:

1. Combinatorial nature of the kinetic parameter estimation problem: The number of reactions for which kinetic parameter estimation must be performed is at least an order of magnitude higher than the number of species for which thermodynamic parameter estimation must be performed.
2. Automated saddle-point location: Kinetic parameter estimation may require automated routines to reliably and robustly locate saddle-points on a potential energy surface. This is analogous to the issues associated with generating an initial guess geometry,

optimizing to a minimum, and checking the resulting structure, which are addressed by the QMTP system, for thermodynamic parameter estimation.

3. Reliable method for reactive potential energy surface (PES): Availability of methods that produce a potential energy surface that is sufficiently reliable for treating chemical reactions and transition states may need to be considered.

The minimum-based approach is to use properties of a potential energy surface minimum (with structural properties resembling those of the transition state of interest) to estimate kinetic parameters. Efforts towards using such an approach for intramolecular disproportionation reaction barrier estimates were discussed in the previous chapter. This approach avoids the second and third challenges discussed above, and its use of minima rather than saddle-points should make it relatively robust. However, this approach relies on correlations (as opposed to rigorous formulas) of a kinetic parameter, such as the energy barrier, with energetic properties, and the theoretical foundations aren't as solid; these correlations also require additional work to develop and validate. Additionally, although a minimum-based approach for intramolecular disproportionation was discussed in the previous chapter, generalization of this approach to other important reaction families may not be as straightforward.

The alternative, saddle-point-based approach involves use of properties of a saddle-point on the potential energy surface, corresponding to the transition state for the reaction of interest, for estimation of kinetic parameters. Through transition state theory (TST), or one of its extensions, this approach has a much more solid theoretical basis; also this approach can be generalized to any reaction involving a tight transition state. However, this approach also brings in additional challenges through the second and third items mentioned previously: automated saddle point location and the availability of reliable PES methods.

A method to perform robust, automated saddle-point location would be needed with the saddle-point-based approach. As with energy minima, most transition state optimization methods require a good initial guess molecular geometry. It is proposed here that the use of distance geometry could address this issue. For instance, the rules for assigning distance bounds in existing distance geometry programs (such as those discussed previously) could be modified to have special distance, angle, and/or dihedral constraints around breaking and forming bonds for a particular reaction to locate an initial guess to that reaction's transition state. An alternative (or perhaps backup) approach could involve the use of interpolation methods to locate a saddle-point

based only on reactant and product structures.¹⁵⁰ Promising interpolation methods include the Synchronous Transit-Guided Quasi Newton (STQN) method developed by Schlegel and coworkers (implemented in Gaussian03 with keywords QST2)^{26, 151} and the growing string method of Peters *et al.*¹⁵⁰ In either case, in order to avoid use of incorrect saddle-points, one needs to be able to obtain confirmation that the structure obtained by these methods is the transition state for the desired reaction. This is analogous to the issue of connectivity checking for PES minima. Analyzing the eigenvector of the vibrational mode with imaginary frequency is probably the fastest and most straightforward method; the vibrational motion should correspond to motion along the reaction path, which should in many cases correspond to translation associated with the breaking and forming bonds. A more computationally demanding, though more rigorous, approach would be to automate intrinsic reaction coordinate (IRC) calculations to march along the reaction path towards reactants and products to confirm that these structures have the desired connectivity.

The saddle-point-approach also places special demands on the method used to evaluate the potential energy surface, as it must be reliable for modeling reactive processes. As discussed in the previous chapter, all but the most advanced *ab initio* methods appear to be unreliable for modeling intramolecular disproportionation, likely making this reaction family ill-suited to a saddle-point-based approach with the current state of technology and research; however, other reaction families, such as intramolecular hydrogen migration, might be more amenable to treatment using low-level electronic structure methods such as PM3. As with thermodynamic parameter estimation, there is also the opportunity to use force fields to evaluate the potential energy surface. Most force-fields, including MM4, are not designed to model reactions, and therefore, if applied to determining properties of reaction saddle-points, results would be unreliable. An exception to this is ReaxFF, which is a force field specifically designed to model reactive processes.¹⁵² The MMX force field developed by Gilbert and Gajewski (available in PCMODEL from Serena Software) also includes atom types for certain types of transition states.¹⁵³ Alternatively, a method has been proposed to develop novel parametrizations of Allinger's MM2 and MM3 force fields suitable for transition state determination.^{153b, 154}

8.5 Main contributions

The main contributions of this thesis work are summarized below:

- Improved group additivity thermodynamic parameter estimates for branched species in RMG via implementation of a method of accounting for non-nearest neighbor steric effects, partly by developing an approach to map bond-centered groups/group values to atom-centered groups/group values
- Investigated approaches for three-dimensional structure generation and explored tradeoffs between the methods, within the context of automated reaction mechanism generation
- Interfaced automated reaction mechanism generation code with an appropriate three-dimensional structure generation tool to reliably and robustly provide three-dimensional structures for purposes of kinetic modeling
- Designed and implemented a robust next-generation approach to on-the-fly thermochemistry parameter estimation using explicit three-dimensional structures during automated reaction mechanism generation with RMG, making the estimates more reliable and general than previous approaches within RMG
- Developed interface between MM4 and automatic reaction mechanism generation software to improve accuracy of thermochemistry estimates in many cases
- Demonstrated an automatic method of accounting for conformational flexibility within separable hindered rotor model, using automated, on-the-fly rotor scans with MM4 force-field
- Designed and implemented a novel algorithm for testing conformational equivalence, which could potentially be combined with newly-developed statistical-mechanics models to provide an alternative approach to account for conformational flexibility
- Applied explicit-3D-structure-based estimation tools within automated reaction mechanism generation to study combustion of the widely-used jet fuel, JP-10, producing a significantly more comprehensive and reliable model than previously available
- Identified novel concerted pathways for JP-10 decomposition and explored their significance compared to previously-identified biradical pathways
- Obtained insights into intramolecular disproportionation and disproportionation using high-level *ab initio* multireference methods

- Explored potential approaches to on-the-fly kinetic parameter estimation using explicit three-dimensional structures

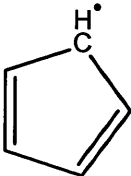
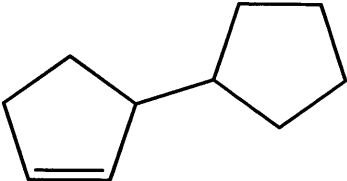
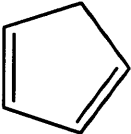
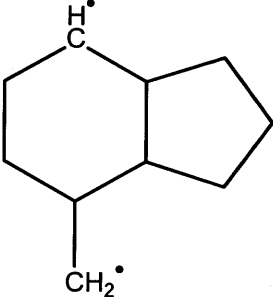
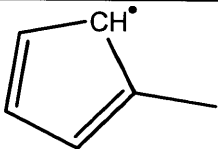
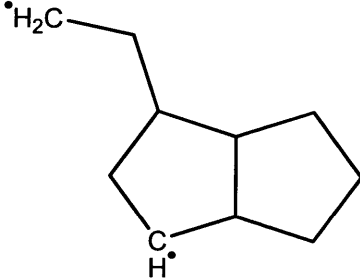
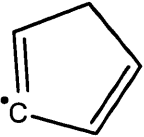
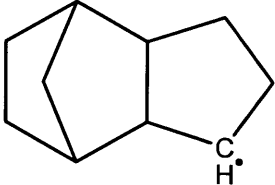
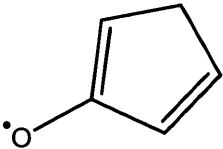
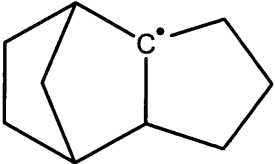
8.6 Conclusions

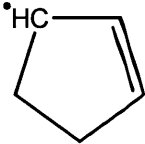
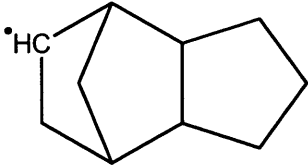
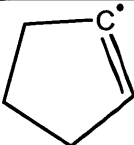
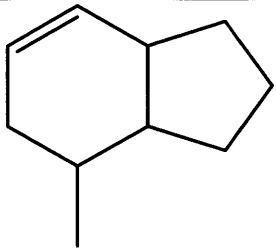
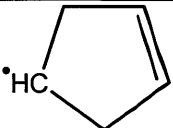
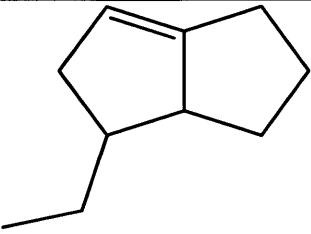
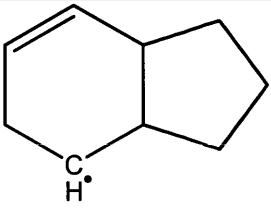
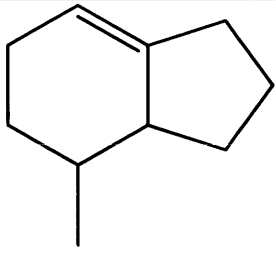
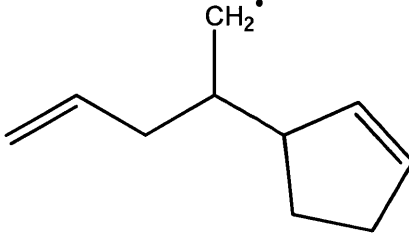
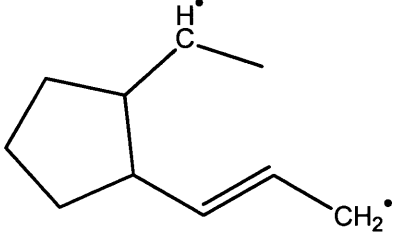
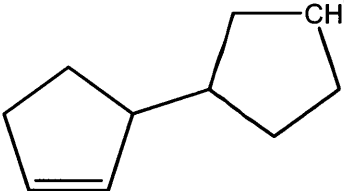
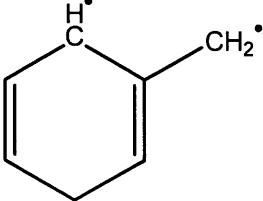
As summarized in the previous section, this work has meaningfully and significantly contributed to the field of automated reaction mechanism generation in multiple ways. New capabilities have been added to RMG, parameter estimates have been refined, and a system has been implemented for robustly and quickly performing on-the-fly thermodynamic calculations using explicit three-dimensional geometries. These refinements allow a wider range of systems to be considered more reliably than previously. These updated tools have been applied to redefine the state-of-the-art in JP-10 combustion modeling. Finally, a number of opportunities for further refinement have been identified and the identification of these opportunities can help to chart a path for developing the next generation of automated reaction mechanism generation tools.

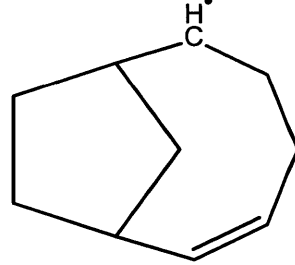
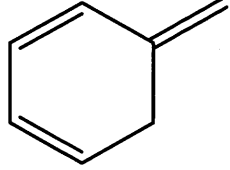
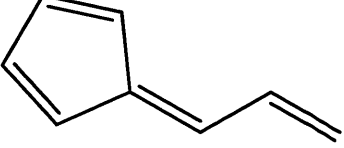
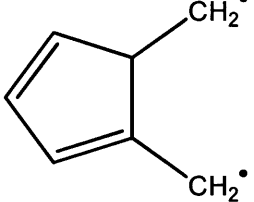
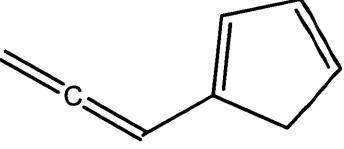
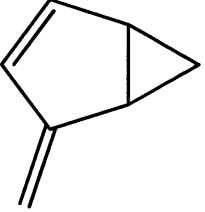
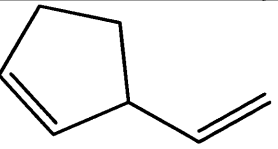
Ultimately, models constructed using the automated reaction mechanism generation tools described here can have a practical impact for systems in a number of areas, including defense, energy, and transportation, by providing insight into the relevant chemistry and assisting in the engineering of improved performance.

9 Appendix I: Glossary for ambiguous species names

The tables discussing JP-10 model sensitivity analysis results in Chapter 6 include some species names for which the corresponding structure is not immediately obvious. The short name and structure for these species are summarized in the table below.

Name	Structure	Name	Structure
cyclopentadienyl		MA110	
cyclopentadiene		BR2	
C ₅ H ₄ CH ₃		BR3	
C ₅ H ₅ _A		JP10R5	
C ₅ H ₅ O		JP10R6	

$C_5H_7_A$		JP10R8	
$C_5H_7_B$		$C_{10}H_{16}_A$	
$C_5H_7_C$		$C_{10}H_{16}_B$	
C_9H_{13}		$C_{10}H_{16}_C$	
$C_{10}H_{15}_A$		$C_{10}H_{16}_D$	
$C_{10}H_{15}_B$		$C_7H_8_A$	

$C_{10}H_{15}_C$		$C_7H_8_B$	
$C_8H_8_A$		$C_7H_8_C$	
$C_8H_8_B$		$C_7H_8_D$	
		C_7H_{10}	

10 Appendix II: Derivation of matrix formulation for variable moment of inertia treatment of hindered rotors

The kinetic energy term in the Hamiltonian for a one-dimensional rotor with reduced moment of inertia that varies with the dihedral angle may be written as follows¹⁵⁵:

$$-\frac{\hbar^2}{8\pi^2} \frac{d}{d\phi} \left[\frac{1}{I_j(\phi)} \frac{d\Psi(\phi)}{d\phi} \right] = -\frac{\hbar^2}{8\pi^2} \left[\frac{1}{I_j(\phi)} \frac{d^2\Psi(\phi)}{d\phi^2} + \frac{d}{d\phi} \left(\frac{1}{I_j(\phi)} \right) \frac{d\Psi(\phi)}{d\phi} \right]$$

We now consider a matrix formulation of the Schrodinger equation.^{155b} The CanTherm v1.0 manual describes the case with a constant reduced moment of inertia. Here, we use the same nomenclature, but consider the effect of a variable reduced moment of inertia. The potential energy term is unaffected by the relaxation of the constant reduced moment of inertia approximation. To facilitate expression of the refined kinetic energy component in the matrix formulation, the inverse of the reduced moment of inertia for the rotor is expanded as a Fourier series:

$$\frac{1}{I_j(\phi)} = B + \sum_{k=1}^N (\alpha_k \cos(k\phi) + \beta_k \sin(k\phi)) = B + \sum_{k=1}^N \left(\frac{(\alpha_k - i\beta_k)}{2} e^{ik\phi} + \frac{(\alpha_k + i\beta_k)}{2} e^{-ik\phi} \right)$$

(Since the development and implementation of this treatment, it was found that the approach of expanding the inverse of the reduced moment of inertia as a Fourier series has been used previously in the literature.^{155a, 156})

Now, the two kinetic energy terms will be considered separately. First, we consider the first kinetic energy term in the product rule expansion above:

$$\frac{\hbar^2}{8\pi^2} \frac{1}{I_j(\phi)} \frac{d^2\Psi(\phi)}{d\phi^2}$$

The first kinetic energy term of the Hamiltonian may be expressed in the matrix formulation, with row n defined as follows:

$$\sum_{m'=-m}^m \left[\frac{\hbar^2 B (m')^2}{8\pi^2} c_{m'} \delta_{n,m'} + \sum_{k=1}^N \left(\frac{\hbar^2 \left(\frac{\alpha_k - i\beta_k}{2} \right) (m')^2}{8\pi^2} c_{m'} \delta_{k,n-m'} + \frac{\hbar^2 \left(\frac{\alpha_k + i\beta_k}{2} \right) (m')^2}{8\pi^2} c_{m'} \delta_{k,m'-n} \right) \right]$$

Now, the second kinetic energy term of the Hamiltonian is:

$$\frac{h^2}{8\pi^2} \frac{d}{d\phi} \left(\frac{1}{I_j(\phi)} \right) \frac{d\Psi(\phi)}{d\phi}$$

This second term may be expressed in the matrix formulation with row n defined as follows:

$$\sum_{m'=-m}^m \left[\sum_{k=1}^N \left(\frac{h^2 \left(\frac{\alpha_k - i\beta_k}{2} \right) (km')}{8\pi^2} c_m \delta_{k,n-m'} + \frac{h^2 \left(\frac{\alpha_k + i\beta_k}{2} \right) (-km')}{8\pi^2} c_m \delta_{k,m'-n} \right) \right]$$

The total kinetic energy term in the matrix formulation is then the sum of these two, with row n defined as follows:

$$\begin{aligned} & \sum_{m'=-m}^m \left[\frac{h^2 B(m')^2}{8\pi^2} c_m \delta_{n,m'} + \sum_{k=1}^N \left(\frac{h^2 \left(\frac{\alpha_k - i\beta_k}{2} \right) ((m')^2 + m'k)}{8\pi^2} c_m \delta_{k,n-m'} + \frac{h^2 \left(\frac{\alpha_k + i\beta_k}{2} \right) ((m')^2 - m'k)}{8\pi^2} c_m \delta_{k,m'-n} \right) \right] \\ & = \sum_{m'=-m}^m \left[\frac{h^2 B(m')^2}{8\pi^2} c_m \delta_{n,m'} + \sum_{k=1}^N \left(\frac{h^2 \left(\frac{\alpha_k - i\beta_k}{2} \right) (m'n)}{8\pi^2} c_m \delta_{k,n-m'} + \frac{h^2 \left(\frac{\alpha_k + i\beta_k}{2} \right) (m'n)}{8\pi^2} c_m \delta_{k,m'-n} \right) \right] \end{aligned}$$

The second and third terms in the brackets in the final expression represent off-diagonal terms in the kinetic energy matrix that are not present in the treatment with constant reduced moment of inertia.

This treatment has been implemented in CanTherm (with $N=5$) to work with one-dimensional rotor scans, as discussed in Chapter 5. The added computational cost of relaxing the assumption of constant moment of inertia is not significant, as the new off-diagonal kinetic energy terms are located at the same positions as the (generally) non-zero potential energy entries, so the sparsity of the Hamiltonian matrix should be preserved. It is noted however, that a cancellation of errors has been noted wherein variation of the reduced moment of inertia is compensated for by variation in vibrational frequencies through rotation about the bond^{80, 157}; this treatment only addresses one of these approximations.

References

1. Allen, J. W.; Ashcraft, R. W.; Beran, G. J.; Goldsmith, C. F.; Harper, M. R.; Jalan, A.; Magoon, G. R.; Matheu, D. M.; Petway, S.; Sumathy, R.; Sharma, S.; Van Geem, K. M.; Song, J.; Wen, J.; West, R. H.; Wong, A.; Wong, H.-W.; Yelvington, P. E.; Yu, J.; Green, W. H. *RMG (Reaction Mechanism Generator)*, 2009.
2. Song, J. *Building Robust Chemical Reaction Mechanisms: Next Generation of Automatic Model Construction Software*. Massachusetts Institute of Technology, Cambridge, MA, 2004.
3. Warth, V.; Stef, N.; Glaude, P. A.; Battin-Leclerc, F.; Scacchi, G.; Côme, G. M., Computer-Aided Derivation of Gas-Phase Oxidation Mechanisms: Application to the Modeling of the Oxidation of n-Butane. *Combust Flame* **1998**, *114* (1–2), 81-102.
4. Susnow, R. G.; Dean, A. M.; Green, W. H.; Peczak, P.; Broadbelt, L. J., Rate-Based Construction of Kinetic Models for Complex Systems. *The Journal of Physical Chemistry A* **1997**, *101* (20), 3731-3740.
5. (a) Allen, J. W.; Goldsmith, C. F.; Green, W. H., Automatic estimation of pressure-dependent rate coefficients. *Phys Chem Chem Phys* **2012**, *14* (3); (b) Matheu, D. M.; Dean, A. M.; Grenda, J. M.; Green, W. H., Mechanism Generation with Integrated Pressure Dependence: A New Model for Methane Pyrolysis. *The Journal of Physical Chemistry A* **2003**, *107* (41), 8552-8565; (c) Matheu, D. M.; Green, W. H.; Grenda, J. M., Capturing pressure-dependence in automated mechanism generation: Reactions through cycloalkyl intermediates. *Int J Chem Kinet* **2003**, *35* (3), 95-119.
6. Harper, M. R. *Automated Reaction Mechanism Generation: Data Collaboration, Heteroatom Implementation and Model Validation*. Massachusetts Institute of Technology, Cambridge, MA, 2011.
7. Van Geem, K. M.; Reyniers, M.-F.; Marin, G. B.; Song, J.; Green, W. H.; Matheu, D. M., Automatic reaction network generation using RMG for steam cracking of n-hexane. *AIChE Journal* **2006**, *52* (2), 718-730.
8. Sharma, S.; Harper, M. R.; Green, W. H., Modeling of 1,3-hexadiene, 2,4-hexadiene and 1,4-hexadiene-doped methane flames: Flame modeling, benzene and styrene formation. *Combust Flame* **2010**, *157* (7), 1331-1345.

9. (a) 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. *Combust Flame* **2011**, *158* (1), 16-41; (b) Van Geem, K. M.; Pyl, S. P.; Marin, G. B.; Harper, M. R.; Green, W. H., Accurate High-Temperature Reaction Networks for Alternative Fuels: Butanol Isomers. *Ind Eng Chem Res* **2010**, *49* (21), 10399-10420.
10. Cohen, N.; Benson, S. W., Estimation of Heats of Formation of Organic Compounds by Additivity Methods. *Chemical Reviews* **1993**, *93*, 2419-2438.
11. (a) Sabbe, M. K.; De Vleeschouwer, F.; Reyniers, M.-F.; Waroquier, M.; Marin, G. B., First Principles Based Group Additive Values for the Gas Phase Standard Entropy and Heat Capacity of Hydrocarbons and Hydrocarbon Radicals. *J Phys Chem A* **2008**, *112* (47), 12235-12251; (b) Khan, S. S.; Yu, X.; Wade, J. R.; Malmgren, R. D.; Broadbelt, L. J., Thermochemistry of Radicals and Molecules Relevant to Atmospheric Chemistry: Determination of Group Additivity Values using G3//B3LYP Theory. *J Phys Chem A* **2009**, *113*, 5176-5194.
12. Domalski, E. S.; Hearing, E. D., Estimation of the Thermodynamic Properties of Hydrocarbons at 298.15 K. *J Phys Chem Ref Data* **1988**, *17* (4), 1637-1678.
13. Benson, S. W.; Buss, J. H., Additivity Rules for the Estimation of Molecular Properties. Thermodynamic Properties. *J Chem Phys* **1958**, *29*.
14. Benson, S. W.; Cruickshank, F. R.; Golden, D. M.; Haugen, G. R.; O'Neal, H. E.; Rodgers, A. S.; Shaw, R.; Walsh, R., Additivity rules for the estimation of thermochemical properties. *Chemical Reviews* **1969**, *69* (3), 279-324.
15. Benson, S. W., In *Thermochemical Kinetics*, 2nd ed., Wiley: New York, NY, 1976.
16. Cohen, N.; Benson, S. W., The thermochemistry of alkanes and cycloalkanes. In *The Chemistry of Alkanes and Cycloalkanes*, Patai, S.; Rappoport, Z., Eds. John Wiley & Sons Ltd: 1992.
17. Green, D. W.; Perry, R. H., *Perry's Chemical Engineers' Handbook* (8th Edition). McGraw-Hill: 2008.
18. Domalski, E. S.; Hearing, E. D., Estimation of the Thermodynamic Properties of C-H-N-O-S-Halogen Compounds at 298.15 K. *J Phys Chem Ref Data* **1993**, *22* (4), 805-1159.
19. (a) Oehlschlaeger, M. A.; Steinberg, J.; Westbrook, C. K.; Pitz, W. J., The autoignition of iso-cetane at high to moderate temperatures and elevated pressures: Shock tube experiments and kinetic modeling. *Combust Flame* **2009**, *156* (11), 2165-2172; (b) Westbrook, C. K.; Pitz, W. J.;

- Mehl, M.; Curran, H. J., Detailed chemical kinetic reaction mechanisms for primary reference fuels for diesel cetane number and spark-ignition octane number. *P Combust Inst* **2011**, *33* (1), 185-192; (c) Ranzi, E.; Frassoldati, A.; Faravelli, T.; Cuoci, A., Lumped Kinetic Modeling of the Oxidation of Isocetane (2,2,4,4,6,8,8-Heptamethylnonane) in a Jet-Stirred Reactor (JSR). *Energ Fuel* **2009**, *23* (10), 5287-5289; (d) Pitz, W. J.; Mueller, C. J., Recent progress in the development of diesel surrogate fuels. *Progress in Energy and Combustion Science* **2011**, *37* (3), 330-350.
20. Design Institute for Physical Properties, S. b. A., DIPPR Project 801 - Full Version. Design Institute for Physical Property Research/AIChE.
21. Pitz, W. J.; Herbinet, O., personal communication. 2008.
22. Somayajulu, G. R.; Zwolinski, B. J., Generalized Treatment of Alkanes. *Transactions of the Faraday Society* **1966**, *62*, 2327-2340.
23. Allinger, N. L.; Chen, K.; Lii, J.-H., An Improved Force Field (MM4) for Saturated Hydrocarbons. *J Comput Chem* **1996**, *17* (5&6), 642-668.
24. Allinger, N. L.; Schmitz, L. R.; Motoc, I.; Bender, C.; Labanowski, J. K., Heats of formation of organic molecules. 2. The basis for calculations using either ab initio or molecular mechanics methods. Alcohols and ethers. *J Am Chem Soc* **1992**, *114* (8), 2880-2883.
25. Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A., A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. *J Chem Phys* **1999**, *110* (6), 2822-2827.
26. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.;

- Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 03, Revision D.01*, Gaussian, Inc.: Wallingford CT, 2003.
27. Sabbe, M. K.; Saeys, M.; Reyniers, M.-F.; Marin, G. B.; Van Speybroeck, V.; Waroquier, M., Group Additive Values for the Gas Phase Standard Enthalpy of Formation of Hydrocarbons and Hydrocarbon Radicals. *J Phys Chem A* **2005**, *109* (33), 7466-7480.
28. Kroenlein, K.; Muzny, C. D.; Kazakov, A. F.; Diky, V.; Magee, R. D. C. J. W.; Abdulagatov, I.; Frenkel, M., NIST/TRC Web Thermo Tables (WTT), NIST Standard Reference Subscription Database 3 - Professional Edition, Version 2-2011-3-Pro. Thermodynamics Research Center (TRC), T. P. D., National Institute of Standards and Technology Ed. Boulder, CO, 2011.
29. 2,2,4,4,6,8,8-Heptamethylnonane model thermodynamic parameters. https://www-pls.llnl.gov/data/docs/science_and_technology/chemistry/combustion/isocetane_2009-06-21_therm_dat_v15.txt.
30. Desai, P. D. Enthalpies of Combustion of Eight Branched Isomeric Alkanes from C-9 to C-16. Texas A&M University, 1968.
31. *TRC thermodynamic tables. Hydrocarbons*. Thermodynamics Research Center, Texas Engineering Experiment Station, Texas A&M University System: College Station, TX, 1985, with various loose-leaf updates.
32. Ashcraft, R. W. Ab initio modeling of complex aqueous and gaseous systems containing nitrogen. Massachusetts Institute of Technology, 2008.
33. Yu, J.; Sumathi, R.; Green, W. H., Accurate and Efficient Method for Predicting Thermochemistry of Polycyclic Aromatic Hydrocarbons – Bond-Centered Group Additivity. *J Am Chem Soc* **2004**, *126* (39), 12685-12700.
34. Broadbelt, L. J.; Stark, S. M.; Klein, M. T., Computer generated reaction networks: on-the-fly calculation of species properties using computational quantum chemistry. *Chemical Engineering Science* **1994**, *49* (24, Part 2), 4991-5010.
35. Broadbelt, L. J., personal communication.
36. Vandewiele, N. Multiple Ring Strain Corrections in single molecule. <https://github.com/GreenGroup/RMG-Java/pull/209>.
37. Sadowski, J.; Gasteiger, J., From atoms and bonds to three-dimensional atomic coordinates: automatic model builders. *Chemical Reviews* **1993**, *93* (7), 2567-2581.

38. Leite, T. B.; Gomes, D.; Miteva, M. A.; Chomilier, J.; Villoutreix, B. O.; Tufféry, P., Frog: a FRee Online druG 3D conformation generator. *Nucleic Acids Res* **2007**, *35* (suppl 2), W568-W572.
39. Molecular Networks, G. *CORINA*, <http://www.molecular-networks.com/software/corina/index.html> Erlangen, Germany.
40. Mizutani, M. Y.; Nakamura, K.; Ichinose, T.; Itai, A., Starting Point to Molecular Design: Efficient Automated 3D Model Builder Key3D. *Chemical & Pharmaceutical Bulletin* **2006**, *54* (12), 1680-1685.
41. Vainio, M. J.; Johnson, M. S., Generating Conformer Ensembles Using a Multiobjective Genetic Algorithm. *Journal of Chemical Information and Modeling* **2007**, *47* (6), 2462-2474.
42. Landrum, G. *RDKit*, <http://rdkit.org>.
43. Gilbert, K.; Guha, R. *smi23d*, <http://www.chembiogrid.org/cheminfo/smi23d/>.
44. (a) Crippen, G. M., Chemical distance geometry: Current realization and future projection. *Journal of Mathematical Chemistry* **1991**, *6* (1), 307-324; (b) Svrcek-Seiler, W. A. 0.02 € on Embedding 2004. <http://www.tbi.univie.ac.at/Bled/Slides04/svrci.pdf>.
45. Spellmeyer, D. C.; Wong, A. K.; Bower, M. J.; Blaney, J. M., Conformational analysis using distance geometry methods. *Journal of Molecular Graphics and Modelling* **1997**, *15* (1), 18-36.
46. (a) Agrafiotis, D. K., Stochastic proximity embedding. *J Comput Chem* **2003**, *24* (10), 1215-1221; (b) Agrafiotis, D. K.; Xu, H.; Zhu, F.; Bandyopadhyay, D.; Liu, P., Stochastic Proximity Embedding: Methods and Applications. *Molecular Informatics* **2010**, *29* (11), 758-770.
47. Xu, H.; Izrailev, S.; Agrafiotis, D. K., Conformational Sampling by Self-Organization. *Journal of Chemical Information and Computer Sciences* **2003**, *43* (4), 1186-1191.
48. Agrafiotis, D. K.; Gibbs, A. C.; Zhu, F.; Izrailev, S.; Martin, E., Conformational Sampling of Bioactive Molecules: A Comparative Study. *Journal of Chemical Information and Modeling* **2007**, *47* (3), 1067-1086.
49. (a) Agrafiotis, D. K.; Lobanov, V. S.; Salemme, F. R. Method, system and computer program product for non-linear mapping of multi-dimensional data, U.S. Pat. No. 6,571,227. 2003; (b) Agrafiotis, D. K.; Xu, H.; Izrailev, S.; Salemme, F. R. Conformational sampling by self-organization, U.S. Pat. App. 20060089808. 2006.

50. Agrafiotis, D. K., Looking for simple conformation analysis tool. In *CCL.NET*, 2007.
51. Gothe, S. A.; Helson, H. E.; Houdaverdis, I.; Lagerstedt, I.; Sinclair, S.; Jorgensen, W. L., Computer-assisted mechanistic evaluation of organic reactions. 22. The generation and use of three-dimensional structures. *The Journal of Organic Chemistry* **1993**, *58* (19), 5081-5094.
52. Ricketts, E. M.; Bradshaw, J.; Hann, M.; Hayes, F.; Tanna, N.; Ricketts, D. M., Comparison of conformations of small molecule structures from the Protein Data Bank with those generated by Concord, Cobra, ChemDBS-3D, and Converter and those extracted from the Cambridge Structural Database. *Journal of Chemical Information and Computer Sciences* **1993**, *33* (6), 905-925.
53. Sadowski, J.; Gasteiger, J.; Klebe, G., Comparison of Automatic Three-Dimensional Model Builders Using 639 X-ray Structures. *Journal of Chemical Information and Computer Sciences* **1994**, *34* (4), 1000-1008.
54. Saunders, M.; Jarret, R. M., A New Method for Molecular Mechanics. *J Comput Chem* **1986**, *7* (4), 578-588.
55. Irikura, K. K. Computed 3-D Structures. <http://webbook.nist.gov/chemistry/3d-structs/#back5>.
56. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G., Open Babel: An open chemical toolbox. *Journal of Cheminformatics* **2011**, *3*, 1-14.
57. RDKit bug tracker. http://sourceforge.net/tracker/?group_id=160139&atid=814650.
58. Ebejer, J.-P.; Morris, G. M.; Deane, C., Freely Available Conformer Generation Methods: How good are they ? *Journal of Chemical Information and Modeling* **2012**.
59. Irikura, K. K.; Frurip, D. J., Computational thermochemistry: prediction and estimation of molecular thermodynamics. American Chemical Society (via <http://cccbdb.nist.gov/thermo.asp>): 1998.
60. (a) Maranas, C. D.; Floudas, C. A., A deterministic global optimization approach for molecular structure determination. *The Journal of Chemical Physics* **1994**, *100* (2), 1247-1261; (b) Lin, Y.; Stadtherr, M. A., Deterministic global optimization of molecular structures using interval analysis. *J Comput Chem* **2005**, *26* (13), 1413-1420.
61. O'Boyle, N. M.; Tenderholt, A. L.; Langner, K. M., cclib: A library for package-independent computational chemistry algorithms. *J Comput Chem* **2008**, *29* (5), 839-845.

62. (a) Stewart, J. J. P., Optimization of parameters for semiempirical methods I. Method. *J Comput Chem* **1989**, *10* (2), 209-220; (b) Stewart, J. J. P., Optimization of parameters for semiempirical methods II. Applications. *J Comput Chem* **1989**, *10* (2), 221-264.
63. Stewart, J. J. P. *MOPAC2009*, <http://OpenMOPAC.net>, Stewart Computational Chemistry: Colorado Springs, CO, USA, 2008.
64. *InChI*, <http://www.iupac.org/inchi/>.
65. Release of InChI Version 1.02 beta; Introducing InChIKey. *Chemistry International* 2007, pp 18-19.
66. Magoon, G. R.; Green, W. H. *MoleCoor*, <https://github.com/gmagoon/MoleCoor>, 2011.
67. (a) Baber, J. C.; Hodgkin, E. E., Automatic assignment of chemical connectivity to organic molecules in the Cambridge Structural Database. *Journal of Chemical Information and Computer Sciences* **1992**, *32* (5), 401-406; (b) Sayle, R., PDB: Cruft to Content. In *MUG 2001*, Santa Fe, NM, 2001.
68. Patchkovskii, S. *SYMMETRY*, <http://www.cobalt.chem.ucalgary.ca/ps/symmetry/>, 2003.
69. Allinger, N. L.; Lii, J.-H. *MM4(2008) and MM4(2003)*.
70. Nevins, N.; Chen, K.; Allinger, N. L., Molecular mechanics (MM4) calculations on alkenes. *J Comput Chem* **1996**, *17* (5-6), 669-694.
71. Nevins, N.; Lii, J.-H.; Allinger, N. L., Molecular mechanics (MM4) calculations on conjugated hydrocarbons. *J Comput Chem* **1996**, *17* (5-6), 695-729.
72. Sharma, S.; Harper, M. R.; Green, W. H. *CANTHERM v1.0*, <https://github.com/GreenGroup/CanTherm>, 2010.
73. Lay, T. H.; Bozzelli, J. W.; Dean, A. M.; Ritter, E. R., Hydrogen Atom Bond Increments for Calculation of Thermodynamic Properties of Hydrocarbon Radical Species. *The Journal of Physical Chemistry* **1995**, *99* (39), 14514-14527.
74. (a) Feng, Y.; Liu, L.; Wang, J.-T.; Zhao, S.-W.; Guo, Q.-X., Homolytic C–H and N–H Bond Dissociation Energies of Strained Organic Compounds. *The Journal of Organic Chemistry* **2004**, *69* (9), 3129-3138; (b) Tian, Z.; Fattahi, A.; Lis, L.; Kass, S. R., Cycloalkane and Cycloalkene C–H Bond Dissociation Energies. *J Am Chem Soc* **2006**, *128* (51), 17087-17092; (c) Agapito, F.; Nunes, P. M.; Costa Cabral, B. J.; Borges dos Santos, R. M.; Martinho Simões, J. A., Energetic Differences between the Five- and Six-Membered Ring Hydrocarbons: Strain

- Energies in the Parent and Radical Molecules. *The Journal of Organic Chemistry* **2008**, *73* (16), 6213-6223.
75. Osmont, A.; Catoire, L.; Gökalp, I., Physicochemical Properties and Thermochemistry of Propellanes. *Energ Fuel* **2008**, *22* (4), 2241-2257.
76. Lii, J.-H., personal communication. 2010.
77. Magoon, G. R.; Aguilera-Iparraguirre, J.; Green, W. H.; Lutz, J. J.; Piccuch, P.; Wong, H.-W.; Oluwole, O. O., Detailed chemical kinetic modeling of JP-10 (exo-tetrahydrodicyclopentadiene) high-temperature oxidation: Exploring the role of biradical species in initial decomposition steps. *Int J Chem Kinet* **2012**, *44* (3), 179-193.
78. Davidson, D. F.; Horning, D. C.; Herbon, J. T.; Hanson, R. K., Shock tube measurements of JP-10 ignition. *P Combust Inst* **2000**, *28*, 1687-1692.
79. (a) Sharma, S.; Raman, S.; Green, W. H., Intramolecular Hydrogen Migration in Alkylperoxy and Hydroperoxyalkylperoxy Radicals: Accurate Treatment of Hindered Rotors. *The Journal of Physical Chemistry A* **2010**, *114* (18), 5689-5701; (b) Speybroeck, V. V.; Vansteenkiste, P.; Neck, D. V.; Waroquier, M., Why does the uncoupled hindered rotor model work well for the thermodynamics of n-alkanes? *Chem Phys Lett* **2005**, *402* (4-6), 479-484.
80. Zheng, J.; Yu, T.; Papajak, E.; Alecu, I. M.; Mielke, S. L.; Truhlar, D. G., Practical methods for including torsional anharmonicity in thermochemical calculations on complex molecules: The internal-coordinate multi-structural approximation. *Phys Chem Chem Phys* **2011**, *13* (23), 10885-10907.
81. (a) Connolly, M. L., Computation of molecular volume. *J Am Chem Soc* **1985**, *107* (5), 1118-1124; (b) Pascual-Ahuir, J. L.; Silla, E., GEPOL: An improved description of molecular surfaces. I. Building the spherical surface set. *J Comput Chem* **1990**, *11* (9), 1047-1060; (c) Silla, E.; Tuñón, I.; Pascual-Ahuir, J. L., GEPOL: An improved description of molecular surfaces II. Computing the molecular area and volume. *J Comput Chem* **1991**, *12* (9), 1077-1088; (d) Pascual-Ahuir, J. L.; Silla, E.; Tuñón, I., GEPOL: An improved description of molecular surfaces. III. A new algorithm for the computation of a solvent-excluding surface. *J Comput Chem* **1994**, *15* (10), 1127-1138.
82. Tomasi, J.; Mennucci, B.; Cammi, R., Quantum Mechanical Continuum Solvation Models. *Chemical Reviews* **2005**, *105* (8), 2999-3094.

83. Curtiss, L. A.; Redfern, P. C.; Raghavachari, K., Gaussian-4 theory. *The Journal of Chemical Physics* **2007**, *126* (8), 084108-12.
84. Herbinet, O.; Sirjean, B.; Bounaceur, R.; Fournet, R.; Battin-Leclerc, F.; Scacchi, G.; Marquaire, P. M., Primary mechanism of the thermal decomposition of tricyclodecane. *J Phys Chem A* **2006**, *110* (39), 11298-11314.
85. Chung, H. S.; Chen, C. S. H.; Kremer, R. A.; Boulton, J. R.; Burdette, G. W., Recent Developments in High-Energy Density Liquid Hydrocarbon Fuels. *Energ Fuel* **1999**, *13* (3), 641-649.
86. (a) Bruno, T. J.; Huber, M. L.; Laesecke, A.; Lemmon, E. W.; Perkins, R. A., Thermochemical and thermophysical properties of JP-10. **2006**, *NISTIR 6640*; (b) Joppin, C. Cooling Performance of Storable Propellants for a micro rocket engine. Massachusetts Institute of Technology, Cambridge, MA, 2002.
87. (a) Mikolaitis, D. W.; Segal, C.; Chandy, A., Ignition delay for jet propellant 10/air and jet propellant 10/high-energy density fuel/air mixtures. *J Propul Power* **2003**, *19* (4), 601-606; (b) Colket, M. B.; Spadaccini, L. J., Scramjet fuels autoignition study. *J Propul Power* **2001**, *17* (2), 315-323; (c) Cooper, M.; Shepherd, J. E. *Thermal and Catalytic Cracking of JP-10 for Pulse Detonation Engine Applications. GALCIT Report FM 2002.002*; Explosion Dynamics Laboratory Report FM2002.002.; Thermal and Catalytic Cracking of JP-10 for Pulse Detonation Engine Applications, California Institute of Technology: Pasadena, 2002; (d) Green, R. J.; Anderson, S. L. In *Pyrolysis Chemistry of JP-10*, Proceedings of the 13th ONR Propulsion Meeting, Salt Lake City, UT, Salt Lake City, UT, 2000; (e) He, K. Y.; Androulakis, I. P.; Ierapetritou, M. G., Multi-element Flux Analysis for the Incorporation of Detailed Kinetic Mechanisms in Reactive Simulations. *Energ Fuel* **2010**, *24*, 309-317; (f) Li, S. C.; Varatharajan, B.; Williams, F. A., Chemistry of JP-10 ignition. *Aiaa J* **2001**, *39* (12), 2351-2356; (g) Magoon, G. R.; Green, W. H.; Oluwole, O. O.; Wong, H.-W.; Albo, S. E.; Lewis, D. K., Updating Our Understanding of JP-10 Decomposition Chemistry: A Detailed JP-10 Combustion Mechanism Constructed Using RMG – an Automatic Reaction Mechanism Generator. In *46th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit*, Nashville, TN, 2010; (h) Van Devener, B.; Anderson, S. L., Breakdown and combustion of JP-10 fuel catalyzed by nanoparticulate CeO₂ and Fe₂O₃. *Energ Fuel* **2006**, *20* (5), 1886-1894.

88. Auzmendi-Murua, I.; Hudzik, J.; Bozzelli, J. W., Chemical activation reactions of cyclic alkanes and ethers and tricyclodecane ring-opened diradicals with O₂: Thermochemistry, reaction paths, kinetics, and modeling. *Int J Chem Kinet* **2012**, *44* (4), 232-256.
89. <http://www.chem.leeds.ac.uk/Combustion/methane.htm>, Leeds methane combustion mechanism.
90. GRI-Mech 3.0. http://www.me.berkeley.edu/gri_mech/.
91. Sharma, S.; Green, W. H., Computed Rate Coefficients and Product Yields for c-C₅H₅ + CH₃ --> Products. *The Journal of Physical Chemistry A* **2009**, *113* (31), 8871-8882.
92. Magoon, G. R.; Green, W. H.; Oluwole, O. O.; Wong, H.-W.; Albo, S. E.; Lewis, D. K., Updating Our Understanding of JP-10 Decomposition Chemistry: A Detailed JP-10 Combustion Mechanism Constructed Using RMG – an Automatic Reaction Mechanism Generator. In *46th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit (AIAA 2010-6825)*, Nashville, TN, 2010.
93. (a) Baulch, D. L.; Bowman, C. T.; Cobos, C. J.; Cox, R. A.; Just, T.; Kerr, J. A.; Pilling, M. J.; Stocker, D.; Troe, J.; Tsang, W.; Walker, R. W.; Warnatz, J., Evaluated kinetic data for combustion modeling: Supplement II. *J Phys Chem Ref Data* **2005**, *34* (3), 757-1397; (b) Frank, P.; Bhaskaran, K. A.; Just, T., In *21st Symp. Int. Combustion*, 1988; p 885.
94. (a) Wagner, H. G.; Zabel, F., Thermal Decay of Ketene in Gas Phase. *Berich Bunsen Gesell* **1971**, *75* (2), 114-&; (b) Frank, P.; Bhaskaran, K. A.; Just, T., High-Temperature Reactions of Triplet Methylene and Ketene with Radicals. *J Phys Chem-Us* **1986**, *90* (10), 2226-2231.
95. Glass, G. P.; Kumaran, S. S.; Michael, J. V., Photolysis of ketene at 193 nm and the rate constant for H+HCCOat 297 K. *J Phys Chem A* **2000**, *104* (36), 8360-8367.
96. Bacskay, G. B.; Mackie, J. C., The pyrolysis of cyclopentadiene: quantum chemical and kinetic modelling studies of the acetylene plus propyne/allene decomposition channels. *Physical Chemistry Chemical Physics* **2001**, *3* (12), 2467-2473.
97. Roy, K. Kinetische Untersuchungen zur Hochtemperaturpyrolyse und-oxidation von Cyclopentadien und Cyclopentadienyl mit Hilfe der Stoßwellentechnik. Doctoral, Doctoral Thesis, 1999.

98. Moskaleva, L. V.; Lin, M. C., Unimolecular isomerization/decomposition of cyclopentadienyl and related bimolecular reverse process: ab initio MO/statistical theory study. *J Comput Chem* **2000**, *21* (6), 415-425.
99. (a) Caldwell, R. A., Intersystem crossing in organic photochemical intermediates. *Pure and Applied Chemistry* **1984**, *56* (9), 1167-1177; (b) Ni, T.; Caldwell, R. A.; Melton, L. A., The relaxed and spectroscopic energies of olefin triplets. *J Am Chem Soc* **1989**, *111* (2), 457-464; (c) Caldwell, R. A., Laser flash photolysis studies of intersystem crossing in biradicals and alkene triplets. In *Kinetics and spectroscopy of carbenes and biradicals*, Platz, M. S., Ed. Plenum Press: New York, NY, 1990; p 110; (d) Unett, D.; Caldwell, R., The triplet state of alkenes: Structure, dynamics, energetics and chemistry. *Research on Chemical Intermediates* **1995**, *21* (7), 665-709.
100. Sabbe, M. K.; Van Geem, K. M.; Reyniers, M.-F.; Marin, G. B., First principle-based simulation of ethane steam cracking. *AIChE Journal* **2011**, *57* (2), 482-496.
101. Chae, K.; Violi, A., Thermal Decomposition of Decalin: An Ab Initio Study. *The Journal of Organic Chemistry* **2007**, *72* (9), 3179-3185.
102. Fascella, S.; Cavallotti, C.; Rota, R.; Carrà, S., The Peculiar Kinetics of the Reaction between Acetylene and the Cyclopentadienyl Radical. *The Journal of Physical Chemistry A* **2005**, *109* (33), 7546-7557.
103. Kislov, V. V.; Mebel, A. M., Ab Initio G3-type/Statistical Theory Study of the Formation of Indene in Combustion Flames. I. Pathways Involving Benzene and Phenyl Radical†. *The Journal of Physical Chemistry A* **2007**, *111* (19), 3922-3931.
104. Kislov, V. V.; Mebel, A. M., An Ab Initio G3-Type/Statistical Theory Study of the Formation of Indene in Combustion Flames. II. The Pathways Originating from Reactions of Cyclic C5 Species Cyclopentadiene and Cyclopentadienyl Radicals. *The Journal of Physical Chemistry A* **2008**, *112* (4), 700-716.
105. Murakami, Y.; Saejung, T.; Ohashi, C.; Fujii, N., Investigation of a New Pathway Forming Naphthalene by the Recombination Reaction of Cyclopentadienyl Radicals. *Chemistry Letters* **2003**, *32* (12), 1112-1113.
106. Wang, D.; Violi, A.; Kim, D. H.; Mullholland, J. A., Formation of Naphthalene, Indene, and Benzene from Cyclopentadiene Pyrolysis: A DFT Study. *The Journal of Physical Chemistry A* **2006**, *110* (14), 4719-4725.

107. Li, G.; Rabitz, H.; Yelvington, P. E.; Oluwole, O. O.; Bacon, F.; Kolb, C. E.; Schoendorf, J., Global Sensitivity Analysis for Systems with Independent and/or Correlated Inputs. *The Journal of Physical Chemistry A* **2010**, *114* (19), 6022-6032.
108. Hong, Z.; Davidson, D. F.; Hanson, R. K., An improved H₂/O₂ mechanism based on recent shock tube/laser absorption measurements. *Combust Flame* **2011**, *158* (4), 633-644.
109. Jasper, A. W.; Klippenstein, S. J.; Harding, L. B.; Ruscic, B., Kinetics of the Reaction of Methyl Radical with Hydroxyl Radical and Methanol Decomposition†. *The Journal of Physical Chemistry A* **2007**, *111* (19), 3932-3950.
110. Robinson, R. K.; Lindstedt, R. P., On the chemical kinetics of cyclopentadiene oxidation. *Combust Flame* **2011**, *158* (4), 666-686.
111. Lewis, D. K.; Sarr, M.; Keil, M., Cyclopentene decomposition in shock waves. *The Journal of Physical Chemistry* **1974**, *78* (4), 436-439.
112. Mati, K.; Ristori, A.; Pengloan, G.; Dagaut, P., Oxidation of 1-methylnaphthalene at 1-13 atm: experimental study in a JSR and detailed chemical kinetic modeling. *Combustion Science and Technology* **2007**, *179* (7), 1261-1285.
113. Frank, P.; Herzler, J.; Just, T.; Wahl, C., High-temperature reactions of phenyl oxidation. *Symposium (International) on Combustion* **1994**, *25* (1), 833-840.
114. Emdee, J. L.; Brezinsky, K.; Glassman, I., A kinetic model for the oxidation of toluene near 1200 K. *The Journal of Physical Chemistry* **1992**, *96* (5), 2151-2161.
115. Chenoweth, K.; van Duin, A. C. T.; Dasgupta, S.; Goddard Iii, W. A., Initiation Mechanisms and Kinetics of Pyrolysis and Combustion of JP-10 Hydrocarbon Jet Fuel. *The Journal of Physical Chemistry A* **2009**, *113* (9), 1740-1746.
116. Hudzik, J. M.; Asatryan, R.; Bozzelli, J. W., Thermochemical Properties of exo-Tricyclo[5.2.1.0^{2,6}]decane (JP-10 Jet Fuel) and Derived Tricyclodecyl Radicals. *The Journal of Physical Chemistry A* **2010**, *114* (35), 9545-9553.
117. (a) Lutz, A. E.; Kee, R. J.; Miller, J. A. *SENKIN: A FORTRAN program for predicting homogeneous gas phase chemical kinetics with sensitivity analysis*, in Sandia National Lab report 87-8248., SENKIN: A FORTRAN program for predicting homogeneous gas phase chemical kinetics with sensitivity analysis, in Sandia National Lab report 87-8248.: 1987; (b) Kee, R. J.; Rupley, F. M.; Miller, J. A.; Coltrin, M. E.; Grcar, J. F.; Meeks, E.; Moffat, H. K.; Lutz, A. E.; Dixon-Lewis, G.; Smooke, M. D.; Warnatz, J.; Evans, G. H.; Larson, R. S.;

- Mitchell, R. E.; Petzold, L. R.; Reynolds, W. C.; Caracotsios, M.; Stewart, W. E.; Glarborg, P.; Wang, C.; McLellan, C. L.; Adigun, O.; Houf, W. G.; Chou, C. P.; Miller, S. F.; Ho, P.; Young, P. D.; Young, D. J.; Hodgson, D. W.; Petrova, M. V.; Puduppakkam, K. V. *CHEMKIN Release 4.1.1*, CHEMKIN Release 4.1.1, Reaction Design: San Diego, CA, 2007.
118. (a) Roos, B. O., The Complete Active Space Self-Consistent Field Method and its Applications in Electronic Structure Calculations. *Advances in Chemical Physics* **1987**, *69*, 399-445, and references therein; (b) Roos, B. O.; Taylor, P. R., A Complete Active Space Scf Method (Casscf) Using a Density-Matrix Formulated Super-Ci Approach. *Chem Phys* **1980**, *48* (2), 157-173; (c) Schmidt, M. W.; Gordon, M. S., The construction and interpretation of MCSCF wavefunctions. *Annu Rev Phys Chem* **1998**, *49*, 233-266.
119. (a) Nakano, H., Mcscf Reference Quasi-Degenerate Perturbation-Theory with Epstein-Nesbet Partitioning. *Chem Phys Lett* **1993**, *207* (4-6), 372-378; (b) Nakano, H., Quasi-Degenerate Perturbation-Theory with Multiconfigurational Self-Consistent-Field Reference Functions. *J Chem Phys* **1993**, *99* (10), 7983-7992.
120. (a) Piecuch, P.; Wloch, M., Renormalized coupled-cluster methods exploiting left eigenstates of the similarity-transformed Hamiltonian. *J Chem Phys* **2005**, *123*, Art. No. 224105 (22), [10 pages]; (b) Piecuch, P.; Wloch, M.; Gour, J. R.; Kinal, A., Single-reference, size-extensive, non-iterative coupled-cluster approaches to bond breaking and biradicals. *Chem Phys Lett* **2006**, *418* (4-6), 467-474; (c) Wloch, M.; Gour, J. R.; Piecuch, P., Extension of the renormalized coupled-cluster methods exploiting left eigenstates of the similarity-transformed Hamiltonian to open-shell systems: A benchmark study. *J Phys Chem A* **2007**, *111* (44), 11359-11382; (d) Wloch, M.; Lodriguito, M. D.; Piecuch, P.; Gour, J. R., Two new classes of non-iterative coupled-cluster methods derived from the method of moments of coupled-cluster equations (vol 104, pg 2149, 2006). *Mol Phys* **2006**, *104* (18), 2149-2172.
121. (a) Purvis, G. D.; Bartlett, R. J., A Full Coupled-Cluster Singles and Doubles Model - the Inclusion of Disconnected Triples. *J Chem Phys* **1982**, *76* (4), 1910-1918; (b) Cramer, C. J.; Kinal, A.; Wloch, M.; Piecuch, P.; Gagliardi, L., Theoretical Characterization of End-On and Side-On Peroxide Coordination in Ligated Cu₂O₂ Models. *The Journal of Physical Chemistry A* **2006**, *110* (40), 11557-11568; (c) Cramer, C. J.; Wloch, M.; Piecuch, P.; Puzzarini, C.; Gagliardi, L., Theoretical models on the Cu₂O₂ torture track: Mechanistic implications for oxytyrosinase and small-molecule analogues. *J Phys Chem A* **2006**, *110* (5), 1991-2004; (d) Ge,

Y. B.; Gordon, M. S.; Piecuch, P., Breaking bonds with the left eigenstate completely renormalized coupled-cluster method. *J Chem Phys* **2007**, *127*, Art. No. 174106 (17), [6 pages]; (e) Ge, Y. B.; Gordon, M. S.; Piecuch, P.; Wloch, M.; Gour, J. R., Breaking Bonds of Open-Shell Species with the Restricted Open-Shell Size Extensive Left Eigenstate Completely Renormalized Coupled-Cluster Method. *J Phys Chem A* **2008**, *112* (46), 11873-11884; (f) Kinal, A.; Piecuch, P., Computational investigation of the conrotatory and disrotatory isomerization channels of bicyclo[1.1.0]butane to buta-1,3-diene: A completely renormalized coupled-cluster study. *J Phys Chem A* **2007**, *111* (4), 734-742; (g) Song, Y. Z.; Kinal, A.; Caridade, P. J. S. B.; Varandas, A. J. C.; Piecuch, P., A comparison of single-reference coupled-cluster and multi-reference configuration interaction methods for representative cuts of the H₂S((1)A') potential energy surface. *J Mol Struct-Theochem* **2008**, *859* (1-3), 22-29; (h) Piecuch, P.; Wloch, M.; Varandas, A. J. C., Application of renormalized coupled-cluster methods to potential function of water. *Theor Chem Acc* **2008**, *120* (1-3), 59-78.

122. (a) Becke, A. D., DENSITY-FUNCTIONAL THERMOCHEMISTRY .3. THE ROLE OF EXACT EXCHANGE. *J Chem Phys* **1993**, *98* (7), 5648-5652; (b) Lee, C. T.; Yang, W. T.; Parr, R. G., Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density. *Phys Rev B* **1988**, *37* (2), 785-789.

123. Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A., Self-Consistent Molecular-Orbital Methods .20. Basis Set for Correlated Wave-Functions. *J Chem Phys* **1980**, *72* (1), 650-654.

124. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision A.02*, Gaussian 09, Revision A.02, Gaussian, Inc.: Wallingford CT, 2009.

125. (a) Gordon, M. S.; Schmidt, M. W., Advances in electronic structure theory: GAMESS a decade later. In *Theory and Applications of Computational Chemistry, the first forty years*, Dykstra, C. E.; Frenking, G.; Kim, K. S.; Scuseria, G. E., Eds. Elsevier: Amsterdam, 2005; pp 1167-1189; (b) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., General Atomic and Molecular Electronic-Structure System. *J Comput Chem* **1993**, *14* (11), 1347-1363.
126. Piecuch, P.; Kucharski, S. A.; Kowalski, K.; Musial, M., Efficient computer implementation of the renormalized coupled-cluster methods: The R-CCSD[T], R-CCSD(T), CR-CCSD[T], and CR-CCSD(T) approaches. *Comput Phys Commun* **2002**, *149* (2), 71-96.
127. Bode, B. M.; Gordon, M. S., MacMolPlt: A graphical user interface for GAMESS. *J Mol Graph Model* **1998**, *16* (3), 133-138.
128. Malick, D. K.; Petersson, G. A.; Montgomery, J. A., Transition states for chemical reactions I. Geometry and classical barrier height. *J Chem Phys* **1998**, *108* (14), 5704-5713.
129. Zheng, J.; Gour, J. R.; Lutz, J. J.; Włoch, M.; Piecuch, P.; Truhlar, D. G., A Comparative Assessment of the Perturbative and Renormalized Coupled Cluster Theories with a Non-iterative Treatment of Triple Excitations for Thermochemical Kinetics, Including a Study of Basis Set and Core Correlation Effects. *J Chem Phys* **2008**, *128*, 044108-1 - 044108-7.
130. Mousavipour, S. H.; Homayoon, Z., A theoretical study on the kinetics of disproportionation versus association reaction of CH₃+C₂H₅. *J Phys Chem A* **2003**, *107* (41), 8566-8574.
131. Zhu, R. S.; Xu, Z. F.; Lin, M. C., Ab initio studies of alkyl radical reactions: Combination and disproportionation reactions of CH₃ with C₂H₅, and the decomposition of chemically activated C₃H₈. *J Chem Phys* **2004**, *120* (14), 6566-6573.
132. Sirjean, B.; Glaude, P. A.; Ruiz-Lopez, M. F.; Fournet, R., Detailed kinetic study of the ring opening of cycloalkanes by CBS-QB3 calculations. *J Phys Chem A* **2006**, *110* (46), 12693-12704.
133. Kiefer, J. H.; Gupte, K. S.; Harding, L. B.; Klippenstein, S. J., Shock Tube and Theory Investigation of Cyclohexane and 1-Hexene Decomposition. *J Phys Chem A* **2009**, *113* (48), 13570-13583.

134. Izgorodina, E. I.; Coote, M. L.; Radom, L., Trends in R–X Bond Dissociation Energies (R = Me, Et, i-Pr, t-Bu; X = H, CH₃, OCH₃, OH, F): A Surprising Shortcoming of Density Functional Theory. *The Journal of Physical Chemistry A* **2005**, *109* (33), 7558-7566.
135. Ventura, E.; Dallos, M.; Lischka, H., Revisiting the stationary points on the potential energy surface of tetramethylene at the MR-AQCC level using analytic gradients. *J Chem Phys* **2003**, *118* (24), 10963-10972.
136. Szalay, P. G.; Bartlett, R. J., Multireference Averaged Quadratic Coupled-Cluster Method - a Size-Extensive Modification of Multireference Ci. *Chem Phys Lett* **1993**, *214* (5), 481-488.
137. Lee, T. J.; Taylor, P. R., A diagnostic for determining the quality of single-reference electron correlation methods. *Int J Quantum Chem* **1989**, *36* (S23), 199-207.
138. Dames, E.; Krylov, A.; Wang, H., Theory and Kinetic Modeling of Initiation Reactions for Cyclohexane and Several of its Mono-Alkylated Derivatives. . In *7th US National Technical Meeting of the Combustion Institute*, Atlanta, GA, 2011.
139. DeTar, D. F., Theoretical ab Initio Calculation of Entropy, Heat Capacity, and Heat Content. *The Journal of Physical Chemistry A* **1998**, *102* (26), 5128-5141.
140. (a) Berne, B. J.; Thirumalai, D., ON THE SIMULATION OF QUANTUM-SYSTEMS - PATH INTEGRAL METHODS. *Annu Rev Phys Chem* **1986**, *37*, 401-424; (b) Lynch, V. A.; Mielke, S. L.; Truhlar, D. G., Accurate vibrational-rotational partition functions and standard-state free energy values for H₂O₂ from Monte Carlo path-integral calculations. *The Journal of Chemical Physics* **2004**, *121* (11), 5148-5162; (c) Chandler, D.; Wolynes, P. G., Exploiting the isomorphism between quantum theory and classical statistical mechanics of polyatomic fluids. *The Journal of Chemical Physics* **1981**, *74* (7), 4078-4095; (d) Parrinello, M.; Rahman, A., Study of an F center in molten KCl. *The Journal of Chemical Physics* **1984**, *80* (2), 860-867.
141. (a) Tafipolsky, M.; Schmid, R., Calculation of rotational partition functions by an efficient Monte Carlo importance sampling technique. *J Comput Chem* **2005**, *26* (15), 1579-1591; (b) Ellingson, B. A.; Lynch, V. A.; Mielke, S. L.; Truhlar, D. G., Statistical thermodynamics of bond torsional modes: Tests of separable, almost-separable, and improved Pitzer--Gwinn approximations. *The Journal of Chemical Physics* **2006**, *125* (8), 084305-17; (c) Vereecken, L.; Peeters, J., The 1,5-H-shift in 1-butoxy: A case study in the rigorous implementation of transition state theory for a multimer system. *The Journal of Chemical*

- Physics* **2003**, *119* (10), 5159-5170; (d) Vansteenkiste, P.; Neck, D. V.; Speybroeck, V. V.; Waroquier, M., An extended hindered-rotor model with incorporation of Coriolis and vibrational-rotational coupling for calculating partition functions and derived quantities. *The Journal of Chemical Physics* **2006**, *124* (4), 044314; (e) Vansteenkiste, P.; Van Speybroeck, V.; Pauwels, E.; Waroquier, M., How should we calculate multi-dimensional potential energy surfaces for an accurate reproduction of partition functions? *Chem Phys* **2005**, *314* (1-3), 109-117.
142. Tresadern, G.; Agrafiotis, D. K., Conformational Sampling with Stochastic Proximity Embedding and Self-Organizing Superimposition: Establishing Reasonable Parameters for Their Practical Use. *Journal of Chemical Information and Modeling* **2009**, *49* (12), 2786-2800.
143. Saunders, M., Stochastic exploration of molecular mechanics energy surfaces. Hunting for the global minimum. *J Am Chem Soc* **1987**, *109* (10), 3150-3152.
144. Kabsch, W., A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A* **1976**, *32* (5), 922-923.
145. Weng, Z. F.; Motherwell, W. D. S.; Cole, J. M., Tormat: a program for the automated structural alignment of molecular conformations. *Journal of Applied Crystallography* **2008**, *41* (5), 955-957.
146. Munkres, J., Algorithms for the Assignment and Transportation Problems. *Journal of the Society for Industrial and Applied Mathematics* **1957**, *5* (1), 32-38.
147. Bond, A., Automated least-squares molecular overlay without a priori atomic correspondence or connectivity information. *Journal of Applied Crystallography* **2010**, *43* (1), 53-57.
148. Collins, A.; Cooper, R. I.; Watkin, D. J., Structure matching: measures of similarity and pseudosymmetry. *Journal of Applied Crystallography* **2006**, *39* (6), 842-849.
149. Sato, K.; Goto, M.; Yano, J.; Honda, K.; Kodali, D. R.; Small, D. M., Atomic resolution structure analysis of β' polymorph crystal of a triacylglycerol: 1,2-dipalmitoyl-3-myristoyl-sn-glycerol. *Journal of Lipid Research* **2001**, *42* (3), 338-345.
150. Peters, B.; Heyden, A.; Bell, A. T.; Chakraborty, A., A growing string method for determining transition states: Comparison to the nudged elastic band and string methods. *The Journal of Chemical Physics* **2004**, *120* (17), 7877-7886.

151. Peng, C. Y.; Schlegel, H. B., Combining Synchronous Transit and Quasi-Newton Methods to Find Transition-States. *Isr. J. Chem.* **1993**, *33* (4), 449-454.
152. van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A., ReaxFF: A Reactive Force Field for Hydrocarbons. *The Journal of Physical Chemistry A* **2001**, *105* (41), 9396-9409.
153. (a) Gilbert, K. E.; Gajewski, J. J. *MMX and PCModel*, Serena Software, Ltd.: Bloomington, IN; (b) Eksterowicz, J. E.; Houk, K. N., Transition-state modeling with empirical force fields. *Chemical Reviews* **1993**, *93* (7), 2439-2461.
154. Sánchez-Ruiz, X.; Jaime, C.; Marco-Contelles, J., The first analysis of the steric control of the 6-exo free-radical cyclization of acyclic sugar derivatives—MM3 transition state modelling. *Journal of Molecular Structure* **1998**, *471* (1–3), 209-218.
155. (a) Barker, J. R.; Ortiz, N. F.; Preses, J. M.; Lohr, L. L.; Maranzana, A.; Stimac, P. J.; Nguyen, T. L.; Dhilip Kumar, T. J., *MultiWell Program Suite User Manual (MultiWell-2011.2d)*. 2011; (b) Meyer, R., Trigonometric Interpolation Method for One-Dimensional Quantum-Mechanical Problems. *The Journal of Chemical Physics* **1970**, *52* (4), 2053-2059.
156. Godunov, I. A.; Alekseev, V. N.; Abramnikov, A. V.; Tatevskii, V. M., Bicyclopropyl : potential function for the internal rotation of the molecule and thermodynamic properties. *Anglais* **1991**, *65* (12), 3254-3259.
157. Pitzer, K. S., Energy Levels and Thermodynamic Functions for Molecules with Internal Rotation: II. Unsymmetrical Tops Attached to a Rigid Frame. *The Journal of Chemical Physics* **1946**, *14* (4), 239-243.