

**Chemistry *in silico*:  
Organic Reactions, Ionic Liquids, Thermochemistry**

by

Billy Wayne McCann

A dissertation submitted to the Graduate Faculty of  
Auburn University  
in partial fulfillment of the  
requirements for the Degree of  
Doctor of Philosophy

Auburn, Alabama  
May 5, 2013

Keywords: chemistry, computational, organic,  
ionic liquids, hydantoins, tautomerization

Copyright 2013 by Billy Wayne McCann

Approved by

Orlando Acevedo, Chair, Associate Professor of Chemistry and Biochemistry  
Vincent Ortiz, Ruth W. Molette Professor of Department of Chemistry and Biochemistry  
Michael L. McKee, Professor of Department of Chemistry and Biochemistry  
Edward J. Parish, Professor of Department of Chemistry and Biochemistry  
Stuart Loch, Professor of Department of Physics

## Abstract

Computational chemistry is a collection of methods that seek to interpret and predict chemical phenomena. In this work, three distinct areas of chemistry are investigated: (1) the reactive pathways of UVA irradiated *N*-chlorohydantoins; (2) alternative methods for calculating long-range electrostatics in room temperature ionic liquids; (3) benchmarking implicit solvent models and cavity sizes to experimentally determined free energies of protomerization in solution phase.

Chapter 1 presents a brief introduction to these three subjects. Chapter 2 presents the methodologies used in this research. Chapter 3 reports the results of a joint experimental and computational investigation into the UVA stabilities of the *N*-chlorohydantoins 1-chloro-3,5,5-trimethylhydantoin and 3-chloro-1,5,5-trimethylhydantoin. Chapter 4 details the exhaustive testing of pair-wise Ewald alternatives on a wide range of ionic liquids reveals that shifted forces, shifted potentials, and Coulombic interactions all damped by  $0.2 \text{ \AA}^{-1}$  with an interaction cutoff of  $15 \text{ \AA}$  consistently accurately reproduce Ewald sum electrostatics. Additionally, a new methodology is presented, SF3, whose accuracy is on par with the above-mentioned methods, yet at a much-reduced computational cost. Finally, chapter 5 details the findings of the tautomerization benchmarking work.

## Acknowledgments

“No man is an island.” ~ John Donne, *Meditation XVII* To all those who have in any way helped me along my path, thank you. This mountain which I have climbed would have been insurmountable without the support my family, my wife Rachel, my kind advisor Orlando Acevedo, and my coworkers Samba, Caley, Kira, Symon, Nicole, and Bin. I am especially grateful to Dr. Acevedo and Dr. Ortiz, who allowed me to continue my Ph.D. pursuit even while teetering on the edge of indecision. I’d like to thank Ann Marie DelSignore, my counselor at the Auburn Student Counseling Center, for giving me perspective at critical moments. Most of all, my greatest gratitude to my loving wife for helping me through doubt, trials, and tribulations.

ד הַאֱלֹהִים הֵיאֵ: אִין־טוֹב בְּאָדָם שִׂיאֲכַל וְשָׁתָה וְהִרְאָה אֶת־נַפְשׁוֹ טוֹב בַּעֲמָלוֹ גַּם־זֶה רְאִיתִי אֲנִי כִּי מִן־

~ Solomon

## Table of Contents

Abstract.....	ii
Acknowledgments.....	iii
List of Tables .....	ix
List of Figures.....	xi
List of Schemes.....	xiv
List of Abbreviations .....	xvi
Chapter 1 Introduction.....	1
1.1 Inter- and Intra-molecular Mechanisms for Chlorine Rearrangements in Trimethyl-substituted <i>N</i> -chlorohydantoins .....	1
1.2 Pairwise Alternatives to Ewald Summation for Calculating Long-Range Electrostatics in Ionic Liquids.....	1
1.3 Free Energy Differences of Protomers in Solution .....	4
1.4 References .....	8
Chapter 2 Methods.....	13
2.1 <i>ab initio</i> Quantum Chemistry.....	13
2.1.1 The Schrödinger Equation.....	13
2.1.2 The Time-Independent Schrödinger Equation .....	15
2.1.3 The Variational Principle.....	18
2.1.4 The Born-Oppenheimer Approximation .....	18

2.1.5 Finding the Electronic Wavefunction.....	19
2.2 Composite Thermochemical Model Chemistries .....	35
2.3 Semiempirical Quantum Mechanics .....	36
2.3.1 NDDO.....	37
2.4 Density Functional Theory (DFT).....	38
2.4.1 Kohn-Sham.....	38
2.4.2 Local Density and Local Spin Density Approximations .....	39
2.4.3 Generalized and meta-Generalized Gradient Approximation .....	40
2.4.4 Hybrid Functionals .....	40
2.5 Molecular Mechanics.....	41
2.5.1 OPLS-AA .....	42
2.6 Mixed Quantum and Molecular Mechanics Methods (QM/MM).....	44
2.7 Thermodynamic quantities .....	46
2.7.1 Molecular Thermodynamics.....	47
2.7.2 Many Body Thermodynamics .....	49
2.7.3 Free Energy Perturbation.....	52
2.8 The Ewald Summation.....	55
2.9 References .....	58
Chapter 3 Inter- and Intra-molecular Mechanisms for Chlorine Rearrangements in Trimethyl-substituted <i>N</i> -chlorohydantoins .....	65

3.1 Abstract.....	65
3.2 Computational Methods.....	66
3.3 Results and Discussion .....	67
3.3.1 Heterolytic Dissociation .....	67
3.3.2 Homolytic Dissociation under UVA Irradiation .....	69
3.3.3 Mechanistic Considerations.....	75
3.3.4 Final Product Ratios .....	89
3.4 Conclusions .....	90
3.5 References .....	91
Chapter 4 Pairwise Alternatives to Ewald Summation for Calculating Long-Range Electrostatics in Ionic Liquids .....	92
4.1 Abstract.....	92
4.2 Theory .....	93
4.3 Computational Methods.....	97
4.4 Results and Discussion .....	98
4.5 Additional smoothing.....	102
4.6 Timings .....	103
4.7 Conclusions .....	104
4.8 References .....	106
Chapter 5 Solution Phase Tautomerization Thermodynamics .....	108

5.1 Abstract.....	108
5.2 Benchmarking Continuum Solvation Models and Cavities for the Accurate Calculation of Condensed Phase Free Energy of Tautomerization.....	109
5.2.1 Introduction .....	109
5.2.2 Methods .....	110
5.2.3 Results and Discussion .....	112
5.3 Application to an Enolization Reaction: 2-nitrocyclohexanone .....	128
5.3.1 Methods .....	130
5.3.2 Results and Discussion .....	134
5.4 Conclusions .....	143
5.5 References .....	144

## List of Tables

Table 1.1 Average bond enthalpies ( $\text{kcal}\cdot\text{mol}^{-1}$ ) for bonds distinguishing keto and enol tautomers.....	6
Table 3.1 Experimental Final Product Ratios Obtained under UVA Light Exposure.....	75
Table 4.1 Statistical analysis of the quality of the configurational energy differences for a given electrostatic method applied across all ionic liquid simulations compared to the reference Ewald sum. <sup>a</sup> .....	100
Table 5.1 Radii of atomic spheres for each cavity type used in the construction of the solute cavities. Units given in Å.....	111
Table 5.2 Unsigned errors in calculated $\Delta G_{gas}$ of Tautomerization. Negative $\Delta G$ values indicate preference of the enol form. Values given in $\text{kcal}\cdot\text{mol}^{-1}$ .....	113
Table 5.3 Unsigned errors of basis sets using the M06 density functional. Values given in $\text{kcal}\cdot\text{mol}^{-1}$ .....	114
Table 5.4. Statistical analysis of unsigned errors across all molecules and all solvents. Values given in $\text{kcal}\cdot\text{mol}^{-1}$ .....	116
Table 5.5. Statistical analysis of errors across all molecules in water. Values given in $\text{kcal}\cdot\text{mol}^{-1}$ .....	117
Table 5.6. Statistical analysis of errors across all molecules in nonaqueous solvents. Values given in $\text{kcal}\cdot\text{mol}^{-1}$ .....	118
Table 5.7. Statistical analysis of unsigned errors for aqueous monocarbonyls. Values given in $\text{kcal}\cdot\text{mol}^{-1}$ .....	119
Table 5.8. Statistical analysis of unsigned errors for $\beta$ -dicarbonyls in all solvents. Values given in $\text{kcal}\cdot\text{mol}^{-1}$ .....	120
Table 5.9. Statistical analysis of unsigned errors of $\beta$ -dicarbonyls in water. Values given in $\text{kcal}\cdot\text{mol}^{-1}$ .....	121
Table 5.10. Statistical analysis of unsigned errors for $\beta$ -dicarbonyls in nonaqueous solvent. Values given in $\text{kcal}\cdot\text{mol}^{-1}$ .....	122

Table 5.11. Statistical analysis of unsigned errors for heterocycles in all solvents. Values given in kcal·mol <sup>-1</sup> .....	123
Table 5.12. Statistical analysis of unsigned errors for heterocycles in water. Values given in kcal·mol <sup>-1</sup> .....	124
Table 5.13. Statistical analysis of unsigned errors for heterocycles in nonaqueous solvents. Values given in kcal·mol <sup>-1</sup> .....	125
Table 5.14. Summary of statistical analysis of unsigned errors for the most accurate method, solvation model and cavity type. Values given in kcal·mol <sup>-1</sup> .....	126
Table 5.15 Experimental Free Energies of Activation $\Delta G^\ddagger$ (kcal·mol <sup>-1</sup> at 298.15 K) for the Base-Catalyzed Enolization of 2-NCH in Various Solvents. All values given in kcal·mol <sup>-1</sup> .....	129
Table 5.16 Experimental and calculated $\Delta G^\ddagger$ for the stepwise TS1 of enolization of 2NCH. Values calculated using the PAULING/CPCM/M06/6-31+G(d,p) method and given in kcal·mol <sup>-1</sup> .....	135
Table 5.17 Experimental and calculated $\Delta G^\ddagger$ for the stepwise TS1 of enolization of 2NCH. Values calculated using the COULOMB/SMD/M06/6-31+G(d,p) method and given in kcal·mol <sup>-1</sup> .....	135

## List of Figures

Figure 1.1 Structures of 1-chloro-3,5,5-trimethylhydantoin (1) and 3-chloro-1,5,5-trimethylhydantoin (2). .....	2
Figure 1.2 Ionic liquid forming cations. R = M (methyl), E (ethyl), B (butyl), H (hexyl), and O (octyl). .....	4
Figure 1.3 Illustration of constitutional isomers ethanol and dimethyl ether. ....	4
Figure 1.4 Keto-enol tautomerization illustrating the differing hydrogen connectivity....	5
Figure 1.5 Phenol keto-enol tautomerization. The keto form loses aromaticity (colored blue), a highly stabilizing factor. ....	6
Figure 1.6 Tautomerization of the $\beta$ -diketone acetylacetone. The $\alpha$ -carbon and conjugated $\pi$ -system is colored red and blue, respectively. The hydrogen bond is indicated with a dashed line. ....	7
Figure 2.1 Illustration of the separation of spatial orbitals into separate $\alpha$ and $\beta$ spin orbitals for open shell systems. ....	20
Figure 2.2 Illustration of the hydrogen $1s$ atomic orbital expressed as a STO and a variationally optimized single GTO. ....	24
Figure 2.3 Illustration of the Lennard-Jones potential. ....	43
Figure 2.4 Partitioning scheme used in the QM/MM methodology. ....	45
Figure 2.5 Periodic Boundary Conditions. If a molecule exits one side of the cell, it reappears on the other side. ....	52
Figure 2.6 Double-wide sampling. The system begins in state $M$ and is perturbed to states $X$ and $Y$ simultaneously. ....	54
Figure 2.7. The Ewald summation in diagrammatic form. Point charges are combined with charge screening Gaussian charge distributions and their reciprocals. ....	56
Figure 3.1 Structures of 1-chloro-3,5,5-trimethylhydantoin (1) and 3-chloro-1,5,5-trimethylhydantoin (2). ....	65

Figure 3.2 Oxidative chlorine loss in water (above) and for the solid state (below) over time for 1-chloro-3,5,5-trimethylhydantoin (1) and 3-chloro-1,5,5-trimethylhydantoin (2). .....	68
Figure 3.3 UVA stability of solid-state 1-chloro-3,5,5-trimethylhydantoin (1) and 3-chloro-1,5,5-trimethyl hydantoin (2). .....	70
Figure 3.4 <sup>1</sup> H NMR spectrum for the photodecomposition products of compound 1 in CDCl <sub>3</sub> . .....	73
Figure 3.5 <sup>1</sup> H NMR spectrum for the photodecomposition products of compound 2 in CDCl <sub>3</sub> . .....	74
Figure 3.6 UB3LYP/6-311++G(2d,p) calculated transition structures in the <b>1</b> to <b>1a</b> intramolecular reaction pathway for (A) the hydrogen atom transfer from 3-methyl to the 2-oxygen, and (B) the subsequent tautomerization between the 2-oxygen and 1-nitrogen. Optimized distances given in angstroms. ....	77
Figure 3.7 UB3LYP/6-311++G(2d,p) calculated transition structure for the intramolecular hydrogen atom transfer from 1· to 5-methylene-1,5-dimethylhydantoin. Optimized distances given in angstroms. ....	78
Figure 3.8 Calculated UB3LYP/6-311++G(2d,p) transition structures for the HATRs of (A) 5-methyl in 2-O4· to oxygen radical, (B) the tautomerization of 2-hydroxyl-5-methylene radical, and (C) 1-methyl to the 5-methylene radical group. Distances given in angstroms. ....	81
Figure 3.9 Calculated UB3LYP/6-311++G(2d,p) intramolecular transition structures for (A) the hydrogen transfer from the 1-methyl to the 2-oxygen and (b) the tautomerization between the 2-oxygen and 3-nitrogen. Distances given in angstroms. ....	82
Figure 3.10 Calculated UB3LYP/6-311++G(2d,p) intermolecular transition structures for the HATR between N1-radical-3-chloromethyl-5,5-dimethylhydantoin and either the starting material 1 or 1b. Optimized distances given in angstroms. ....	85
Figure 4.1 Least squares regressions of the configurational energy differences for the [EPyr][BF <sub>4</sub> ] system using a 15 Å cutoff distance. ....	99
Figure 4.2 Graphical representation of the correlation coefficient ( $R^2$ ) of the configurational energy differences for a given electrostatic method applied across all ionic liquid simulations at three cutoff distances -- 9, 12, and 15 Å. Correlation coefficients closest to 1.0 indicate $\Delta E$ values most similar to the reference Ewald sum. ....	101
Figure 4.3 Timings for ionic liquid simulations relative to the full Ewald sums. ....	104

Figure 5.1 Geometries for the first transition structures in the enolization of 2NCH in dichloromethane as calculated by (a) COULOMB/SMD and (b) PAULING/CPCM. The orientation of the methyl groups of (a) causes significant steric hindrance. ....	137
Figure 5.2 Potential energy surface for TS1 of the stepwise pyridine catalyzed deprotonation of 2NCH in acetonitrile. ....	139
Figure 5.3 Potential energy surface for TS1 of the stepwise pyridine catalyzed deprotonation of 2NCH in dichloromethane.....	140
Figure 5.4 Potential energy surface for TS1 of the stepwise pyridine catalyzed deprotonation of 2NCH in carbontetrachloride. ....	141
Figure 5.5 Potential energy surface for the concerted enolization of 2NCH catalyzed by pyridine in cyclohexane. ....	142

## List of Schemes

Scheme 3.1 Intramolecular photorearrangement of acyclic <i>N</i> -halamines. ....	69
Scheme 3.2 Three Potential Reaction Products for the Photodecomposition under UVA Irradiation of 1-chloro-3,5,5-trimethylhydantoin (1) and 3-chloro-1,5,5-trimethylhydantoin (2) .....	71
Scheme 3.3 Proposed Intramolecular Hydrogen Atom Transfer Mechanism for the Formation of 1a from 1. $\Delta G^\ddagger$ (kcal/mol) Values Computed using UB3LYP/6-311++G(2d,p). ....	76
Scheme 3.4 Alternative Pathway for the Proposed Intramolecular Mechanism Given in Scheme 5 that Leads to the Formation of 1c. $\Delta G^\ddagger$ (kcal/mol) Calculated using UB3LYP/6-311++G(2d,p).....	77
Scheme 3.5 Three Resonance Structures for 2 $\cdot$ .....	79
Scheme 3.6 Proposed Intramolecular Hydrogen Atom Transfer Mechanism for the Formation of 2a and 2c from 2-O4 $\cdot$ . $\Delta G^\ddagger$ Values (kcal/mol) Computed using UB3LYP/6-311++G(2d,p).....	80
Scheme 3.7 Proposed Intramolecular Hydrogen Atom Transfer Mechanism for the Formation of 2a from 2-O2 $\cdot$ . $\Delta G^\ddagger$ Values (kcal/mol) Computed using UB3LYP/6-311++G(2d,p). ....	82
Scheme 3.8 Intermolecular mechanism for 1 yielding products 1a and 1b. $\Delta G^\ddagger$ values (kcal/mol) computed using UB3LYP/6-311++G(2d,p).....	84
Scheme 3.9 Intermolecular Mechanism of 1 Yielding Products 1b and 1-chloro-3-methylene-5,5-dimethylhydantoin. $\Delta G^\ddagger$ values (kcal/mol) computed using UB3LYP/6-311++G(2d,p).....	86
Scheme 3.10 Intermolecular mechanism of 2 yielding products 2a and 2b. $\Delta G^\ddagger$ values (kcal/mol) computed using UB3LYP/6-311++G(2d,p).....	88
Scheme 3.11 Intermolecular hydrogen transfer between 2 and 2 $\cdot$ at the 5-methyl position. $\Delta G^\ddagger$ value (kcal/mol) computed using UB3LYP/6-311++G(2d,p). ....	89

Scheme 5.1. Proposed shift in the enolization mechanism of 2NCH from (a) concerted in less polar solvents to (b) stepwise in more polar solvents. ....	129
Scheme 5.2 Stationary points characterizing the stepwise enolization of 2NCH. ....	131
Scheme 5.3 Reaction coordinates (bold) used in creating the 2D free energy map for the stepwise enolization of 2NCH. ....	134
Scheme 5.4 Reaction coordinates (bold) used in creating the 2D free energy map for the concerted enolization of 2NCH. ....	134

## List of Abbreviations

[HMIM][Tf <sub>2</sub> N]	1-n-hexyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide
[RMIM]	1-alkyl-3-methylimidazolium
[RPYR]	N-alkylpyridinium
Å	Angstrom
B	Butyl
BOSS	Biochemical and Organic Simulation Software
CBS	Complete Basis Set
CC	Coupled Cluster
CCS	Coupled Cluster Singles
CCSD	Coupled Cluster Singles Doubles
CCSD(T)	Coupled Cluster Singles Doubles with Perturbative Triples
CCSDT	Coupled Cluster Singles Doubles Triples
CHARMM	Chemistry at Harvard Molecular Mechanics
CHARMM-SF	Chemistry at Harvard Molecular Mechanics Shifted Force
CI	Configuration Interaction
CM1	Charge Model 1
CM3	Charge Model 3
DC	Damped Coulombic Potential
DFT	Density Functional Theory

DSF	Damped Shifted Force
DSP	Damped Shifted Potential
E	Ethyl
FEP	Free Energy Perturbation
GGA	Generalized Gradient Approximation
GTO	Gaussian Type Orbital
H	Hexyl
HATR	Hydrogen Atom Transfer Reaction
HF	Hartree-Fock
kcal	Kilocalories
LCAO-MO	Linear Combination of Atomic Orbitals Molecular Orbitals
LDA	Local Density Approximation
LSDA	Local Spin Density Approximation
M	Methyl
MC	Monte Carlo
MD	Molecular Dynamics
meta-GGA	meta Generalized Gradient Approximation
mol	Mole
MP2	Second Order Møller-Plesset Perturbation Theory
NDDO	Neglect of Diatomic Differential Overlap
NMR	Nuclear Magnetic Resonance
NPT	Isobaric-Isothermal Ensemble
O	Octyl
OPLS-AA	Optimized Parameters for Liquid Simulations All Atoms
P3M	Particle-Particle Particle-Mesh Ewald

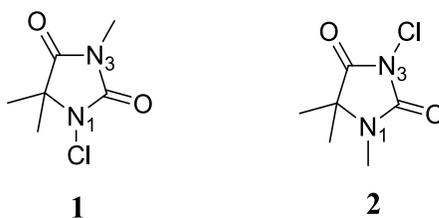
PME	Particle Mesh Ewald
QM	Quantum Mechanical
QM/MM	Mixed Quantum Mechanics and Molecular Mechanics
R	Alkyl
SF	Shifted Force
SF3	Shifted Force 3rd Derivative
SFG	Shifted Force Gradient
SP	Shifted Potential
STO	Slater Type Orbital
UVA	Ultraviolet Radiation A

## Chapter 1 Introduction

### 1.1 Inter- and Intra-molecular Mechanisms for Chlorine Rearrangements in Trimethyl-substituted *N*-chlorohydantoins

*N*-halamine monomers and polymers have been shown to be excellent antimicrobials with rates of inactivation of bacteria, viruses, fungi, and protozoa from seconds at low concentrations in aqueous solution to several minutes for polymeric coatings containing the materials.<sup>1-3</sup> They also possess the advantages of being non-corrosive, non-toxic to humans and animals, and the capability of regeneration *in situ* after loss of oxidative halogen by exposure to dilute solutions of aqueous free chlorine, e.g., household bleach, or free bromine. One such polymer, an *N*-halogenated poly-styrenehydantoin, is currently being employed for potable water disinfection in developing nations.<sup>4</sup> There are numerous potential applications for the materials in the medical fields.<sup>5-8</sup> However, some *N*-halamine derivatives employed in antimicrobial coatings are subject to probable free-radical halogen migration processes in the presence of UVA irradiation which can lead to decomposition causing loss of the regeneration capability.<sup>7</sup> For this reason this thesis combined an experimental and computational effort designed to elucidate mechanisms for the migration processes.

In this work a computational study has been performed for 1-chloro-3,5,5-trimethylhydantoin (**1**) and 3-chloro-1,5,5-trimethylhydantoin (**2**) (see Figure 1). The focus of the study was to rationalize experimentally determined products and rates of photolytic decomposition of the two model compounds employing density functional theory (DFT) calculations to determine optimum predicted mechanistic pathways for the formation of the resulting products. Both intramolecular and intermolecular processes have been explored. Prior computational studies relevant to *N*-halamine chemistry have been reported, i.e., a study of the stabilities of N-Cl bonds,<sup>8</sup> the mechanism of formation of biocidal imidazolidin-4-one derivatives,<sup>9</sup> and the mechanism of monochlorination of 5,5-dimethylhydantoin through a dichloro intermediate.<sup>10</sup>



**Figure 1.1** Structures of 1-chloro-3,5,5-trimethylhydantoin (**1**) and 3-chloro-1,5,5-trimethylhydantoin (**2**).

## 1.2 Pairwise Alternatives to Ewald Summation for Calculating Long-Range Electrostatics in Ionic Liquids

The computational cost of Monte Carlo (MC) and molecular dynamics (MD) methods can present a major obstacle for exhaustive simulations containing thousands or millions of atoms. Periodic boundary conditions minimize the number of atoms required to simulate a bulk environment, but also introduce the challenge of evaluating potentials beyond the minimum-image convention. A simple spherical cutoff that only computes nonbonded interactions within a specified distance can improve the efficiency. However, the potential for error is particularly large in the electrostatics portion, as the potentials generally decay to zero more slowly with increasing distance than the Lennard-Jones function.<sup>9-13</sup>

Ewald (or lattice) summation techniques can provide an accurate electrostatic treatment by using a cutoff distance for the quickly decaying short-ranged real-space summation and by performing a second long-ranged reciprocal-space summation.<sup>14,15</sup> As the number of particles,  $N$ , increase, a simple Ewald implementation can increase the simulation effort as  $O(N^2)$ . Optimization of the reciprocal-space summation with Fourier-based approaches, e.g., particle-particle particle-mesh (P3M)<sup>16</sup> and particle-mesh Ewald (PME),<sup>17</sup> has enabled the Ewald sums to scale as low as  $O(N \log(N))$ .<sup>15,17,18</sup>

Alternatively, truncation artifacts could be reduced by the use of a *shifted potential* scheme that provides a continuous shifting of the potential at all distances such that the value of the potential (or the value and first derivative for a *shifted force potential*)

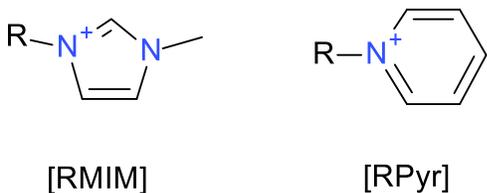
becomes zero at the cutoff distance.<sup>19</sup> For example, the CHARMM program<sup>20</sup> implements an efficient shifted force potential (Eq. 1.2.1),<sup>21</sup> where  $R_C$  is the cutoff distance,  $r_{ij}$  is the distance between particles, and  $q_i$  and  $q_j$  are the fixed atomic charges. Recent protein folding MD simulations over long time scales, e.g., up to 125  $\mu$ s, have shown the CHARMM-SF method to be reasonably effective at treating long-range electrostatics interactions at cutoffs greater than 9 Å.<sup>22</sup>

$$V_{CHARM}^{SF}(r_{ij}) = \begin{cases} \frac{q_i q_j}{r_{ij}} \left[ 1 - \left( \frac{r_{ij}}{R_C} \right)^2 \right]^2, & r_{ij} \leq R_C \\ 0, & r_{ij} > R_C \end{cases} \quad (1.2.1)$$

Of particular interest is the application of pairwise electrostatic interaction methods to the simulation of room temperature ionic liquids. Ionic liquids are a unique class of solvents, defined as a material containing only ionic species. They can be fluid at temperatures as low as 204 K, have low viscosities and vapor pressures, excellent thermal and chemical stabilities, are recyclable, and tolerate impurities such as water.<sup>23,24</sup> Ion components can be fine-tuned through different functional groups to enhance the degree of localized structuring in the liquid phase, which distinguishes ionic liquids from molecular solvents and solutions containing dissociated ions. Consequently, a large amount of computer time is required for the convergence of solvent properties due to the extensive sampling required and the costly evaluation of Ewald summations. Some pairwise electrostatic alternatives to Ewald have been applied to individual room temperature ionic liquid simulations. For example, Shi and Maginn employed the *damped shifted force* (DSF) method of Fennell and Gezelter<sup>25</sup> on the ionic liquid 1-n-hexyl-3-

methylimidazolium bis(trifluoromethylsulfonyl)imide [HMIM][Tf<sub>2</sub>N] with good reproducibility of the Ewald electrostatics by using a cutoff value of 12 Å and adjusting the damping parameter to 0.2022 Å<sup>-1</sup>.<sup>26</sup> Subsequent ionic liquid studies have implemented the Wolf et al. method<sup>27</sup> or used Shi and Maginn's recommendation for a small number of imidazolium-based ionic liquids.<sup>28-31</sup>

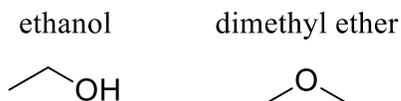
In this thesis, 59 unique ionic liquid combinations of 1-alkyl-3-methylimidazolium [RMIM], where R = M (methyl), E (ethyl), B (butyl), H (hexyl), and O (octyl), and N-alkylpyridinium [RPyr] cations (Scheme 1), along with Cl<sup>-</sup>, PF<sub>6</sub><sup>-</sup>, BF<sub>4</sub><sup>-</sup>, NO<sub>3</sub><sup>-</sup>, AlCl<sub>4</sub><sup>-</sup>, Al<sub>2</sub>Cl<sub>7</sub><sup>-</sup>, and TfO<sup>-</sup> anions have been tested at 25 °C using the *shifted potential* (SP) and *damped shifted potential* (DSP) methods of Wolf et al.,<sup>27</sup> the *shifted force* (SF) method of Levitt et al.,<sup>32</sup> the DSF method of Fennell and Gezelter,<sup>25</sup> the *shifted force gradient* (SFG) method of Kale and Herzfeld,<sup>33</sup> the *damped Columbic potential* (DC) method of Zahn et al.,<sup>34</sup> the CHARMM-SF method,<sup>21</sup> and our proposed *shifted force 3<sup>rd</sup> derivative* (SF3) method. The full Ewald sums energies are used as the benchmark to gauge the accuracy of each method. Multiple cutoffs, i.e., 9, 12, and 15 Å, and damping values ( $\alpha = 0.1, 0.2, \text{ and } 0.3 \text{ \AA}^{-1}$ ) were examined resulting in over 5000 molten salt/pairwise interaction combinations tested. This extensive study provides evidence that simple pairwise electrostatic interaction alternatives to Ewald for computing electrostatics can be an efficient and accurate method for use in the simulation of room temperature ionic liquids. Specific recommendations are given based on the findings.



**Figure 1.2** Ionic liquid forming cations. R = M (methyl), E (ethyl), B (butyl), H (hexyl), and O (octyl).

### 1.3 Free Energy Differences of Protomers in Solution

Isomers are compounds that possess the same simple chemical formula but vary in their molecular makeup. Two types of isomers are constitutional isomers and stereoisomers. Constitutional isomers possess different atom connectivities, such as ethanol,  $\text{CH}_3\text{CH}_2\text{OH}$ , and dimethyl ether,  $\text{CH}_3\text{OCH}_3$ , shown in figure 1.3. Stereoisomers, on the other hand, share the same atomic connectivities but oriented differently in space.



**Figure 1.3** Illustration of constitutional isomers ethanol and dimethyl ether.

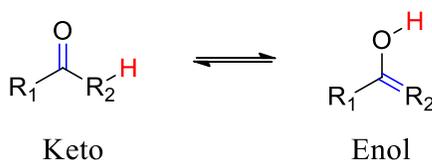
Isomerization is the process by which one isomer converts into another. This is of great importance to chemists because isomers do not have the same chemical properties. A particularly important type of isomerization is that of tautomerization, in which the two forms of the isomer are in rapid equilibrium with one another. Tautomers establish an

equilibrium constant  $K_T$ , which is proportional to their energy difference  $\Delta G$ , according to the equation 1.3.1,

$$K_T = \frac{[\textit{isomer A}]}{[\textit{isomer B}]} = e^{\frac{-\Delta G}{RT}} \quad (1.3.1)$$

where [ ] represents the concentration of the isomer,  $e$  is a constant of value approximately 2.718,  $R$  is the gas constant  $1.987 \times 10^{-3} \text{ kcal}\cdot\text{K}^{-1}\cdot\text{mol}^{-1}$ , and  $T$  is the absolute temperature. The tautomer that possesses the least free energy will be the dominant form.

The particular type of tautomerization explored in this thesis is protomerization, in which a proton is changing bond connectivity. Keto-enol tautomerization is illustrated in figure 1.4.



**Figure 1.4** Keto-enol tautomerization illustrating the differing hydrogen connectivity.

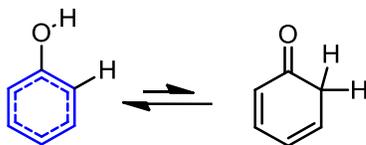
The keto form typically dominates due to its greater bond enthalpy. All other factors being equal, the difference between the keto and the enol are three bonds: the keto possesses a C-H, a C-C, and a C=O bond; the enol, a O-H, a C=C, and a C-O bond. An analysis of the different average bond enthalpies for these bonds, given in  $\text{kcal}\cdot\text{mol}^{-1}$ , is displayed in the table 1.1. On average, the keto form is more stable by approximately 12

$\text{kcal}\cdot\text{mol}^{-1}$ . However, there are cases where the enol is present to a significant degree and may even predominate.

**Table 1.1** Average bond enthalpies ( $\text{kcal}\cdot\text{mol}^{-1}$ ) for bonds distinguishing keto and enol tautomers.

Keto Form		Enol Form	
<b>C-H</b>	96 – 99	<b>O-H</b>	110 – 111
<b>C-C</b>	83 – 85	<b>C=C</b>	146 – 151
<b>C=O</b>	173 – 181	<b>C-O</b>	85 – 91
<b>Sum</b>	<b>352 – 365</b>		<b>341 – 353</b>

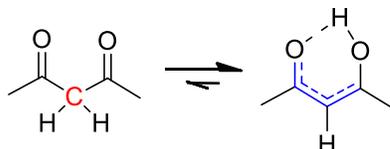
For phenol, tautomerization from the enol leads to loss of aromaticity. The  $K_T$  for this compound is  $4\times 10^{13}$ , signifying practical exclusivity of the enol form.



**Figure 1.5** Phenol keto-enol tautomerization. The keto form loses aromaticity (colored blue), a highly stabilizing factor.

Other cases include molecules with steric hindrance at the  $\alpha$ -carbon. The  $\alpha$ -carbon is tetrahedral in the keto form, placing substituents approximately  $109.5^\circ$  apart, while the enol  $\alpha$ -carbon is trigonal planar, with substituents approximately  $120^\circ$  apart.

Bulky substituents shift the equilibrium to the planar, sterically relaxed configuration of the enol form.  $\beta$ -diketones, in which the enol possesses a conjugated  $\pi$ -system and intramolecular hydrogen bond, also possess significant concentrations of the enol form. The phenomenon is observed for acetylacetone, shown in figure 1.6.



**Figure 1.6** Tautomerization of the  $\beta$ -diketone acetylacetone. The  $\alpha$ -carbon and conjugated  $\pi$ -system is colored red and blue, respectively. The hydrogen bond is indicated with a dashed line.

By equation 1.3.1, free energy differences dictate  $K_T$ . In the isothermal-isobaric ensemble, free energy difference is calculated by

$$\Delta G = \Delta H - T\Delta S \quad (1.3.2)$$

where  $G$  is the Gibbs free energy,  $H$  is the enthalpy,  $T$  is the absolute temperature, and  $S$  is the entropy of the isomer. As indicated by equation 1.3.2, temperature affects  $K_T$  via the change in the entropy term. In the case of acetylacetone, figure 1.6, the gas phase enol content is 95 percent at 22°C (295K) but only 44 percent at 275°C (548K), marking a shift back to the predominance of the keto form.<sup>35</sup> Additionally, the presence of solvent may dramatically affect  $K_T$ . The gas phase free energy of tautomerization  $\Delta G_{taut}$  for the acetylacetone is  $-2.2 \text{ kcal}\cdot\text{mol}^{-1}$ , favoring the enol form.<sup>36</sup> In aqueous solution, however,  $\Delta G_{taut}$  changes to  $1.11 \text{ kcal}\cdot\text{mol}^{-1}$ , favoring the keto form.<sup>37</sup> In water, the intermolecular

hydrogen bonding between the two carbonyl groups and multiple hydrogen bonding donors of the solvent stabilizes the keto form more than the conjugated  $\pi$ -system and intramolecular hydrogen bonding of the enol form.

It is desirable to determine quickly which protomer is favored in order to perform a synthesis with desired selectivity or designing a biologically active molecule. Protomerization may also occur traversing a cell membrane, affecting the reactivity of pharmaceuticals. Chemical calculations have the potential of determining the predominant form, yet prediction of solution phase protomeric equilibrium constants remains a challenge to computational chemistry.<sup>38–41</sup> The focus of this investigation is to find the most accurate computational model for calculating  $\Delta G_{taut}$  for various protomerizations in several solvents. Our strategy is to find the most accurate model for the calculation of gas phase  $\Delta G_{taut}$  compared to experimental data. Then, the best model is used in conjunction with several implicit solvation models and solvent cavity sizes to benchmark the calculated solution phase  $\Delta G_{taut}$  to experimental data.

## 1.4 References

- (1) Worley, S. D.; Williams, D. E.; Crawford, R. A. *Critical Reviews in Environmental Control* **1988**, *18*, 133–175.
- (2) Worley, S. D.; Sun, G. *Trends in Polymer Science* **1996**, *4*, 364–370.
- (3) Kenawy, E.-R.; Worley, S. D.; Broughton, R. *Biomacromolecules* **2007**, *8*, 1359–84.

- (4) Chen, Y.; Worley, S. D.; Kim, J.; Wei, C.-I.; Chen, T.-Y.; Santiago, J. I.; Williams, J. F.; Sun, G. *Industrial & Engineering Chemistry Research* **2003**, *42*, 280–284.
- (5) Sun, G.; Xu, X. *Textile Chemist and Colorist* **1998**, *30*, 26–30.
- (6) Luo, J.; Porteous, N.; Sun, Y. *ACS Applied Materials & Interfaces* **2011**, *3*, 2895–903.
- (7) Zhao, N.; Zhanel, G. G.; Liu, S. *Journal of Applied Polymer Science* **2011**, *120*, 611–622.
- (8) Makal, U.; Wood, L.; Ohman, D. E.; Wynne, K. J. *Biomaterials* **2006**, *27*, 1316–26.
- (9) Madura, J.; Pettitt, B. M. *Chemical Physics Letters* **1998**, *150*, 105–108.
- (10) Schreiber, H.; Steinhauser, O. *Chemical Physics* **1992**, *168*, 75–89.
- (11) Guenot, J.; Kollman, P. A. *Journal of Computational Chemistry* **1993**, *14*, 295–311.
- (12) Aufinger, P.; Beveridge, D. L. *Chemical Physics Letters* **1995**, *234*, 413–415.
- (13) Shirts, M. R.; Mobley, D. L.; Chodera, J. D.; Pande, V. S. *Journal of Physical Chemistry B* **2007**, *234*, 413–415.
- (14) Ewald, P. *Annalen der Physik* **1921**, *64*, 253–287.

- (15) Toukmaji, A. Y.; Boars Jr., J. A. *Computational Physics Communications* **1996**, *95*, 73–92.
- (16) Deserno, M.; Holm, C. *Journal of Chemical Physics* **1998**, *109*, 7678–7693.
- (17) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *Journal of Chemical Physics* **1995**, *103*, 8557–8593.
- (18) Darde, T.; York, D.; Pedersen, L. *Journal of Chemical Physics* **1993**, *98*, 10089–10092.
- (19) Brooks III, C. L.; Pettitt, B. M.; Karplus, M. *Journal of Chemical Physics* **1985**, *83*, 5897–5908.
- (20) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal of Computational Chemistry* **1983**, *4*, 187–217.
- (21) Steinbach, P. J.; Brooks, B. R. *Journal of Computational Chemistry* **1994**, *15*, 667–683.
- (22) Piana, S.; Lindorff-Larsen, K.; Dirks, R. M.; Salmon, J. K.; Dror, R. O.; Shaw, D. E. *PLoS One* **2012**, *7*, e39918.
- (23) Hallett, J. P.; Welton, T. *Chemical Reviews* **2001**, *111*, 3508–3576.
- (24) Welton, T. *Chemical Reviews* **1999**, *99*, 2071–2083.
- (25) Fennell, C. J.; Gezelter, J. D. *Journal of Chemical Physics* **2006**, *124*, 234104.

- (26) Shi, W.; Maginn, E. J. *Journal of Physical Chemistry B* **2008**, *112*.
- (27) Wolf, D.; Koblinski, P.; Phillpot, S. R.; Eggebrecht, J. *Journal of Chemical Physics* **1999**, *110*, 8254–8282.
- (28) Shi, W.; Myers, C. R.; Luebke, D. R.; Steckel, J. A.; Sorescu, D. C. *Journal of Physical Chemistry B* **2012**, *116*, 283–295.
- (29) Ghobadi, A. F.; Taghikhani, V.; Elliot, J. R. *Journal of Physical Chemistry B* **2011**, *115*, 13599–13607.
- (30) Shi, W.; Sorescu, D. C. *Journal of Physical Chemistry B* **2010**, *114*, 15029–15041.
- (31) Shi, W.; Sorescu, D. C.; Luebke, D. R.; Keller, M. J.; Wickramanayake, S. *Journal of Physical Chemistry B* **2010**, *114*, 6531–6541.
- (32) Levitt, M.; Hirshberg, M.; Sharon, R.; Daggett, V. *Computational Physics Communications* **1995**, *91*, 215–291.
- (33) Kale, S.; Herzfeld, J. *Journal of Chemical Theory and Computation* **2011**, *7*, 215–231.
- (34) Zahn, D.; Schilling, B.; Kast, S. M. *Journal of Physical Chemistry B* **2002**, *106*.
- (35) Smith, M.; March, J. *March's Advanced Organic Chemistry*; 6th ed.; John Wiley and Sons Ltd.: Hoboken, New Jersey, 2007.

- (36) Folkendt, M. M.; Weiss-Lopez, B. E.; Chauvel, J. P.; True, N. S. *The Journal of Physical Chemistry* **1985**, *89*, 3347–3352.
- (37) Toullec, J. In *The chemistry of enols*; Rappoport, Z., Ed.; John Wiley and Sons Ltd.: West Sussex, England, 1990; pp. 323–398.
- (38) Klamt, A.; Diedenhofen, M. *Journal of Computer-Aided Molecular Design* **2010**, *24*, 621–5.
- (39) Kwiatkowski, J. S.; Zielinski, T. J.; Rein, R. In *Advances in Quantum Chemistry*; Lowdin, P.-O., Ed.; Academic Press: Orlando, 1986; pp. 85–130.
- (40) Sayle, R. *Journal of Computer-Aided Molecular Design* **2010**, *24*, 485–96.
- (41) Szafran, M.; Karelson, M. M.; Katritzky, A. R.; Koput, J.; Zerner, M. C. *Journal of Computational Chemistry* **1993**, *14*, 371–377.

## Chapter 2 Methods

Computational chemistry is a collection of methods that use physics and mathematics to simulate molecular structures, reactions, and environments in order to predict and interpret chemical information. When simulating molecular systems, two broad methods used are classical mechanics (or molecular mechanics), and quantum mechanics, referred to as *ab initio* methods. Statistical thermodynamics provides the necessary means to bridge the results of these two methods with thermodynamic physical observables.

### 2.1 *ab initio* Quantum Chemistry

#### 2.1.1 The Schrödinger Equation

The rigorous mathematical description of atoms and molecules requires the use of the Schrödinger equation. The time dependent form must be used when external potentials change with time (equation 2.1.1).

$$\hat{H}\Psi(\bar{x}, t) = -\frac{\hbar}{i} \frac{\partial \Psi(\bar{x}, t)}{\partial t} \quad (2.1.1)$$

$\Psi(\bar{x}, t)$  is the wavefunction of the system and characterizes the position and momenta of the particles within the system,  $\bar{x}$  is the spatial coordinate,  $t$  is the time coordinate,  $i$  is the square root of negative one, and  $\hbar$  is Planck's constant divided by  $2\pi$ .<sup>1</sup> While  $\Psi(\bar{x}, t)$  arguably has no physical significance, the absolute value of its square gives the probability  $P$  at time  $t$  of finding the particle in question in the region of space between  $\bar{x}$  and  $\bar{x} + d\bar{x}$ , as illustrated in equation 2.1.2.

$$P = |\Psi(\bar{x}, t)|^2 d\bar{x} \quad (2.1.2)$$

$\hat{H}$  is the Hamiltonian operator and represents the forces associated the system. It comprises a sum of the kinetic energy operator  $\hat{T}$  and potential energy operator  $\hat{V}$ .

$$\hat{H} = \hat{T} + \hat{V} \quad (2.1.3)$$

The kinetic energy operator  $\hat{T}$  is summed over all of the  $N$  nuclei and  $M$  electrons for the system.  $\hat{T}$  is separated into a nuclear kinetic operator  $\hat{T}_N$ , equation 2.1.4, and an electronic kinetic operator  $\hat{T}_e$ , equation 2.1.5, the mathematical forms of which are

$$\hat{T}_N = \sum_{A=1}^N -\frac{\hbar^2}{2m_A} \frac{\partial^2}{\partial \bar{x}_A^2} \quad (2.1.4)$$

$$\hat{T}_e = \sum_{a=1}^M -\frac{\hbar^2}{2m_a} \frac{\partial^2}{\partial \bar{x}_a^2} \quad (2.1.5)$$

where  $m$  is the mass of the individual nuclei or electron, upper cased indices refer to individual nuclei, and lower cased indices refer to individual electrons.

The potential energy operator  $\hat{V}$  of equation 2.1.3 is a pair-wise summation over all of the electrons and nuclei found in the system and corresponds to the electrostatic interactions between these particles.  $\hat{V}$  includes separate types of interactions, as shown in equation 2.1.6: nucleus-nucleus,  $\hat{V}_{NN}$ , nucleus-electron,  $\hat{V}_{Ne}$ , and electron-electron,  $\hat{V}_{ee}$ , shown in equations 2.1.7, 2.1.8, and 2.1.9, respectively.

$$\hat{V} = \hat{V}_{NN} + \hat{V}_{Ne} + \hat{V}_{ee} \quad (2.1.6)$$

The mathematical forms of these operators for a system of  $N$  number of nuclei and  $M$  number of electrons are

$$\hat{V}_{NN} = \sum_{A=1}^N \sum_{B>A}^N \frac{e^2 Z_A Z_B}{4\pi\epsilon_0 r_{AB}} \quad (2.1.7)$$

$$\hat{V}_{Ne} = - \sum_{A=1}^N \sum_{a=1}^M \frac{e^2 Z_A}{4\pi\epsilon_0 r_{aA}} \quad (2.1.8)$$

$$\hat{V}_{ee} = \sum_{a=1}^M \sum_{b>a}^M \frac{e^2}{4\pi\epsilon_0 r_{ab}} \quad (2.1.9)$$

where upper cased indices indicate individual nuclei, lower cased indices indicate individual electrons,  $e$  is the electrical charge associated with an electron,  $Z$  is the electrical charge of the nuclei,  $\epsilon_0$  is the permittivity of free space, and  $r$  is the radius between the two particles in question. The potential energy operator would also include any potential external to the system, such as the potential created by solvent.

Typically, in dealing with the above equations, atomic units are used such that the unit of charge is the absolute value of charge of the electron, the unit of mass is the mass of the electron, the unit of length is 1 Bohr (the radius of the 1s orbital of hydrogen), and the unit of energy is the Hartree.<sup>2</sup> This simplifies the equations to become e.g. equation 2.1.10.

$$\hat{V}_{ee} = \sum_{a=1}^M \sum_{b>a}^M \frac{1}{r_{ab}} \quad (2.1.10)$$

## 2.1.2 The Time-Independent Schrödinger Equation

When the potential external to the system does not change with time, the wavefunction is separable into a spatial function and a time function. (Equation 2.1.11)

$$\Psi(\bar{x}, t) = \psi(\bar{x})\phi(t) \quad (2.1.11)$$

The Schrödinger equation is algebraically manipulated to gather all time-dependent terms on one side of the equation and all position-dependent terms on the other side of the equation. (Equations 2.1.12-2.1.15)

$$\hat{H}\Psi(\bar{x}, t) = -\frac{\hbar}{i} \frac{\partial \Psi(\bar{x}, t)}{\partial t} \quad (2.1.12)$$

$$\hat{H}\psi(\bar{x})\phi(t) = -\frac{\hbar}{i} \frac{\partial \psi(\bar{x})\phi(t)}{\partial t} \quad (2.1.13)$$

$$\phi(t)\hat{H}\psi(\bar{x}) = -\frac{\hbar}{i} \frac{\partial \phi(t)}{\partial t} \psi(\bar{x}) \quad (2.1.14)$$

$$\frac{\hat{H}\psi(\bar{x})}{\psi(\bar{x})} = -\frac{\hbar}{i} \frac{\partial \phi(t)}{\partial t} \frac{1}{\phi(t)} \quad (2.1.15)$$

The left hand side of the of equation 2.1.15 is independent of the time coordinate while the right hand side of the equation is independent of the spatial coordinate. Therefore, the overall equation is independent of both the time coordinate and the spatial coordinate and hence is constant. This constant is postulated to be the energy  $E$  of the system.<sup>1</sup> Setting the left hand side of the equation to  $E$ , yields

$$\frac{\hat{H}\psi(\bar{x})}{\psi(\bar{x})} = E \quad (2.1.16)$$

$$\hat{H}\psi(\bar{x}) = E\psi(\bar{x}) \quad (2.1.17)$$

Equation 2.1.17 is the time-independent Schrödinger equation and characterizes stationary states of the wavefunction, in which the total energy remains constant. Formally, the definition of a stationary state is one wherein the energy of the system is an eigenvalue of the Hamiltonian.<sup>2</sup> All *ab initio* simulations conducted in this research use this time-independent form of the Schrödinger equation to calculate the molecular energy.

Algebraic manipulation of the Schrödinger equation isolates the energy term, viz. pre-multiplication of equation 2.1.17 by  $\psi^*$ , the complex conjugate of  $\psi$ , and then integrating over all space.

$$\int_{-\infty}^{\infty} \psi^*(\bar{x}) \hat{H} \psi(\bar{x}) d\bar{x} = \int_{-\infty}^{\infty} \psi^*(\bar{x}) E \psi(\bar{x}) d\bar{x} \quad (2.1.18)$$

The energy  $E$  is constant and thus removed from the integral on the right hand side.

$$\int_{-\infty}^{\infty} \psi^*(\bar{x}) \hat{H} \psi(\bar{x}) d\bar{x} = E \int_{-\infty}^{\infty} \psi^*(\bar{x}) \psi(\bar{x}) d\bar{x} \quad (2.1.19)$$

Isolating  $E$  yields the formula for calculating the energy of the system. (Equation 2.1.20)

$$E = \frac{\int_{-\infty}^{\infty} \psi^*(\bar{x}) \hat{H} \psi(\bar{x}) d\bar{x}}{\int_{-\infty}^{\infty} \psi^*(\bar{x}) \psi(\bar{x}) d\bar{x}} \quad (2.1.20)$$

Additionally, the wavefunction is normalized, setting the denominator equal to one. (Equation 2.1.21)

$$\int_{-\infty}^{\infty} \psi^*(\bar{x}) \psi(\bar{x}) d\bar{x} = 1 \quad (2.1.21)$$

Therefore, the energy is solved from a simplified form. (Equation 2.1.22)

$$E = \int_{-\infty}^{\infty} \psi^*(\bar{x}) \hat{H} \psi(\bar{x}) d\bar{x} \quad (2.1.22)$$

Or, represented in the more compact bra-ket Dirac notation

$$E = \langle \psi(\bar{x}) | \hat{H} | \psi(\bar{x}) \rangle \quad (2.1.23)$$

It is not possible, however, to solve the Schrödinger equation for a system of three or more bodies. Many strategies for approximating this intractable problem exist. Specific sets of approximations are referred to as a theoretical model or level of theory.

### 2.1.3 The Variational Principle

The variation principle states that the energy of an approximate wavefunction  $\psi_{approx}$  will always be greater than the energy of the true wavefunction  $\psi_{exact}$ .

$$\langle \psi_{approx} | \hat{H} | \psi_{approx} \rangle = E_{approx} \geq E_{exact} \quad (2.1.24)$$

Therefore, the wavefunction that gives a lower expectation value of the energy is a better approximation of the true wavefunction. Over the course of a calculation, the wavefunction is varied systematically until the lowest energy wavefunction is found.

### 2.1.4 The Born-Oppenheimer Approximation

Because the mass of a proton is approximately eighteen hundred times the mass of an electron, the nuclei are relatively stationary compared to the electrons. Hence, the electrons equilibrate instantaneously to any movement of the nuclei. This approximation simplifies the Schrödinger equation to reflect stationary nuclei and dynamic electrons. The full wavefunction  $\psi$  is separated into a nuclear wavefunction  $\psi_N$  and an electronic wavefunction  $\psi_e$ . (Equation 2.1.25)

$$\psi = \psi_N \psi_e \quad (2.1.25)$$

Fixed nuclei have zero kinetic energy and constant nuclear-nuclear potential energy. The resultant Hamiltonian is referred to as the electronic Hamiltonian  $\hat{H}_e$ . (Equation 2.1.26)

$$\hat{H}_e = \hat{T}_e + \hat{V}_{Ne} + \hat{V}_{ee} \quad (2.1.26)$$

The electronic Schrödinger equation thus becomes

$$\hat{H}_e \psi_e = E_e \psi_e \quad (2.1.27)$$

where the eigenvalue  $E_e$  is the electronic energy of the molecule to which may be added the energy of the electrostatic repulsion between the nuclei (equation 2.1.7).

Optimization of the nuclear coordinates requires successive energy calculations until a lowest energy nuclear geometry is found.

### 2.1.5 Finding the Electronic Wavefunction

The electronic wavefunction is approximated by means of molecular orbitals, which are described not only in terms of the space they occupy, but also in terms of the spin which its occupying electron possesses, either spin up  $\alpha$  or spin down  $\beta$ . These are referred to as spin orbitals. For example, ground-state molecular hydrogen possesses two electrons  $1$  and  $2$  occupying two spin orbitals  $\psi_1$  and  $\psi_2$  which both in turn occupy the same spatial orbital  $\psi_\sigma$ , such that they may be represented as

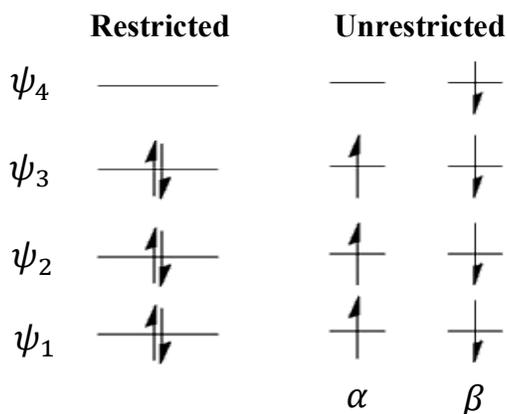
$$\psi_1 = \psi_\sigma(1)\alpha(1) \quad (2.1.28)$$

$$\psi_2 = \psi_\sigma(2)\beta(2) \quad (2.1.29)$$

#### 2.1.5.1 Open and Closed Shell Systems

A system that possesses electrons paired into spatial orbitals is a closed or restricted shell system. Being fermions, electrons in the same spatial orbital must have opposite spin. The closed shell form allows the determination of the wavefunction as the variationally optimized spatial aspect of the spin orbitals.

Alternatively, open shell systems possess spatially unpaired electrons. Calculations on these systems are called unrestricted and each spin orbital must be variationally optimized.



**Figure 2.1** Illustration of the separation of spatial orbitals into separate  $\alpha$  and  $\beta$  spin orbitals for open shell systems.

Another method of calculating open shell systems is to divide the molecular orbitals into doubly-occupied and singly-occupied orbitals, which are then treated in a closed shell and open shell fashion, respectively.

### 2.1.5.2 The Hartree Product and Slater Determinants

A first approximation to the relationship between the wavefunctions of the molecular orbitals and the total wavefunction is the Hartree product, in which the total electronic wavefunction  $\Psi$  is the product of the individual orbital wavefunctions  $\psi_i$ . For a molecule of  $M$  electrons,

$$\Psi = \psi_1\psi_2\psi_3 \dots \psi_M \quad (2.1.30)$$

For molecular hydrogen, the Hartree product would be

$$\Psi = \psi_\sigma(1)\alpha(1)\psi_\sigma(2)\beta(2) \quad (2.1.31)$$

The Hartree product indicates that the total electronic energy of the molecule is the sum of the energies of the individual one-electron spin orbitals and that the electrons

do not interact with one another. However, the motions of the electrons correlate, minimizing their electrostatic repulsions from one another. Therefore, the more electrons in a given molecule, the more the Hartree product will fail to produce an accurate energy. Additionally, this simple summation of orbital energies double-counts the electron repulsion. The Hartree product also violates the Pauli antisymmetry principle: if two electrons exchange position, the new wavefunction must have the opposite sign of the former wavefunction.<sup>2</sup> If  $\hat{P}_{1,2}$  is an operator that exchanges the positions of two electrons, then

$$\hat{P}_{1,2} \Psi = -\Psi \quad (2.1.32)$$

The solution is to construct the total electronic wavefunction as a Slater determinant comprising the molecular orbitals representative of the electronic configuration. Thus, for the ground state of molecular hydrogen, the Slater determinant would be

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_{\sigma}(1)\alpha(1) & \psi_{\sigma}(1)\beta(1) \\ \psi_{\sigma}(2)\alpha(2) & \psi_{\sigma}(2)\beta(2) \end{vmatrix} \quad (2.1.33)$$

which expands to

$$\Psi = \frac{1}{\sqrt{N!}} (\psi_{\sigma}(1)\alpha(1)\psi_{\sigma}(2)\beta(2) - \psi_{\sigma}(1)\beta(1)\psi_{\sigma}(2)\alpha(2)) \quad (2.1.34)$$

where  $\frac{1}{\sqrt{N!}}$  is the normalization factor and  $N$  is the number of electrons. This form of the electronic wavefunction meets the antisymmetry requirement of Pauli's principle.<sup>2</sup>

Additionally, the full electronic wavefunction is better approximated through a linear combination of Slater determinants. A single determinant confines the electrons to the forms of the molecular orbital wavefunctions used in the determinant. Using multiple determinants allows the electrons extra spatial degrees of freedom and, therefore, allows

them to correlate their motion. If an infinite number of Slater determinants are used, with an infinite basis set (*vide infra*), then the calculated wavefunction is the exact solution to the Schrödinger equation.

### 2.1.5.3 Basis sets

The molecular orbitals that comprise the Slater determinant are constructed as a linear combination of the atomic orbitals present within the atoms that make up the molecule. For example, the doubly occupied spatial molecular orbital  $\psi_\sigma$  of  $H_2$  is a linear combination of the two  $1s$  atomic orbitals present upon each hydrogen atom, represented

$$\psi_\sigma = A(1s + 1s) \quad (2.1.35)$$

where  $A$  is a normalization factor. For a general polyatomic, molecular orbitals are formulated as

$$\psi_i = A \sum_{v=1}^K c_{vi} \phi_v \quad (2.1.36)$$

where  $i$  is the molecular orbital being constructed, the subscript  $v$  represents an individual atomic orbital,  $K$  is the total number of atomic orbitals,  $c$  is the coefficient for the individual atomic orbital and represents its contribution to the molecular orbital, and  $\phi$  is the atomic orbital wavefunction. This approach is referred to as the linear combination of atomic orbital molecular orbital, or LCAO-MO.

The set of atomic orbital mathematical functions used for the construction of the molecular orbitals constitute a basis set. There are generally three types of basis sets. Valence basis sets use only the valence atomic orbitals to construct molecular orbitals. Minimal basis sets included core and valence electrons, but utilize only the minimum

number of atomic orbitals to represent the valence shells. Minimal basis sets also express the atomic orbital as a single mathematical function, the relevance of which is discussed below. If a basis set uses more than the minimal set and expresses them with more than one mathematical function, it is referred to as an extended basis set.

The atomic orbitals themselves are wavefunctions which consist of a radial function  $R(r)$  and an angular function  $\gamma(\omega, \rho)$ .

$$\phi_v = R_{nl}(r)\gamma_{lm}(\omega, \rho) \quad (2.1.37)$$

where  $n$  represents the principal quantum number, or the atomic shell,  $l$  the azimuthal quantum number, or the atomic subshell commonly referred to as s, p, d, etc.,  $m$  the magnetic quantum number, or the spatial orientation of the subshell such as  $p_x$ ,  $d_z^2$ , etc, and  $r$ ,  $\omega$ ,  $\rho$  are spherical polar coordinates.<sup>3</sup>

For atomic hydrogen, the radial function is a decaying exponential multiplied by a polynomial  $L(\rho)$ .

$$R_{nl}(r) = - \left[ \left( \frac{2Z}{na_o} \right)^3 \frac{(n-l-1)!}{2n[(n+1)!]^3} \right]^{\frac{1}{2}} e^{-\frac{\rho}{2}} \rho^l L_{n+1}^{2l+1}(\rho) \quad (2.1.38)$$

where  $Z$  is the nuclear charge,  $a_o$  is the Bohr radius, and  $\rho = \frac{2Zr}{na_o}$ . The bracketed term is the normalization factor. For polyelectronic atoms, the radial function is the solution for the hydrogen atom if the orbital exponent is adjusted to reflect the effective nuclear charge.

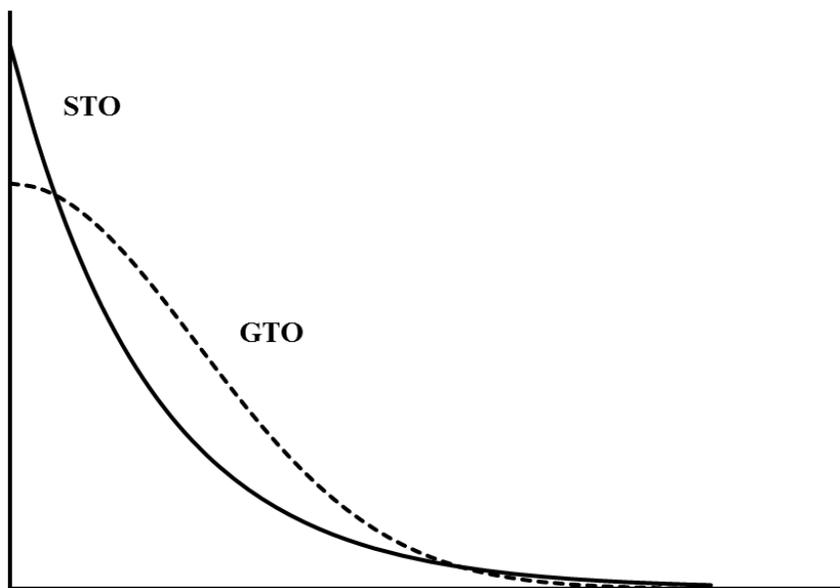
However, the hydrogenic function is too mathematically complex for molecular orbital calculations. Slater suggested a simpler radial function for the orbitals.<sup>4</sup>

$$R_{nl}(r) = (2\zeta)^{n+1/2} [(2n)!]^{-1/2} r^{n-1} e^{-\zeta r} \quad (2.1.39)$$

where  $\zeta$  represents the effective nuclear charge associated with the atomic orbital. While these Slater-type orbitals (STO) most accurately represent the shape of atomic orbitals, the  $e^{-\zeta r}$  term is difficult to solve in *ab initio* calculations that involve three- and four-center integrals.<sup>5</sup> An alternative is to use Gaussian-type orbitals (GTO) which exchange the  $e^{-\zeta r}$  term to an analytically integrable  $e^{-\alpha r^2}$ , where  $\alpha$  determines the width of the Gaussian curve. The full form of the radial portion of the GTO is

$$R_{nl}(r) = \left[ \frac{2^{2n+3/2}}{(2n-1)! \sqrt{\pi}} \right]^{1/2} \alpha^{(2n+1)/4} r^{n-1} e^{-\alpha r^2} \quad (2.1.40)$$

where  $n$  and  $\alpha$  are terms which are optimized to produce the desired shape and energy of the wavefunction.<sup>6</sup>



**Figure 2.2** Illustration of the hydrogen *1s* atomic orbital expressed as a STO and a variationally optimized single GTO.

The change in the form of the exponential in going from STO's to GTO's has two effects which can be seen in figure 2.2, in which the hydrogen 1s orbital is plotted as an exact STO

$$\phi_{STO} = \frac{1}{\sqrt{\pi}} e^{-r} \quad (2.1.41)$$

and the best single variationally optimized GTO

$$\phi_{GTO} = \left(\frac{9\pi^2}{16}\right)^{\frac{3}{2}} e^{-\frac{8}{9\pi}r^2} \quad (2.1.42)$$

The first effect is that, at  $r = 0$ , the STO has the shape of a cusp while the GTO has zero slope. The second is that the GTO decays more rapidly at larger values of  $r$ . In terms of its effects upon the representation of the atomic orbital, the GTO poorly represents the gradient of the electron density at the nucleus and underestimates electron density both at the nucleus and at larger distances from the nucleus. Fortunately, Gaussian wavefunctions may be summed to better resemble a Slater wavefunction.

The number and types of Gaussians used to create the atomic wavefunctions define a basis set. An example of a minimal basis set is STO-3G, in which three GTOs are used to mimic a single STO corresponding to a particular atomic wavefunction  $\phi_i$

$$\phi_i = a_1 G_1 + a_2 G_2 + a_3 G_3 \quad (2.1.43)$$

where  $a_i$  is a constant optimized to fit variational *ab initio* calculations. The  $G$ 's are primitive Gaussians while their sum,  $\phi_i$  in the case of equation 2.1.43, is a contracted Gaussian.

Extended basis sets may use more than one contracted Gaussian to represent an atomic orbital. For example, split valence basis sets will use two or more contracted Gaussians to represent the valence atomic orbitals of an atom. Such would be the 3-21G

basis set which uses a single contracted function comprised of three primitives for core orbitals (equation 2.1.44) and two contracted functions for the valence orbitals (equations 2.1.45 – 2.1.47), referred to as a double-zeta basis set.

$$\phi_{core} = a_1 G_1 + a_2 G_2 + a_3 G_3 \quad (2.1.44)$$

$$\phi_{valence} = c_1 \varphi_1 + c_2 \varphi_2 \quad (2.1.45)$$

$$\varphi_1 = a_1 G_4 + a_2 G_5 \quad (2.1.46)$$

$$\varphi_2 = a_3 G_6 \quad (2.1.47)$$

The valence atomic orbitals contain any number of contracted functions: two for a double-zeta, three for a triple-zeta, four for a quadruple-zeta, etc. Adjusting the coefficient for a particular primitive function adds flexibility to the basis set and results in a better approximation of an atomic function.

A basis set may also have more atomic orbitals than the minimum number required for all of its electrons. For example, hydrogen may have p-orbitals added to it, while carbon may have d-orbitals. This has the effect of polarizing the existing atomic orbitals, introducing asymmetry into them and better representing the anisotropy that exists in the electron distribution between atoms. These additional atomic wavefunctions are polarization functions and are necessary in energy calculations of molecules with three-center two-electron bonds, such as diborane, or molecules with high bond angle strain, such as cyclopropane.

Because GTO's underestimate electron density far from the nucleus, it is necessary to include a more spatially diffuse function into the Gaussian summation in order to represent larger atomic orbitals. This more accurately represents species such as

anions, as well as those possessing lone-pairs of electrons, and molecules in highly excited electronic states.

However, a basis set is not only the number of atomic orbitals used to represent molecular orbitals and the number and types of basis functions used to represent those atomic orbitals. A basis set also specifies the particular values of the  $n$  and  $\alpha$  terms used in the Gaussian functions (equation 2.1.39).

Dunning introduced a class of basis sets for which the coefficients and exponents of the functions have been parameterized using levels of theory which account for electron correlation, referred to as correlation consistent basis sets.<sup>7</sup> Polarization functions of higher angular momentum are added in a systematic manner such that each subsequent set of functions added recover a certain amount of correlation energy. Using these basis sets allows for an accurate extrapolation to the complete basis set.

#### 2.1.5.4 The Hartree-Fock approximation

The Hartree-Fock approximation represents an all-electron approximation to solving the Schrödinger equation. The wavefunction is represented as a single Slater determinant of spin orbitals that is variationally optimized. The electronic Hamiltonian operator is a sum of one-electron and two-electron operators.

$$\hat{H}_{elec} = \sum_{a=1}^N \hat{\mathcal{H}}_{core} + \sum_{i=1}^N \sum_{b>a}^N \frac{1}{r_{ab}} \quad (2.1.48)$$

In equation 2.1.48,  $\hat{\mathcal{H}}_{core}$  represents the core one-electron Hamiltonian corresponding to the kinetic energy of electron  $a$  and its electrostatic attraction to the  $N$  nuclei within the system. (Equation 2.1.49)

$$\hat{\mathcal{H}}_{core} = -\frac{1}{2} \frac{\partial^2}{\partial \bar{x}_a^2} \frac{\partial^2}{\partial \bar{y}_a^2} \frac{\partial^2}{\partial \bar{z}_a^2} - \sum_{A=1}^N \frac{Z_A}{r_{aA}} \quad (2.1.49)$$

The full derivation of the Hartree-Fock equations are not presented here, yet a brief discussion of the results is important for understanding not only Hartree-Fock but also the post-Hartree-Fock, density functional theory, and semi-empirical quantum mechanical methods that are discussed below.

First, the variational principle is invoked such that the derivative of the energy with respect to the nuclear coordinates,  $\partial E$ , must be zero. Second, constraining Lagrangian multipliers are imposed such that the molecular orbitals comprising the Slater determinant must remain orthonormal, expressed in terms of an overlap integral  $S$ .

$$S_{ij} = \langle \psi_i | \psi_j \rangle = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} = \delta_{ij} \quad (2.1.50)$$

where  $\delta_{ij}$  is the Kronecker delta. The resulting integro-differential equation is more easily illustrated with the introduction of the Coulomb (equation 2.1.51) and exchange (equation 2.1.52) operators,  $\mathcal{J}$  and  $\mathcal{K}$ , respectively.<sup>3</sup>

$$\mathcal{J}_j(1) = \left\langle \psi_j(2) \left| \frac{1}{r_{12}} \right| \psi_j(2) \right\rangle \quad (2.1.51)$$

$$\mathcal{K}_j(1)\phi_i(1) = \left\langle \psi_j(2) \left| \frac{1}{r_{12}} \right| \phi_i(2) \right\rangle \psi_j(1) \quad (2.1.52)$$

The exchange operator  $\mathcal{K}_j$  must be defined in terms of its effect upon  $\phi_i$  because it exchanges the location of electrons 1 and 2.

The Hartree-Fock equation for one electron thus becomes

$$\left[ \mathcal{H}_{core}(1) + \sum_{j=1}^N \{ \mathcal{J}_j(1) - \mathcal{K}_j(1) \} \right] \psi_i(1) = \sum_{j=1}^N \varepsilon_{ij} \psi_j(1) \quad (2.1.53)$$

The term in brackets in equation 2.1.53 is the Fock operator, denoted  $\hat{f}$ . Equation 2.1.53, however, is not an eigenvalue equation; the Fock operator operating upon orbital  $\psi_i$  does not yield the energy of the orbital multiplied by the orbital. This reflects the fact that the solutions to the Hartree-Fock equations are not unique. Equation 2.1.53 results from imposition of Lagrange multipliers; one may be canonical form of the Hartree Fock equations by requiring the  $\epsilon$  matrix to be diagonal. In the canonical form of the Hartree-Fock equation, a pseudo-eigenvalue equation yields the energy of the orbital  $\epsilon_i$  and the canonical orbital  $\psi_i$ . (Equation 2.1.54)

$$\hat{f}_i \psi_i = \epsilon_i \psi_i \quad (2.1.54)$$

Because the energy of each spin-orbital depends upon every other spin-orbital in the system and because each electron interacts with all other electrons only in a mean-field, non-dynamical fashion, the spin-orbitals must already be solved before the energy of any particular spin-orbital can be found. This is remedied by guessing the spin-orbitals and then systematically adjusting them until the energy reaches a minimum value, referred to as the self-consistent field method. The variational principle assures that the wavefunction that gives the lowest energy is the best representation of the wavefunction. The lowest energy achievable by Hartree-Fock is referred to as the Hartree-Fock limit and is attained by employing an infinite basis set. To move beyond this limit, approaches incorporating electron correlation must be used. It is worth noting, though, that Hartree-Fock lies at the foundation of all of these approximations.

### 2.1.5.5 The Roothaan-Hall equations

The derivation of the Hartree-Fock equations for closed shell systems in a fixed basis set were first independently proposed by Roothaan and Hall. Instead of the integro-

differential form of equation 2.1.53, Roothaan and Hall formulated the equations into matrix form. Briefly, substituting the LCAO-MO expansion of equation 2.1.36 into equation 2.1.54, yields equation 2.1.55.

$$\widehat{f}_i(1) \sum_{\nu=1}^K c_{\nu i} \phi_{\nu}(1) = \varepsilon_i \sum_{\nu=1}^K c_{\nu i} \phi_{\nu}(1) \quad (2.1.55)$$

Pre-multiplication of a second basis function  $\phi_{\mu}$  and integration yields the matrix equation 2.1.56.

$$\sum_{\nu=1}^K c_{\nu i} \langle \phi_{\mu}(1) | \widehat{f}_i(1) | \phi_{\nu}(1) \rangle = \varepsilon_i \sum_{\nu=1}^K c_{\nu i} \langle \phi_{\mu}(1) | \phi_{\nu}(1) \rangle \quad (2.1.56)$$

While molecular orbitals within Hartree-Fock are constrained to be orthonormal, the basis functions are not;  $\langle \phi_{\mu}(1) | \phi_{\nu}(1) \rangle$  does not have to yield the Kronecker delta. As in the case for molecular orbitals, this integral is termed the overlap integral  $S_{\mu\nu}$ .

Each term in equation 2.1.56 corresponds to an element in its corresponding matrix in equation 2.1.57, e.g.  $\widehat{f}_i$  is an element of the  $\mathbb{F}$  matrix,  $c_{\nu i}$  of  $\mathbb{C}$ ,  $S_{\nu\mu}$  of  $\mathbb{S}$ , and  $\varepsilon_i$  of  $\mathbb{E}$ .

$$\mathbb{F}\mathbb{C} = \mathbb{S}\mathbb{C}\mathbb{E} \quad (2.1.57)$$

For a closed shell system, the  $\mu\nu^{th}$  element of  $\mathbb{F}$  has the form

$$F_{\mu\nu} = \langle \mu | \widehat{\mathcal{H}}_{core} | \nu \rangle + \sum_{\lambda, \sigma=1}^{N/2} P_{\lambda\sigma} \left[ (\mu\nu | \lambda\sigma) - \frac{1}{2} (\mu\lambda | \nu\sigma) \right] \quad (2.1.58)$$

where the bond order (or charge density) matrix  $P_{\lambda\sigma}$  (equation 2.1.59) arises from the LCAO-MO expansion of the molecular orbitals, which are assumed to be real, within the Coulomb and exchange operators, wherein the molecular orbitals are assumed to be real.

$$P_{\lambda\sigma} = 2 \sum_{i=1}^{N/2} c_{\lambda i} c_{\sigma i} \quad (2.1.59)$$

Note that equations 2.1.58-59 have used Greek minuscules for basis functions and has employed Mulliken notation, in which the basis functions associated with electron one and two are grouped on the left and right sides within the parentheses, respectively, for the two-electron integrals.

The object then becomes to solve equation 2.1.57 and to diagonalize the  $\mathbb{E}$  matrix, whose eigenvalues will be the energies of the molecular orbitals. Once the energy of every molecular orbital  $\varepsilon_i$  is known, the total electronic energy is found by summing the individual molecular orbital energies and then subtracting the Coulomb and exchange interactions from each pair to avoid double counting of these interactions. For a closed shell molecule, the total electronic energy is

$$E_{elec} = \sum_{i=1}^N \varepsilon_i - \sum_{i=1}^{N/2} \sum_{j=1}^{N/2} (2J_{ij} - K_{ij}) \quad (2.1.60)$$

where

$$J_{ij} = (\phi_i^2 | \phi_j^2) \quad (2.1.61)$$

$$K_{ij} = (\phi_i \phi_j | \phi_i \phi_j) \quad (2.1.62)$$

Inclusion of the nuclear-nuclear repulsion energy (equation 2.1.7) to equation 2.1.60 yields the total molecular energy. For a geometry optimization, new nuclear coordinates are chosen and the Roothaan-Hall equations are resolved. This is repeated until a minimum molecular energy is found, or, in the case of a geometry optimization to a transition structure, the second derivative of the energy with respect to the nuclear

coordinates is negative (corresponding to a repulsive force constant) in one dimension on the potential energy hypersurface and positive in all others.<sup>8</sup>

### 2.1.5.6 Post-Hartree-Fock methods

The deficiency of the Hartree-Fock method is the lack of instantaneous correlation of the movements of electrons, which lowers the total electron-electron potential energy. Hence, the exact energy of the molecule will always be lower than that calculated by the Hartree-Fock method.

$$E_{exact} = E_{HF} + E_{correlation} \quad (2.1.63)$$

Three methods that include electron correlation are: (1) configuration interaction (CI), which utilizes a linear combination of Slater determinants to represent the wavefunction, (2) many body perturbation theory, in which a perturbation operator is added to the Hartree-Fock Hamiltonian, and (3) the coupled-cluster (CC) method, wherein the full wavefunction is a product of the ground state SCF wavefunction and an operator  $e^{\hat{T}}$ .

In configuration interaction (CI), the exact wavefunction is a linear combination of determinantal wavefunctions, each of which corresponds to an electronic state of the molecule. In this way, the ground electronic state is represented as a mixture of interacting electronic configurations. (Equation 2.1.64)

$$\Psi_{exact} = c_0\Psi_0 + c_1\Psi_1 + c_2\Psi_2 + \dots \quad (2.1.64)$$

where  $\Psi_0$  is the ground electronic configuration of the Fock operator,  $\Psi_1$  the first electronic excited configuration,  $\Psi_2$  the second, etc., and the coefficients  $c$  represent the

contribution of that particular electronic configuration to the full description of the wavefunction and which are variationally optimized.

In Møller-Plesset many body perturbation theory, the exact Hamiltonian is represented as combination of a zeroth order Hamiltonian  $\hat{H}_0$ , for which molecular orbitals are calculated, plus a parameterized perturbation operator  $\lambda\hat{P}$ , where the parameter  $\lambda$  may vary from 0 to 1 (equation 2.1.65). The perturbation operator corresponds to the potential energy of electron correlation.

$$\hat{H}_{exact} = \hat{H}_0 + \lambda\hat{P} \quad (2.1.65)$$

The exact wavefunction and energy are expressed as a power series of  $\lambda$ . (Equations 2.1.66-67)

$$\Psi_{exact} = \lim_{n \rightarrow \infty} \sum_{i=0}^n \frac{\lambda^i \Psi^{(i)}}{i!} \quad (2.1.66)$$

$$E_{exact} = \lim_{n \rightarrow \infty} \sum_{i=0}^n \frac{\lambda^i E^{(i)}}{i!} \quad (2.1.67)$$

The value of  $i$  determines the order of correction to the zeroth order term. Due to the formalism of this method, electron correlation energy is not estimated until  $n=2$ , corresponding to second order Møller-Plesset perturbation theory (MP2). The wavefunction and energy are thus expressed as a power series to the third term. (Equations 2.1.68-69)

$$\Psi_{MP2} = \Psi^{(0)} + \lambda\Psi^{(1)} + \lambda^2\Psi^{(2)} \quad (2.1.68)$$

$$E_{MP2} = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} \quad (2.1.69)$$

Substitution of these terms into the Schrödinger equation yields equation 2.1.70.

$$\hat{H}_0\Psi^{(2)} + \hat{P}\Psi^{(1)} = E^{(0)}\Psi^{(2)} + E^{(1)}\Psi^{(1)} + E^{(2)}\Psi^{(0)} \quad (2.1.70)$$

Premultiplying by  $\Psi^{(0)}$  and integrating over all space yields the following forms for the energy. (Equations 2.1.71-73)

$$E^{(0)} = \langle \Psi^{(0)} | \hat{H}_0 | \Psi^{(0)} \rangle \quad (2.1.71)$$

$$E^{(1)} = \langle \Psi^{(0)} | \hat{P} | \Psi^{(0)} \rangle \quad (2.1.72)$$

$$E^{(2)} = \langle \Psi^{(0)} | \hat{P} | \Psi^{(1)} \rangle \quad (2.1.73)$$

The sum of  $E^{(0)}$  and  $E^{(1)}$  is the Hartree-Fock energy.  $E^{(2)}$  is the first approximation to the electron correlation energy, expanded to have the form of equation 2.1.74.

$$E^{(2)} = \sum_i^{occ.} \sum_{j>i}^{occ.} \sum_a^{vir.} \sum_{b>a}^{vir.} \frac{\langle \phi_i(1)\phi_j(2) | \frac{1}{r_{12}} | [\phi_a(1)\phi_b(2) - \phi_b(1)\phi_a(2)] \rangle}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j} \quad (2.1.74)$$

where  $i$  and  $j$  are occupied spin orbitals,  $a$  and  $b$  are unoccupied virtual orbitals, and  $\varepsilon$  represents the orbital energy.

The central equation to CC theory is

$$\Psi = e^{\hat{T}} \psi_0 \quad (2.1.75)$$

where  $\Psi$  is the exact ground-state wavefunction,  $\psi_0$  is the ground state HF reference wavefunction.<sup>1</sup> The operator  $e^{\hat{T}}$  is a Maclaurin series

$$e^{\hat{T}} \equiv \sum_{i=0}^{\infty} \frac{\hat{T}^i}{i!} \quad (2.1.76)$$

where  $i$  is summed to the number of electrons of the system and  $\hat{T}^i$  is the particle excitation operator which excites electrons from a filled spin orbital to a virtual spin orbital.<sup>1,5,9</sup>  $\hat{T}$  in equation 2.1.75 is the cluster operator defined as

$$T \equiv \sum_{j=1}^n \hat{T}_j \quad (2.1.77)$$

where  $\hat{T}_j$  creates a linear combination of all possible determinants having  $j$  excitations from the reference HF wavefunction. In practice, the frozen core approximation is typically used such that only valence electron excitations are considered. The subscript  $i$  determines the nomenclature associated with a particular level of CC, e.g.  $j = 1$  corresponds to single excitations and is termed coupled-cluster singles (CCS),  $j = 2$  corresponds to double excitations, termed coupled-cluster doubles (CCD), etc. Combinations of cluster operators yield such methods as CCSD, CCSDT, and so on. The most popular and robust of these methods is CCSD(T), which includes single and double excitations and perturbative inclusion of triple excitations.<sup>10</sup> Its accuracy has made it the so-called gold standard of *ab initio* quantum chemistry.

## 2.2 Composite Thermochemical Model Chemistries

Composite *ab initio* quantum methods aim for highly accurate stationary point energy determinations by combining a series of calculations of ever increasing levels of theory.<sup>5</sup> The initial calculation is a geometry optimization and frequency calculation at a relatively low level of theory. A series of post-Hartree-Fock calculations is then used to recover the effects of basis-set incompleteness and electron correlation, with the assumption made that these two properties can be recovered in an additive fashion. For example, the additivity assumption implies that a Hartree-Fock calculation with a 6-311++G(d,p) basis set can be equivalently accomplished using equation 2.2.1.

$$\begin{aligned}
E[\text{HF}/6\text{-}311\text{++G}(\text{d},\text{p})] &\approx E[\text{HF}/6\text{-}31\text{G}] \\
&+ \{E[\text{HF}/6\text{-}31\text{G}(\text{d},\text{p})] - E[\text{HF}/6\text{-}31\text{G}]\} \\
&+ \{E[\text{HF}/6\text{-}311\text{G}] - E[\text{HF}/6\text{-}31\text{G}]\} \\
&+ \{E[\text{HF}/6\text{-}31\text{++G}] - E[\text{HF}/6\text{-}31\text{G}]\}
\end{aligned}
\tag{2.2.1}$$

The second term of equation 2.2.1 corrects the lack of polarization functions in the 6-31G basis set, while the third and fourth terms correct for triple zeta basis effects and diffuse functions, respectively. This approach is more computationally efficient in that the left hand side of 2.2.1 requires 542 basis sets for  $\text{C}_{11}\text{H}_{26}\text{NO}_2\text{PS}$  and, since Hartree-Fock calculations scales at most quartically with the number of basis sets,  $542^4 \approx 7.54 \times 10^{10}$  calculations. The right hand side of equation 2.2.1, on the other hand, requires  $204^4 + 378^4 + 294^4 + 294^4 \approx 3.71 \times 10^{10}$  calculations, approximately two fifths the computational cost.<sup>5</sup> The  $G_n$  models, with the exception of G4, adopt this additivity assumption.<sup>11,12</sup> The Complete Basis Set (CBS) models, however, do not. Instead, the Hartree-Fock limit is extrapolated using correlation consistent basis sets.<sup>13-15</sup>

The recovery of electron correlation energy is performed through a succession of post HF methods that introduce higher order terms with every calculation. Afterward, empirical energy corrections for spin-orbital coupling, radical species, and relativistic effects are made. Composite thermochemical models achieve kilocalorie mean unsigned errors for heats of formation across a large test set of molecules at a much-reduced computational cost.

### 2.3 Semiempirical Quantum Mechanics

The large computational cost associated with quantum mechanical treatments of chemical systems within the LCAO-MO approximation is the calculation of the many

integrals within the Hartree-Fock equation. Approximating or neglecting certain integrals would allow molecular orbital calculations to be less computationally expensive, making the evaluation of larger and larger systems feasible. This philosophy embodies semiempirical quantum mechanics. Semiempirical theory also parameterizes integrals, which are fitted to either *ab initio* or to experimental values.

Though this methodology has had several incarnations throughout the years, beginning with Huckel theory, which dealt only with conjugated  $\pi$  systems, the focus of this section will be on the neglect of diatomic differential overlap (NDDO) formalism.

### 2.3.1 NDDO

Perhaps the greatest departure from Hartree-Fock (HF) within semiempirical theory is the treatment of core electrons. In it, the core electrons are implicit, which is chemically justifiable in that core electrons are virtually invariant to the chemical environment. The core electrons are subsumed into a kernel that also contains the nucleus, and the nuclear-nuclear electrostatic repulsions found in HF are replaced with core-core repulsion functions. For valence electrons, NDDO uses a minimal basis set of STO's .

The term “differential overlap” concerns the overlap of basis functions that are located on different atomic centers. Within NDDO, the overlap of basis functions is zero if they are centered on different atoms. There are no two-electron three- or four-center terms within NDDO. Hence, the overlap matrix  $S$  is replaced with the identity matrix  $I$  and the Roothaan-Hall equations (equation 2.1.57) become

$$FC = ICE \quad (2.3.1)$$

The two electron integrals

$$(\mu\nu|\lambda\sigma) \quad (2.3.2)$$

are retained only if the basis functions  $\mu\nu$  are centered on the same atomic center and if  $\lambda\sigma$  are also located on the same atomic center, although  $\lambda\sigma$  may be located on an atom different than that of  $\mu\nu$ .

## 2.4 Density Functional Theory (DFT)

Thus far, the calculation of chemical properties has relied upon a description of the wavefunction, which is not a physical observable. In 1964, Hohenberg and Kohn showed that the molecular properties of a system in a non-degenerate electronic ground state can be characterized by its electron probability density  $\rho(r)$ , a physical observable.<sup>16</sup> Because there is a one-to-one mapping between ground state  $\rho(r)$  and the ground state energy, the latter is obtained once the former is found. The primary molecular property of interest in these studies is the energy and nuclear geometry of stationary points on the potential energy surface.

### 2.4.1 Kohn-Sham

Kohn and Sham suggested a way to use the Hohenberg-Kohn theorem by approximating the energy functional  $F[\rho(r)]$  in equation 2.4.1.<sup>17</sup>

$$E[\rho(r)] = \int V_{ext}(r)\rho(r)dr + F[\rho(r)] \quad (2.4.1)$$

where  $E[\rho(r)]$  is the energy of the system and  $V_{ext}(r)$  represents the potential of the nuclei acting upon the electron density, and  $F[\rho(r)]$  is

$$F[\rho(r)] = E_{KE}[\rho(r)] + E_{COUL}[\rho(r)] + E_{XC}[\rho(r)] \quad (2.4.2)$$

$F[\rho(r)]$  represents the the electronic kinetic, Coulombic, and exchange-correlation energy contributions.  $E_{KE}[\rho(r)]$  and  $E_{COUL}[\rho(r)]$  are solved with Kohn-

Sham orbitals in the same fashion as Hartree-Fock, utilizing a Slater-determinant constructed from a set of single-electron spin orbitals solved as a self-consistent field. The exchange-correlation term represents an advantage over Hartree-Fock in that dynamic electron correlation energy is built into the functional. However, the form of this functional is unknown and while many approximations have been developed, there remains no way of systematically improving it.<sup>5</sup> The exchange-correlation term is crucial to the accuracy of density functional theory.<sup>18</sup> Approximations to  $E_{XC}[\rho(r)]$  are discussed below.

#### 2.4.2 Local Density and Local Spin Density Approximations

The local density approximation (LDA) assumes that individual volume elements of the system have the density of a uniform electron gas and that the electron density varies smoothly throughout the entire volume. Within LDA, the  $E_{XC}[\rho(r)]$  functional is represented as the interaction between the electron density  $\rho(r)$  and  $\epsilon_{XC}[\rho(r)]$ , the per-electron exchange-correlation energy, integrated over all space (equation 2.4.3).

$$E_{XC}^{LDA}[\rho(r)] = \int \rho(r)\epsilon_{XC}[\rho(r)]dr \quad (2.4.3)$$

Every point  $r$  is surrounded by a volume element  $dr$  of constant electron density. LDA performs well for metallic systems, but overbinds organic compounds.<sup>17</sup>

The local spin density approximation (LSDA) divides the density into spin-up and spin-down electrons and the net spin density is the difference between the two (equation 2.4.4).<sup>19</sup>

$$\sigma(r) = \rho_{\uparrow}(r) - \rho_{\downarrow}(r) \quad (2.4.4)$$

LSDA works well for bond lengths and vibrations, but overestimates dipole moments and molecular binding energies while underestimating reaction barrier heights.<sup>3,20,21</sup>

### 2.4.3 Generalized and meta-Generalized Gradient Approximation

The next generation of exchange-correlation approximations, the generalized gradient approximation (GGA), considers not only the electron density at a certain point but also its gradient. The gradient term is added to the LSDA per-electron exchange-correlation energy as a correction factor. (equations 2.4.5).

$$\epsilon_{XC}^{GGA}[\rho(r)] = \epsilon_{XC}^{LSDA}[\rho(r)] + \Delta\epsilon_{XC} \left[ \frac{|\nabla\rho(r)|}{\rho^{4/3}(r)} \right] \quad (2.4.5)$$

GGA functionals correct many of the short-comings of the LSDA approximations and yield better thermochemical predictions yet still underestimates reaction barriers.<sup>18</sup> Given the improvement of GGA over LSDA, which is a Taylor-series-like expansion, the logical next step would be the inclusion of second derivative corrections, referred to as meta-GGA functionals. In these, the exchange-correlation potential is dependent upon the kinetic energy density,  $\tau$ , formulated as

$$\tau(r) = \sum_{i=1}^N \frac{1}{2} |\nabla\psi_i(r)|^2 \quad (2.4.6)$$

where  $\psi_i$  is solved from a self-consistent field calculation.

### 2.4.4 Hybrid Functionals

Consider a fictitious device that allows one to adjust smoothly the amount of electron correlation within a uniform electron gas. The exact exchange-correlation energy could then be calculated, using the Hellman-Feynman theorem, as

$$E_{XC} = \int_0^1 \langle \Psi(\lambda) | V_{XC} | \Psi(\lambda) \rangle d\lambda \quad (2.4.7)$$

where  $\lambda$  represents the amount of electronic correlation, zero being no correlation and one being full correlation, and  $V_{XC}$  is the exchange-correlation potential operator.<sup>5</sup> At  $\lambda = 0$ , correlation goes to zero and only exchange interaction remains, which is calculated exactly using Hartree-Fock theory. As  $\lambda$  goes from zero to one, the exchange energy is still that calculated using Hartree-Fock, yet there is an unknown amount of correlation, which DFT seeks to calculate. However, DFT exchange-correlation functionals also calculate the exchange energy for the system. Thus, some fraction of the Hartree-Fock exchange must be subtracted from the DFT exchange-correlation. The total exchange-correlation energy for the fully interacting system is then

$$E_{XC} = (1 - a)E_{XC}^{DFT} + aE_X^{HF} \quad (2.4.8)$$

where  $a$  determines the amount of Hartree-Fock exchange that is incorporated into the functional, usually determined empirically.

Due to the incorporation of non-local Hartree-Fock exchange into the local exchange-correlation of DFT, such functionals are referred to as hybrid functionals. These functionals are often more accurate than purely local functionals for main group thermochemistry.<sup>18</sup>

## 2.5 Molecular Mechanics

Molecular mechanics differs drastically from quantum mechanics and density functional theory. Electrons are not explicitly treated and there is neither a wavefunction nor an electron density for which to solve. Instead, molecules are treated as a collection of atoms, bonds, bond angles, and dihedral angles. Atoms are represented as point

charges. Bonds and bond angles are given equilibrium values; deviation from them results in an increase in energy. Torsion angles are treated as a Fourier series of energetic barriers. Non-bonded atoms interact via Coulombic and Lennard-Jones potentials.

A major drawback of molecular mechanics is that it cannot, due to the implicit treatment of electrons, simulate the making and breaking of bonds. Because a bond is treated as a harmonic potential, significant displacement from its equilibrium distance yields quadratically increasing energy, whereas in reality the bond will break once it reaches a certain length.

### 2.5.1 OPLS-AA

The force field used in the present study is the Optimized Parameters for Liquid Simulations All Atoms (OPLS-AA) force field.<sup>22,23</sup> Within OPLS-AA, the total energy of the system  $E_{tot}$  is calculated as

$$E_{tot} = E_{bonds} + E_{angles} + E_{torsions} + E_{nonbonded} \quad (2.5.1)$$

$$E_{bonds} = \sum_{bonds} K_r (r - r_o)^2 \quad (2.5.2)$$

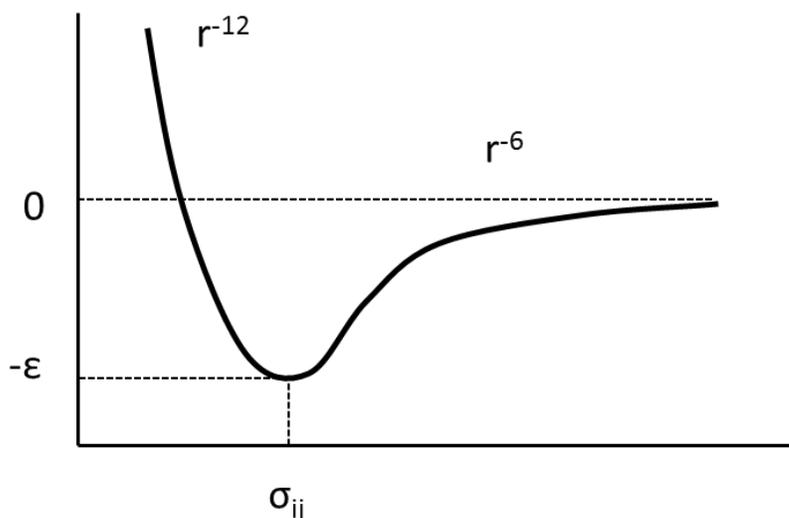
$$E_{angles} = \sum_{angles} K_\theta (\vartheta - \vartheta_0)^2 \quad (2.5.3)$$

$$E_{torsions} = \sum_{torsions} \left[ \frac{V_1}{2} (1 + \cos \varphi) + \frac{V_2}{2} (1 - \cos 2\varphi) + \frac{V_3}{2} (1 + \cos 3\varphi) \right] \quad (2.5.4)$$

$$E_{nonbonded} = \sum_{i < j} \left[ \frac{q_i q_j e^2}{r_{ij}} + 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \right] f_{ij} \quad (2.5.5)$$

where  $r_o$  and  $\vartheta_0$  are the equilibrium bond lengths and angles, respectively,  $K$  is the force constant,  $V$  represents the torsional barrier height, and  $q$  represents the point charge assigned to an atom.<sup>23,24</sup>

Equation 2.5.5 represents the energy from atoms that are not bonded to one another. The  $f_{ij}$  term is 1.0 except in the case of 1,4 intramolecular interactions, where it is 0.5.<sup>25</sup> This damping is necessary so that the same parameters may be used for both intra- and intermolecular interactions. The first term within the brackets in equation 2.5.5 represents the classical Coulombic interaction while the second is the Lennard-Jones potential, which models repulsive (the  $r^{-12}$  term) and attractive van Der Waals (the  $r^{-6}$  term) potentials. The values  $\epsilon$  and  $\sigma$  are more easily expressed by figure 2.3.



**Figure 2.3** Illustration of the Lennard-Jones potential.

The Lennard-Jones potential thus represents a potential energy curve with a well depth of  $-\epsilon$  and an equilibrium radius of  $\sigma$ . The  $\epsilon$ 's and  $\sigma$ 's are first calculated for identical atoms, i.e.  $\sigma_{ii}$ . For interactions between different atoms, the well depth  $-\epsilon_{ij}$  or radius  $\sigma_{ij}$  are calculated via combining rules such that

$$\sigma_{ij} = \sqrt{\sigma_{ii}\sigma_{jj}} \quad (2.5.6)$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}}. \quad (2.5.7)$$

If two or more molecules are included in the calculation, then the energy is a sum of the intermolecular energies and the intramolecular energies (equations 2.5.8-9).

$$E = \sum_{a < b} E_{ab}^{inter} + \sum_a E_a^{intra} \quad (2.5.8)$$

$$E_{ab} = \sum_i^{\in a} \sum_j^{\in b} \left[ \frac{q_i q_j e^2}{r_{ij}} + 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \right] \quad (2.5.9)$$

where  $a$  and  $b$  represent separate molecules and  $i$  and  $j$  represent the atoms within those molecules.

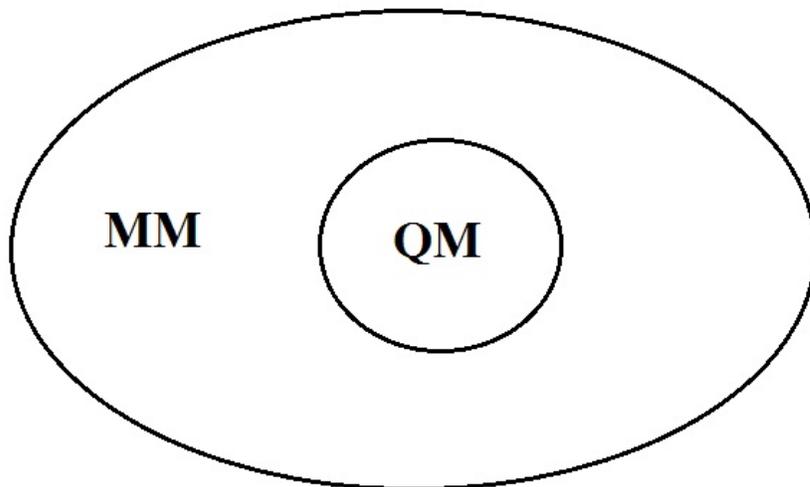
The OPLS-AA force field is parameterized to match experimentally determined solvent properties such as density, free energy of vaporization, and heat capacity. These values are indicative of molecular size and intermolecular interactions. Therefore reproducing them is of particular importance.<sup>26</sup> Because the parameters are found using simulations of hundreds of molecules, many body effects are intrinsically included in the force field. Additionally, the OPLS-AA force field includes the results of accurate quantum mechanical calculations for conformer energetics, such as rotamers, and hydrogen bonding. The OPLS-AA force field has been shown to reproduce experimentally determined properties for a wide range of solvents, organic, and biomolecular systems.<sup>4-12</sup>

## 2.6 Mixed Quantum and Molecular Mechanics Methods (QM/MM)

Atomistic insight into very large systems requires the explicit representation of all atoms. For many systems, such as the condensed phase or proteins, the number of atoms can be on the order of hundreds to tens of thousands. Modeling chemical reactions

within such systems requires the use of quantum mechanics because, as noted above, molecular mechanics cannot simulate bonding making and breaking.

A widely used method is to partition the total space into a non-reactive region, treated with molecular mechanics, and a reactive region, treated with quantum mechanics.



**Figure 2.4** Partitioning scheme used in the QM/MM methodology.

The energy is calculated as the sum of the QM portion, the MM portion, and the interaction between the two.

$$E_{total} = E_{QM} + E_{MM} + E_{QM/MM} \quad (2.6.1)$$

Because of the size of the system and the amount of configuration space that must be sampled, an efficient quantum mechanical method must be used. Within the present study, the QM region is calculated using the semiempirical quantum mechanical NDDO-based AM1 method, while the MM portion is calculated using the OPLS-AA force field.

<sup>5,34,14</sup> The energy of the QM/MM region is calculated similarly to the non-bonded energetics of OPLS-AA. (equation 2.6.2)

$$E_{QM/MM} = \sum_i^{solute} \sum_j^{solvent} \left[ \frac{\alpha q_i^{CMX} q_j e^2}{r_{ij}} + 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \right] \quad (2.6.2)$$

In order to determine the atomic charges on the solute, the QM electrostatic potential is mapped onto atoms using either the Charge Model 1 (CM1) or Charge Model 3 (CM3) method.<sup>36,37,5</sup> Because CM1 and CM3 were designed to reproduce gas phase dipole moments, the charges are scaled by a parameter  $\alpha$  when used in solution in order to reflect the increase in condensed phase dipole moments. The optimal  $\alpha$  was found by fitting it to reproduce experimental free energies of hydration, yielding  $\alpha = 1.14$  when using CM1 and  $\alpha = 1.15$  when using CM3. These factors result in errors on the order of 1 kcal·mol<sup>-1</sup>, with CM1 best reproducing molecules with the biochemically significant amide functionality.<sup>38</sup>

## 2.7 Thermodynamic quantities

Free energy is arguably the most important thermodynamic property.<sup>39</sup> While HF, DFT, and molecular mechanics yield the electronic energies, to obtain free energies the thermal contributions from molecular vibrations, rotations, and translations must be calculated. Additionally, free energy is not a property of a single molecule, but of an ensemble of molecules. If any comparison between the computational and experimental determination of physical observables is to be attained, the microscopic must be connected to the macroscopic. Statistical thermodynamics provides such a connection.

The fundamental value to be calculated is the partition function,  $Q$ , a dimensionless quantity which is a ratio of the total number of molecules in the system to the number of molecules in the ground state.<sup>40,41</sup> Once  $Q$  is determined, all

thermodynamic properties of the system may be obtained. The partition function is calculated as a sum over all microstates  $j$  such that

$$Q = \sum_{j=1}^N \exp(-\beta \epsilon_j) \quad (2.7.1)$$

where  $N$  is the possible number of states,  $\beta$  is  $(k_b T)^{-1}$  where  $k_b$  is the Boltzmann constant and  $T$  is the absolute temperature, and  $\epsilon_j$  is the energy of microstate  $j$ .

### 2.7.1 Molecular Thermodynamics

Within the scope of QM calculations on a single molecule, the partition function is reduced to the molecular partition function  $q$ . Because thermodynamic values are reported in molar quantities, the total  $Q$  is  $q^{N_A}$  where  $N_A$  is Avogadro's number. Many idealizations of the system are made, namely, ideal gas, harmonic oscillator, rigid rotor, and the separability of these components. The molecular partition function is then a product of electronic, vibrational, rotational, and translational partition functions (equation 2.7.2). The total energy is the sum of the individually calculated electronic, vibrational, rotational, and translational energies.<sup>5</sup> (equation 2.7.3)

$$q_{tot} = q_{elec} q_{vib} q_{rot} q_{trans} \quad (2.7.2)$$

$$\epsilon_{tot} = \epsilon_{elec} + \epsilon_{vib} + \epsilon_{rot} + \epsilon_{trans} \quad (2.7.3)$$

Acquiring the electronic energy has been discussed in detail above. Because the first electronic excited state is typically much larger than  $k_b T$ , it is assumed that only the electronic ground state is accessible. This reduces  $q_{elec}$  to the electronic spin multiplicity of the molecule. (This assumption is made within the software package, Gaussian 09, used in this research.<sup>42,43</sup>)

The partition functions for the translational, rotational, and vibrational contributions to the molecular partition function each have a unique form. For translation, it is

$$q_{trans} = \left( \frac{2\pi m k_B T}{h^2} \right)^{\frac{3}{2}} V \quad (2.7.4)$$

where  $m$  is the mass of the molecule,  $h$  is Planck's constant, and  $V$  is the volume.

For rotation, it is

$$q_{rot} = \left( \frac{\pi^{\frac{1}{2}}}{\sigma} \right) \left( \frac{2I_X k_B T}{h^2} \right) \left( \frac{2I_Y k_B T}{h^2} \right) \left( \frac{2I_Z k_B T}{h^2} \right) \quad (2.7.5)$$

where  $I$  is the moment of inertia and  $\sigma$  is the molecular symmetry number.

For vibration, it has the form

$$q_{vib} = \prod_i^{modes} \frac{1}{1 - \exp(-h\omega_i/k_B T)} \quad (2.7.6)$$

where  $\omega = (k/\mu)^{\frac{1}{2}}$ , in which  $k$  is the force constant and  $\mu$  the reduced mass. This form of  $q_{vib}$  is measured relative to the zero-point energy  $U_0$ , which is the electronic energy plus the vibrational energy at zero Kelvin. (Equation 2.7.7)

$$U_0 = E_{elec} + \sum_{i=1}^{modes} \frac{1}{2} h\omega_i \quad (2.7.7)$$

From equations 2.7.4-6, thermodynamic values such as internal energy  $U$  (equation 2.7.8), enthalpy  $H$  (equation 2.7.9), and Gibbs free energy  $G$  (equation 2.7.10) are obtained.<sup>5,40,41,44</sup>

$$U = \frac{k_B T^2}{Q} \left( \frac{\partial Q}{\partial T} \right)_V \quad (2.7.8)$$

$$H = k_B T^2 \left( \frac{\partial \ln Q}{\partial T} \right)_V + k_B T V \left( \frac{\partial \ln Q}{\partial V} \right)_T \quad (2.7.9)$$

$$G = -k_B T \ln Q + k_B T V \left( \frac{\partial \ln Q}{\partial V} \right)_T \quad (2.7.10)$$

## 2.7.2 Many Body Thermodynamics

Monte Carlo is a numerical method for finding solutions to multi-dimensional integrals through random sampling of the integration variables.<sup>45</sup> In the context of molecular simulations, these variables are the momenta and positions of every molecule in the ensemble. For such systems, the key problem is solving the configurational integral when calculating the average of some property  $A$ . The research performed in this work simulates the isobaric-isothermal (NPT) ensemble, the corresponding equations for which are

$$\langle A \rangle = A_k + \iint A(X, V) P(X, V) dX dV \quad (2.7.11)$$

$$P(X, V) = \exp(-\beta H(X, V)) \iint \exp(-\beta H(X, V)) dX dV \quad (2.7.12)$$

where  $\langle A \rangle$  denotes the average of property  $A$ ,  $X$  represents a configuration which is sampled,  $V$  is the volume,  $P(X, V)$  is the Boltzmann factor, and  $H(X, V)$  is the enthalpy. The integrals are taken over all possible configurations and volumes of the system.<sup>46</sup>  $A_k$  in equation 2.7.5 represents the kinetic energy contribution to the average. When calculating the differences in average Gibbs free energy, however, kinetic energy contributions were shown to cancel by W. van Gunsteren, so it is unnecessary to determine this value for Monte Carlo free energy difference calculations.<sup>39</sup>

Simple Monte Carlo, which integrates over completely random configurations, samples spaces of low Boltzmann probability that contribute little to the average, such as those of overlapping particles, and then weights them according to  $\exp(-H/k_B T)$ .<sup>47,48</sup> Such a technique is impractical for large systems. Metropolis et al. modified this procedure to choose only those configurations with probability  $\exp(-H/k_B T)$  and then weight them evenly.<sup>49</sup> This algorithm creates a Markov chain of configurations. When the system in configuration  $i$  attempts to move to configuration  $j$ , the move is either rejected or accepted. The probability  $p = \pi_j/\pi_i$  determines the acceptability of the new configuration, where

$$\pi = \exp(-\beta H) V^N \quad (2.7.13)$$

If  $p \geq 1$ , denoting  $H_j \leq H_i$ , the move is accepted. For the case of  $p < 1$ , denoting an increase in  $H$ ,  $p$  is compared to a random number  $x$  between zero and one. If  $p \geq x$ , the move is accepted. If not, the move is rejected and configuration  $i$  is repeated. Randomly accepting moves of greater enthalpy improves the sampling of configuration space by allowing the algorithm to escape from local minima. High energy barriers, however, are extremely unlikely to be overcome, which may lead to a metastable system.<sup>46,48</sup>

The Metropolis algorithm and the neglect of the kinetic energy term simplifies equation 2.7.4 to

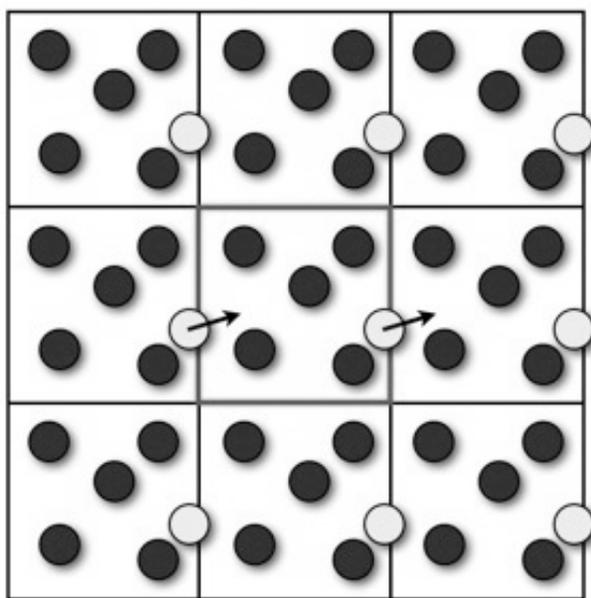
$$\langle A \rangle = \frac{1}{L} \sum_{i=1}^L A_i(X, V) \quad (2.7.14)$$

where  $L$  is the number of considered configurations. Adequate sampling requires configurational changes leading to an acceptance rate of 30 to 50 percent. If fewer are

accepted, the parameters for molecular translations, rotations, or volume movements are too high, creating forbidden movements. If greater, the parameters are too small and the configurations in the Markov chain will not change enough. Either case leads to a Markov chain unrepresentative of a physical system, resulting in a failure to obtain the necessary Boltzmann-weighted configurations for equation 2.7.14 to yield accurate average properties.

A simulation involving a solute and hundreds of solvent atoms can employ preferential sampling, in which solvent atoms closest to the solute are moved more frequently.<sup>50</sup> This allows nearby solvent molecules to respond to movements of the solute. The Monte Carlo simulations in this research perform a solvent movement ninety nine times out of a hundred configurations, allowing the proximal solvent to respond fully to solute perturbations.

A second improvement contributed by Metropolis et al. is the use of periodic boundary conditions.<sup>49</sup> The simulated system composed of  $N$  species is surrounding in three dimensions by identical images, removing surface effects and providing a better approximation to bulk phase space. Should a molecule move out of the central cell, it reappears on the opposite side, so only the contents of the central cell need be tracked.



**Figure 2.5** Periodic Boundary Conditions. If a molecule exits one side of the cell, it reappears on the other side.

When evaluating the intermolecular potentials, spherical cutoff radii allow the avoidance of interactions of one molecule with more than one image of another molecule. A cutoff equal to or less than one half the length of the cell is chosen to avoid self-interactions.<sup>46,47</sup> For long-range interactions neglected beyond the cutoff, full periodic boundary conditions are implemented through the Ewald method (*vide infra*) or by embedding the cutoff sphere within a polarizable continuum.<sup>46</sup>

### 2.7.3 Free Energy Perturbation

Relative energy differences between chemically related systems may be of greater importance than the absolute energy of the independent systems. Free energy differences determine the system's thermal stability, chemical composition, rate of reaction, and reaction pathway. Calculating free energy differences between related chemical systems, e.g. the products and reactants of a reaction or conformational isomers, is performed by

free energy perturbation (FEP). For two defined states  $X$  and  $Y$  the Gibb's free energy difference is

$$\Delta G = G_Y - G_X = -k_B T \ln \left( \frac{Q_X}{Q_Y} \right) \quad (2.7.15)$$

Zanzwig used equation 2.7.15 to derive the central formula to free energy perturbation, equation 2.7.16.<sup>51</sup>

$$\Delta G = -k_B T \ln \left\langle \exp \left( \frac{-(\mathcal{H}_X - \mathcal{H}_Y)}{k_B T} \right) \right\rangle_Y \quad (2.7.16)$$

where  $\mathcal{H}$  is the Hamiltonian of the system, brackets denotes the ensemble average, and the subscript  $Y$  denotes that the average has been taken over the ensemble in state  $Y$ . If the phase space of states  $X$  and  $Y$  do not overlap, simulations using equation 2.7.16 will not converge to an accurate energy.<sup>39,3</sup> The energy between the two states must not be much larger than  $k_B T$ .<sup>3</sup> To avoid this issue, the Hamiltonian is reformulated so that it depends upon a parameter  $\lambda$ . (equation 2.7.17)

$$\mathcal{H}(\lambda) = \lambda \mathcal{H}_Y + (1 - \lambda) \mathcal{H}_X \quad (2.7.17)$$

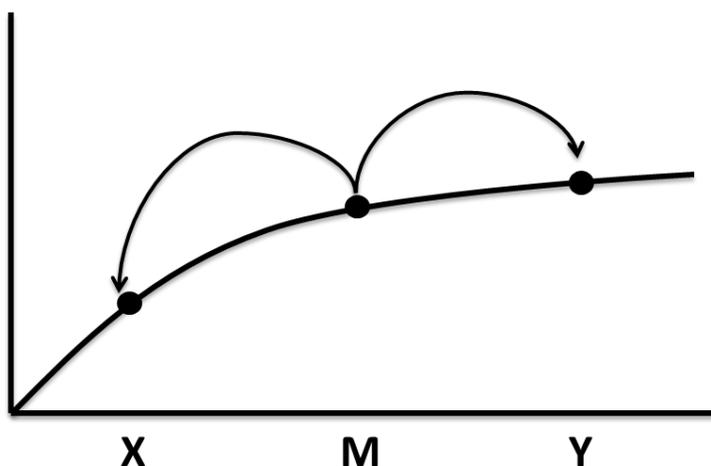
Equation 2.7.16 becomes

$$\Delta G = \sum_{\lambda=0}^1 -k_B T \ln \left\langle \exp \left( \frac{-\Delta \mathcal{H}'}{k_B T} \right) \right\rangle_Y \quad (2.7.18)$$

where  $\Delta \mathcal{H}' = \mathcal{H}_{\lambda+d\lambda} - \mathcal{H}_\lambda$ . Thus, the perturbation between states is divided into as many windows as necessary to guarantee sufficient overlap. In this research, the method of double-wide sampling is performed.<sup>48, 32</sup> Consider states  $X$ , for which  $\lambda = 0$ ,  $Y$ , for which  $\lambda = 1$ , and a middle state  $M$ , for which  $\lambda = 0.5$ . Instead of perturbing the system from state  $X$  to  $Y$  directly, the system begins at state  $M$  and samples to  $X$  and  $Y$  simultaneously, illustrated in figure 2.6.

Consequently, only half the number of simulations needs to be run. Calculating  $\Delta G$  using double-wide sampling requires a modification to equation 2.7.16 to account for the simultaneous perturbations from the initial reference state to two others (equation 2.7.19).<sup>48</sup>

$$\Delta G = k_B T \ln \left( \frac{\langle \exp(-(E_0 - E_M)/k_B T) \rangle_M}{\langle \exp(-(E_1 - E_M)/k_B T) \rangle_M} \right) \quad (2.7.19)$$



**Figure 2.6** Double-wide sampling. The system begins in state **M** and is perturbed to states **X** and **Y** simultaneously.

In the context of organic reactions, states *X* and *Y* correspond to the reactants and products, respectively. Hence, the perturbation is a function of an intra- or intermolecular coordinate, e.g. the carbon-bromine radius of an E1 type reaction or a dihedral rotation. The free energy surface produced by such a calculation is referred to as a potential of mean force.

## 2.8 The Ewald Summation

Full periodic boundary conditions must be applied for systems such as room temperature ionic liquids due to the prevalent long range electrostatics.<sup>52</sup> The Ewald method calculates the exact electrostatic energy of a fully periodic infinite lattice.<sup>3</sup> Consider a lattice with a cubic central cell of length  $L$  and with periodic cells having coordinate vectors  $\mathbf{n} = (n_x L, n_y L, n_z L)$  where  $n$  is an integer. The Coulombic potential from all interactions is

$$V = \frac{1}{2} \sum_{|\mathbf{n}|=0} \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{4\pi\epsilon_o |\mathbf{r}_{ij} + \mathbf{n}|}. \quad (2.8.1)$$

(Note that in equation 2.8.1 the series omits  $i = j$  for  $\mathbf{n} = 0$ .) Unfortunately, equation 2.8.1 diverges for large systems. Ewald solved this by replacing the single diverging series with two more rapidly convergent series by manipulating the  $r^{-1}$  term. (Equation 2.8.2)

$$\frac{1}{r} = \text{erf}\left(\frac{\frac{1}{2}\sqrt{\alpha}r}{r}\right) + \text{erfc}\left(\frac{\frac{1}{2}\sqrt{\alpha}r}{r}\right), \quad (2.8.2)$$

where  $\text{erf}(x)$  and  $\text{erfc}(x)$  are the error function (equation 2.8.3) and its complement (equation 2.8.4), respectively, and the parameter  $\alpha$  controls the width of the Gaussian charge distribution  $\rho(\mathbf{r})$  (equation 2.8.5) discussed below.

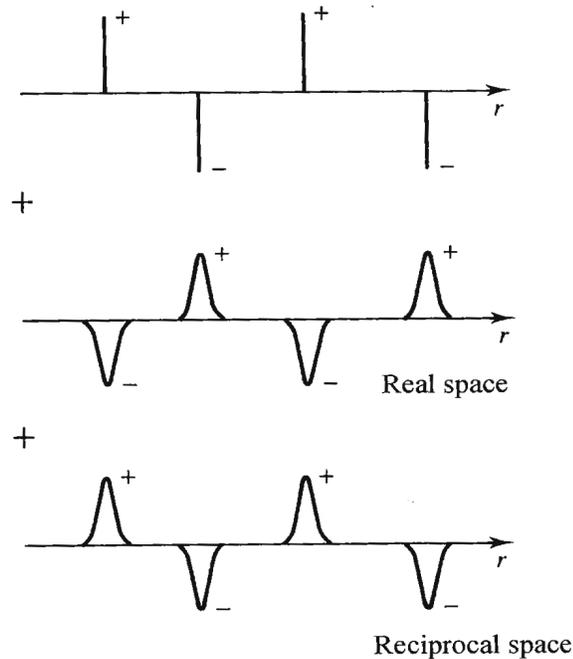
$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (2.8.3)$$

$$\text{erfc}(x) = 1 - \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt \quad (2.8.4)$$

$$\rho_i(\mathbf{r}) = \frac{q_i \alpha^3}{\pi^{3/2}} \exp(-\alpha^2 r^2) \quad (2.8.5)$$

The first term of equation 2.8.2 goes to a constant ( $\sqrt{\alpha/\pi}$ ) as  $r \rightarrow 0$  and handles the rapid variation of equation 2.8.1 at small  $r$ . However, it has a long tail as  $r \rightarrow \infty$ . The second term approaches singularity as  $r \rightarrow 0$ , but decays exponentially as  $r \rightarrow \infty$ . The resulting sum is a well-behaved convergent series.

The point charges are formulated as a combination of screened charges minus a smoothly varying screening background.<sup>53</sup> First, equation 2.8.1 becomes a sum of point charges and neutralizing Gaussian distributions (equation 2.8.5), shown as the first sum in figure 2.7.



**Figure 2.7.** The Ewald summation in diagrammatic form. Point charges are combined with charge screening Gaussian charge distributions and their reciprocals.

The potential given by this “real space” summation is

$$V_{real} = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \sum_{|\mathbf{n}|=0} \frac{q_i q_j}{4\pi\epsilon_0} \frac{\text{erfc}(\alpha|\mathbf{r}_{ij} + \mathbf{n}|)}{|\mathbf{r}_{ij} + \mathbf{n}|}, \quad (2.8.6)$$

where  $\alpha$  controls the width of the Gaussian function. The second “reciprocal space” sum of Gaussians counteracts the first neutralizing distribution. (Equation 2.8.7)

$$V_{reciprocal} = \frac{1}{2} \sum_{\mathbf{k} \neq 0} \sum_{i=1}^N \sum_{j=1}^N \frac{1}{\pi L^3} \frac{q_i q_j}{4\pi\epsilon_0} \frac{4\pi^2}{k^2} \exp\left(-\frac{k^2}{4\alpha^2}\right) \cos(\mathbf{k} \cdot \mathbf{r}_{ij}) \quad (2.8.7)$$

where the reciprocal vectors  $\mathbf{k}$  are given as  $\mathbf{k} = 2\pi\mathbf{n}/L$ . Adding the reciprocal space introduces interactions of the point charges with their own charge distribution, i.e. self-interaction. This potential (equation 2.8.8) must then be subtracted to find the total potential (equation 2.8.9).<sup>53</sup>

$$V_{self} = -\frac{\alpha}{\sqrt{\pi}} \sum_{k=1}^N \frac{q_k^2}{4\pi\epsilon_0} \quad (2.8.8)$$

$$V_{total} = V_{real} + V_{reciprocal} - V_{self} \quad (2.8.9)$$

The choice of  $\alpha$  is critical. Real space summation converges more rapidly with larger  $\alpha$  whereas reciprocal space summation converges more rapidly with smaller  $\alpha$ . Also,  $\alpha$  must be chosen such that, if radial cutoffs distances are employed within the simulation, only charges within the cutoff interact. The software used in this research, the Biochemical and Organic Simulation System (BOSS), automatically sets  $\alpha$  such that the value of the reciprocal sum at the cutoff is  $1.0 \times 10^{-6}$  kcal/mol.<sup>54</sup>

## 2.9 References

- (1) Levine, I. N. *Quantum Chemistry*; 6th ed.; Pearson Education Ltd.: London, 2009.
- (2) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications, Inc.: New York, 1996.
- (3) Leach, A. R. *Molecular Modelling: Principles and Applications*; 2nd ed.; Pearson Education Limited: Essex, 2001.
- (4) Slater, J. *Physical Review* **1930**, *36*, 57–64.
- (5) Cramer, C. J. *Essentials of Computational Chemistry*; 2nd ed.; John Wiley and Sons Ltd.: West Sussex, England, 2004.
- (6) Boys, S. F. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **1950**, *200*, 542–554.
- (7) Dunning, T. H. *The Journal of Chemical Physics* **1989**, *90*, 1007.
- (8) Schlegel, H. B. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 790–809.
- (9) Jensen, F. *Introduction to Computational Chemistry*; 2nd ed.; John Wiley and Sons Ltd.: West Sussex, England, 2007.
- (10) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chemical Physics Letters* **1989**, *157*, 479–483.

- (11) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 810–825.
- (12) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *The Journal of Chemical Physics* **2007**, *126*, 084108.
- (13) Peterson, G. A. In *Computational Thermochemistry: Prediction and Estimation of Molecular Thermodynamics*; Irikura, K. K.; Frurip, D. J., Eds.; American Chemical Society: Washington, DC, 1998; pp. 237–266.
- (14) Ochterski, J. W.; Petersson, G. A.; Montgomery, J. A. *The Journal of Chemical Physics* **1996**, *104*, 2598.
- (15) Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *The Journal of Chemical Physics* **2000**, *112*, 6532.
- (16) Hohenberg, P. *Physical Review* **1964**, *136*, B864–B871.
- (17) Kohn, W.; Sham, L. J. *Physical Review* **1965**, *140*, A1133–A1138.
- (18) Zhao, Y.; Truhlar, D. G. *Accounts of Chemical Research* **2008**, *41*, 157–67.
- (19) Kohn, W.; Becke, A. D.; Parr, R. G. *The Journal of Physical Chemistry* **1996**, *100*, 12974–12980.
- (20) Becke, A. *Physical Review A* **1986**, *33*, 2786–2788.

- (21) Johnson, B. G.; Gill, P. M. W.; Pople, J. a. *The Journal of Chemical Physics* **1993**, *98*, 5612.
- (22) Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. *Journal of the American Chemical Society* **1984**, *106*, 6638–6646.
- (23) Jorgensen, W. L.; Tirado-Rives, J. *Journal of the American Chemical Society* **1988**, *110*, 1657–1666.
- (24) Jorgensen, W. L.; Tirado-Rives, J. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 6665–70.
- (25) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *Journal of the American Chemical Society* **1996**, *118*, 11225–11236.
- (26) Jorgensen, W. L.; Tirado-Rives, J.; Fields, F. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 6665–70.
- (27) Jorgensen, W. L.; Ulmschneider, J. P.; Tirado-Rives, J. *The Journal of Physical Chemistry B* **2004**, *108*, 16264–16270.
- (28) Rizzo, R. C.; Jorgensen, W. L. *Journal of the American Chemical Society* **1999**, *121*, 4827–4836.
- (29) Damm, W.; Frontera, A.; Tirado-Rives, J.; Jorgensen, W. L. *Journal of Computational Chemistry* **1997**, *18*, 1955–1970.

- (30) Jorgensen, W. L.; Severance, D. L. *Journal of the American Chemical Society* **1990**, *112*, 4768–4774.
- (31) Jorgensen, W. L.; McDonald, N. A. *Journal of Molecular Structure: THEOCHEM* **1998**, *424*, 145–155.
- (32) Jensen, K. P.; Jorgensen, W. L. *Journal of Chemical Theory and Computation* **2006**, *2*, 1499–1509.
- (33) McDonald, N. A.; Jorgensen, W. L. *The Journal of Physical Chemistry B* **1998**, *102*, 8049–8059.
- (34) Dewar, M.; Zoebisch, E.; Healy, E. *Journal of the American* **1985**, 3902–3909.
- (35) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926.
- (36) Winget, P.; Thompson, J. D.; Xidos, J. D.; Cramer, C. J.; Truhlar, D. G. *The Journal of Physical Chemistry A* **2002**, *106*, 10707–10717.
- (37) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *Theoretical Chemistry Accounts* **2005**, *113*, 133–151.
- (38) Jorgensen, W. L.; Box, P. O.; Haven, N.; Udier-Blagović, M.; Morales De Tirado, P.; Pearlman, S. a *Journal of computational chemistry* **2004**, *25*, 1322–32.
- (39) Kollman, P. *Chemical Reviews* **1993**, *93*, 2395–2417.

- (40) McQuarrie, D. A.; Simon, J. D. *Molecular Thermodynamics*; University Science Books: USA, 1999.
- (41) Atkins, P. W. *Physical Chemistry*; 2nd ed.; W. H. Freeman and Company: San Francisco, 1982.
- (42) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. S.; M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. M.; G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. H.; A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. H.; M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. N.; Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. M. J.; J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. B.; K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. N.; K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. T.; M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. C.; V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. S.; O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. O.; R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. V.; P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. D.; O. Farkas, J. B. Foresman, J. V. Ortiz, J. C.; Fox, D. J. F. *Gaussian 09, Revision B.01*; Gaussian, Inc.: Wallingford, CT, 2010.
- (43) Ochterski, J. W. Thermochemistry in Gaussian  
[http://www.gaussian.com/g\\_whitepap/thermo.htm](http://www.gaussian.com/g_whitepap/thermo.htm).
- (44) Chandler, D. *Introduction to Modern Statistical Mechanics*; Oxford University Press: New York, 1987.

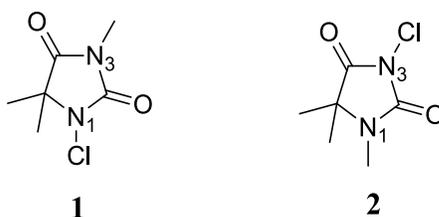
- (45) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes*; Cambridge University Press: Cambridge, 1986.
- (46) Jorgensen, W. L. In *Encyclopedia of Computational Chemistry*; von Rague Shleyer, P., Ed.; John Wiley and Sons Ltd., 1998; pp. 1774–1763.
- (47) Frenkel, D. In *Computational Soft Matter: From Synthetic Polymers to Proteins*; Attig, N.; Binder, K.; Grubmuller, H.; Kremer, K., Eds.; John von Neumann Institute for Computing: Julich, 2004; Vol. 23, pp. 29–59.
- (48) Jorgensen, W. L.; Thomas, L. L. *Journal of chemical theory and computation* **2008**, *4*, 869–876.
- (49) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *The Journal of Chemical Physics* **1953**, *21*, 1087.
- (50) Jorgensen, W. L.; Ravimohan, C. *The Journal of Chemical Physics* **1985**, *83*, 3050.
- (51) Zwanzig, R. *The Journal of Chemical Physics* **1954**, *22*, 1420.
- (52) Ewald, P. *Annalen der Physik* **1921**, *64*, 253–287.
- (53) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press, Inc.: San Diego, 1996.

- (54) Jorgensen, W. L.; Tirado-Rives, J. *Journal of Computational Chemistry* **2005**, *26*, 1689–700.

## Chapter 3 Inter- and Intra-molecular Mechanisms for Chlorine Rearrangements in Trimethyl-substituted *N*-chlorohydantoins

### 3.1 Abstract

The antimicrobial compounds 1-chloro-3,5,5-trimethylhydantoin and 3-chloro-1,5,5-trimethylhydantoin (**1** and **2**, respectively) have been synthesized and examined via a joint experimental and computational study (Figure 3.1).



**Figure 3.1** Structures of 1-chloro-3,5,5-trimethylhydantoin (**1**) and 3-chloro-1,5,5-trimethylhydantoin (**2**).

The measured rate of loss of oxidative chlorine in the absence of and presence of exposure to UVA irradiation determined **2** to be less stable than **1**.

Reprinted with permission from *J. Phys. Chem. A*, 2012, 116 (26), pp 7245–7252.  
Copyright 2012 American Chemical Society.

Two novel hydrogen atom transfer reaction (HATR) mechanisms have been proposed: (1) an intramolecular process in which a hydrogen atom undergoes a series of sigmatropic shifts, and (2) an intermolecular pathway in which a radical abstracts a hydrogen atom from a neighboring molecule. Density functional theory (DFT) calculations at the UB3LYP/6-311G++(2d,p) theory level have been employed to elucidate the preferred reaction pathway. Both proposed HATR mechanisms predicted **2** to possess a lower free energy of activation,  $\Delta G^\ddagger$ , relative to **1** in accordance with the experimental stability measurements. However, the intermolecular route had an overall lower absolute  $\Delta G^\ddagger$  and was more consistent with measured product ratios in solution. The intermolecular reaction pathway, unlike the intramolecular route, also predicted the lack of formation of a migration product featuring a Cl covalently bonded to a methylene group at the 5-position of the hydantoin moiety, which was verified by NMR experiments.

### 3.2 Computational Methods

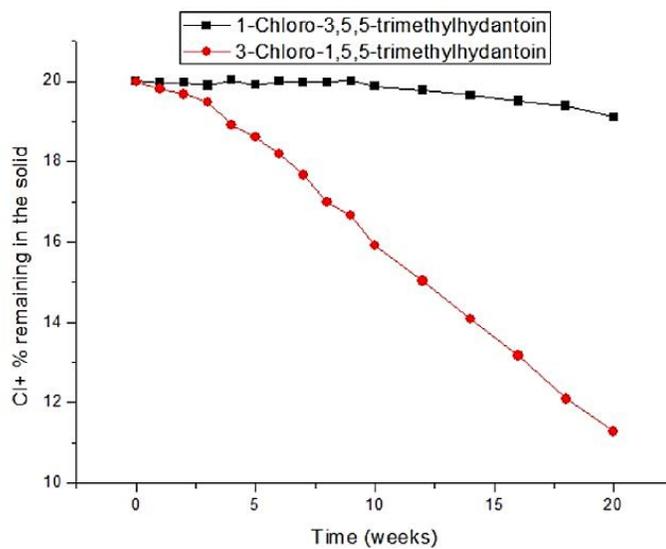
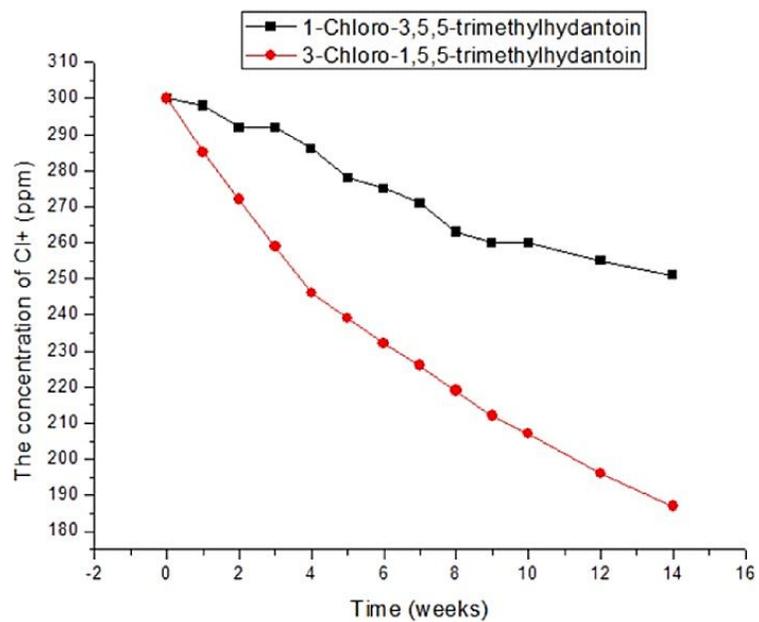
DFT calculations were used to characterize ground states and transition structures at both the restricted and unrestricted (U)B3LYP<sup>1,2</sup> level of theory in conjunction with the 6-311++G(2d,p) basis set. Geometry optimizations were performed for each structure, and frequency calculations were performed in order to verify all stationary points as minima for ground states or a saddle points for transition structures. All calculations were performed for species in vacuum using Gaussian 09<sup>3</sup> on computers located at the Alabama Supercomputer Center.

### 3.3 Results and Discussion

#### 3.3.1 Heterolytic Dissociation

In the absence of UV irradiation, whether in aqueous solution or in the solid state, *N*-halamine compounds dissociate primarily through loss of oxidative halogen, e.g., Cl<sup>+</sup>, to form their un-halogenated precursor amine compounds. This process is illustrated for compounds **1** and **2** in Figure 3.2.

It is evident that compound **1** is much more stable toward chlorine loss than is compound **2**, especially in the solid state. Intuitively, this was as expected because the resulting negative charge on nitrogen should be stabilized inductively and through resonance by two carbonyl moieties for the imide nitrogen in **2**, but only by one for the amide nitrogen in **1**. The enhanced rate of Cl<sup>+</sup> decomposition for **2** as compared to **1** is in accord with recent calculations of base-induced de-chlorinations of 3-chloro-5,5-dimethylhydantoin and 1-chloro-5,5-dimethylhydantoin to yield a hydantoin anion and Cl<sup>+</sup> (i.e., HOCl).<sup>4</sup> In that mechanistic study, B3LYP/6-311+G(2d,p) predicted de-chlorination  $\Delta G^\ddagger$  values of 37.8 and 47.7 kcal/mol for the 3-chloro versus 1-chloro derivatives, respectively,<sup>4</sup> which is indicative of a more facile scission of the imide-Cl bond and consistent with the current results.

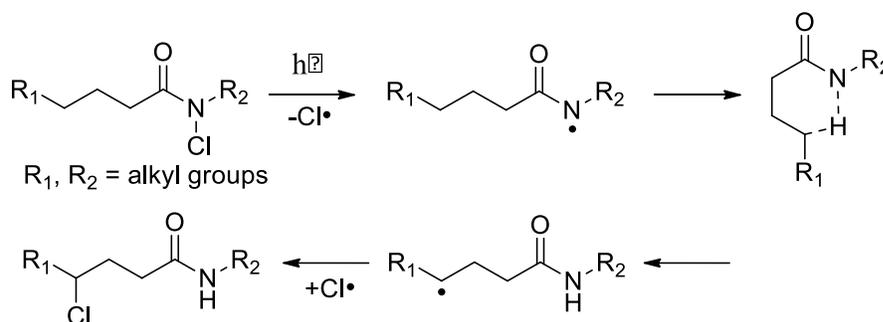


**Figure 3.2** Oxidative chlorine loss in water (above) and for the solid state (below) over time for 1-chloro-3,5,5-trimethylhydantoin (1) and 3-chloro-1,5,5-trimethylhydantoin (2).

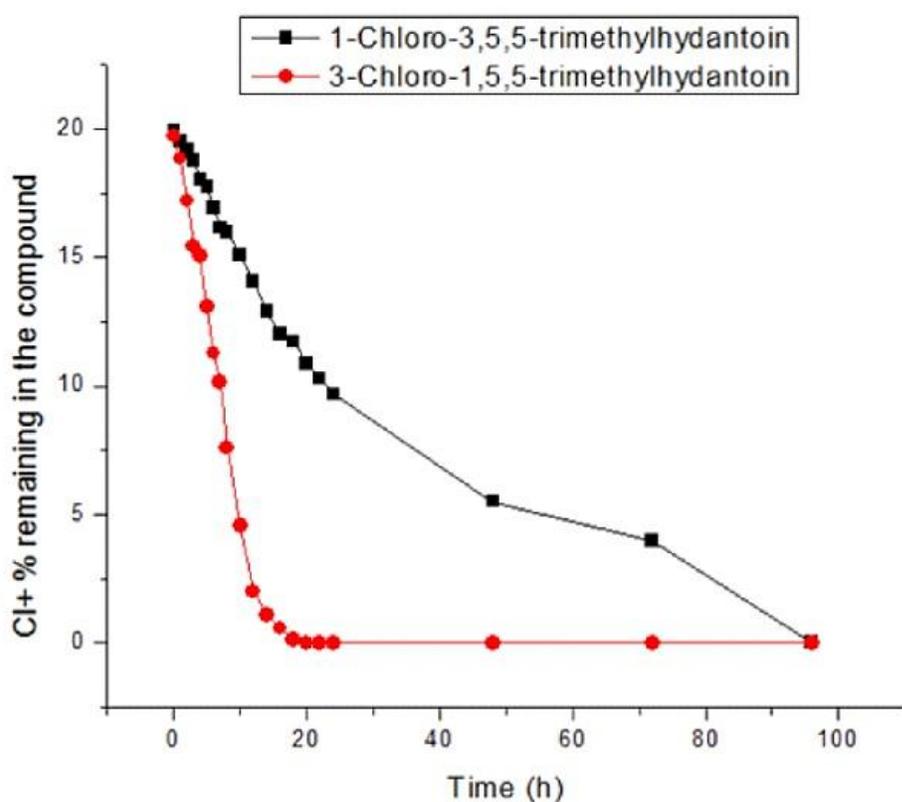
### 3.3.2 Homolytic Dissociation under UVA Irradiation

In the presence of UV irradiation, *N*-halamines dissociate primarily homolytically to form halogen radicals and radical centers on nitrogen atoms. Following this process, products in addition to the precursor amines can form. For example, in a recent study in these laboratories it was shown that 1-chlorohydantoinyl moieties bonded through the imide nitrogen via a three-carbon chain and subsequently through a siloxane moiety to cellulose dissociated from the surface under UVA irradiation.<sup>5</sup> It was suggested that the dissociation from the surface was due to scission of a C-Si bond induced by migration of the amide chlorine to the propyl side chain through a free-radical process via a Hoffmann-Loeffler-type mechanism (Scheme 3.1). DFT computations on the model compound 3-butyl-1-chloro-5,5-dimethylhydantoin supported the experimental observation (NMR and mass spectral data) that the Cl atom migrated to the  $\beta$  and  $\gamma$  carbons of the alkyl chain, but not to the  $\alpha$  carbon.<sup>5</sup> The transition structure for a similar migration to the  $\alpha$  carbon was too strained; thus no product containing the Cl at the  $\alpha$  carbon (the carbon bonded directly to the imide nitrogen of the hydantoin ring) was observed. Therefore, it is apparent that a different mechanism must be proposed for any photoinduced chlorine migrations in compounds **1** and **2**.

**Scheme 3.1** Intramolecular photorearrangement of acyclic *N*-halamines.



While **1** was more stable toward  $\text{Cl}^+$  dissociation than was **2** for the compounds not directly exposed to UVA, it was possible that this would not be the case for homolytic cleavage in the presence of UVA irradiation. However, Figure 3.3 clearly demonstrates that the order of stabilities was the same for the two types of processes.

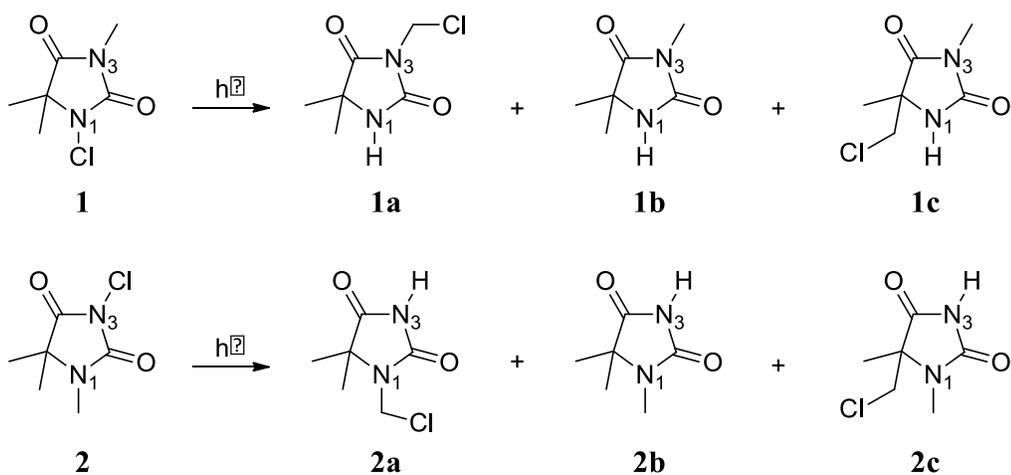


**Figure 3.3** UVA stability of solid-state 1-chloro-3,5,5-trimethylhydantoin (1) and 3-chloro-1,5,5-trimethyl hydantoin (2).

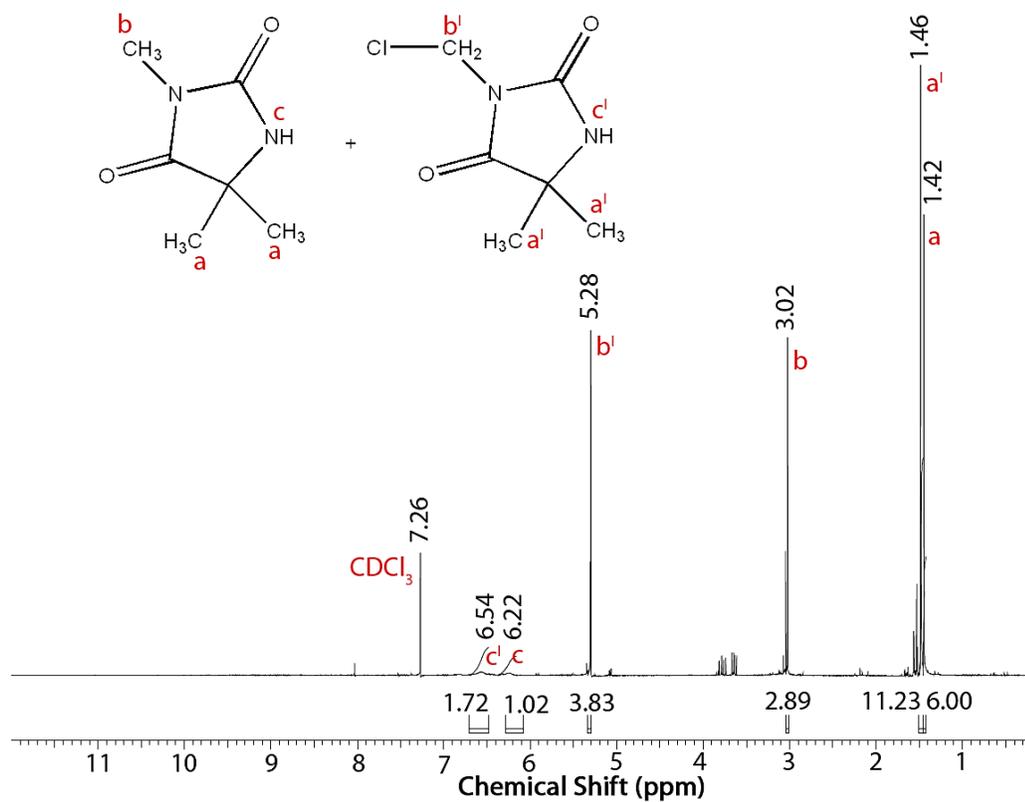
For the data presented in Figure 3.3, solid-state samples were titrated for remaining oxidative chlorine after variable times of exposure to the UVA photons.

Compound **2** lost all of its bound chlorine within 20 h of UVA exposure, while compound **1** retained about 55% of its initial chlorine loading. A comparable experiment, in which solid-state commercial 1,3-dichloro-5,5-dimethylhydantoin and 1-chloro-5,5-dimethylhydantoin (synthesized by a straightforward exchange reaction of 1,3-dichloro-5,5-dimethylhydantoin and unchlorinated 5,5-dimethylhydantoin) were irradiated for 24 h, showed that the dichloro derivative lost only 6.0 % of its chlorine, and the 1-chloro-5,5-dimethyl derivative lost only 4.2 % of its chlorine. Over the same time period of irradiation **1** lost 52 % and **2** lost 100 % of their chlorine contents. Thus clearly the presence of a methyl substituent on the imide nitrogen in **1** and the amide nitrogen in **2** was instrumental in the photolytic decomposition process for these two trimethyl derivatives.

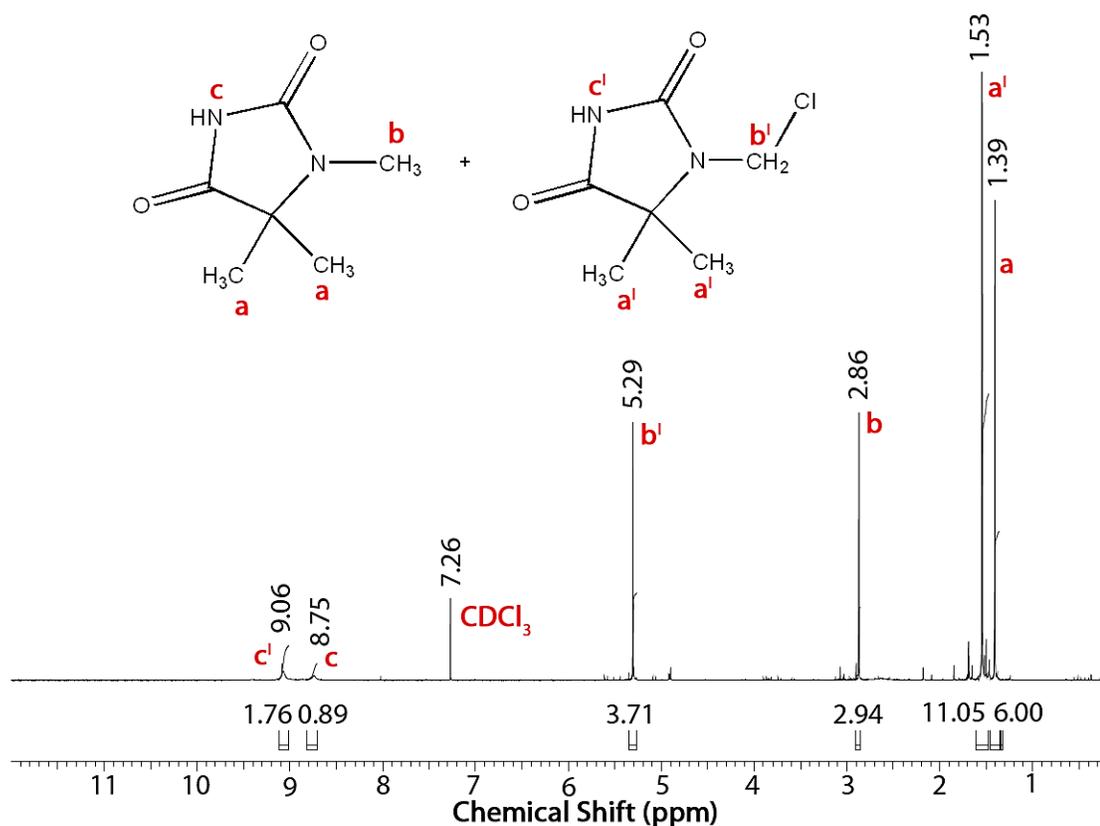
**Scheme 3.2** Three Potential Reaction Products for the Photodecomposition under UVA Irradiation of 1-chloro-3,5,5-trimethylhydantoin (**1**) and 3-chloro-1,5,5-trimethylhydantoin (**2**)



The structures of some logical photodecomposition products of **1** and **2** are shown in Scheme 3.2. Other decomposition products such as ones arising from rupture of the hydantoin ring are unlikely based upon the observations for 1,3-dichloro-5,5-dimethylhydantoin and 1-chloro-5,5-dimethylhydantoin discussed above. Both compounds **1** and **2** were irradiated in the Accelerated Weather Tester in the solid state in open vessels and in CDCl<sub>3</sub> solvent in an NMR tube. The time of irradiation was 144 h for both compounds; at these irradiation times all chlorine was lost from both compounds. Product analyses were accomplished by <sup>1</sup>H NMR. For the solid state decomposition, the product ratio of **1a:1b** was about 10:1 with no **1c** detected; the product ratio of **2a:2b** was 1.5:1.0 with no **2c** detected. The <sup>1</sup>H NMR spectra for the photodecomposition products for the two compounds in CDCl<sub>3</sub> are shown in Figures 3.4 and 3.5.



**Figure 3.4**  $^1\text{H}$  NMR spectrum for the photodecomposition products of compound 1 in  $\text{CDCl}_3$ .



**Figure 3.5**  $^1\text{H}$  NMR spectrum for the photodecomposition products of compound **2** in  $\text{CDCl}_3$ .

From comparison with  $^1\text{H}$  NMR data, the assignments of the resonances to **1a**, **1b**, **2a**, and **2b** are obvious, and an integration of the resonances assigned to the 5,5-dimethyl substituents provides a ratio of about 11:6 for both **1a:1b** and **2a:2b**, again with no **1c** or **2c** detected. Therefore, several questions come to mind. Why is there no chlorine migration to the methyl groups at carbon 5 on the hydantoin ring during decomposition of either N-chloramine? Why are the product ratios different for the solid state as compared to the solution process (Table 3.1)? And most important, what mechanism(s) logically

explain(s) the chlorine migration process? DFT computations have been carried out to assist in the interpretation of the experimental data.

**Table 3.1** Experimental Final Product Ratios Obtained under UVA Light

Exposure

Product ratio	<b>1a:1b</b>	<b>2a:2b</b>
solid state	10:1	1.5:1
in CDCl <sub>3</sub>	11:6	11:6

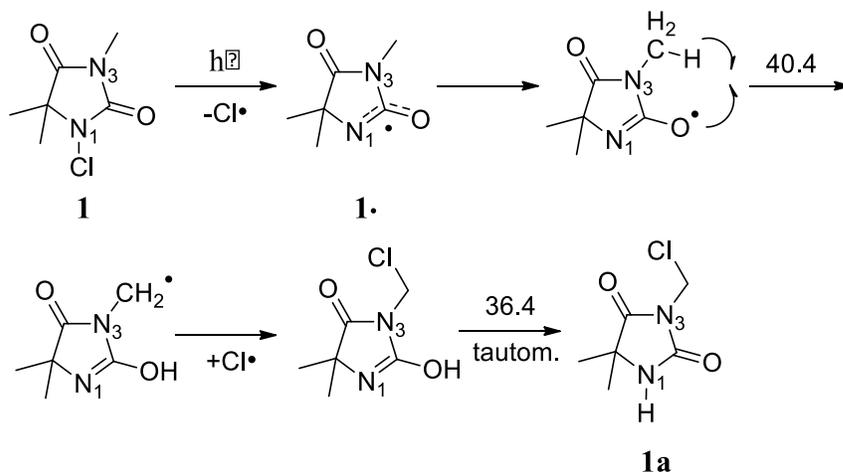
### 3.3.3 Mechanistic Considerations

As stated earlier, a Hoffmann-Loeffler photorearrangement mechanism is not possible for **1** or **2** because of geometry constraints in the necessary cyclic transition structure. Therefore, alternative radical mechanisms must be considered to explain the formation of the photodecomposition products. While the transformations for **1** and **2** likely involve a hydrogen atom transfer reaction (HATR), the exact reaction pathway is intuitively unclear. Two mechanisms for the HATRs are proposed: (1) an intramolecular process in which a hydrogen atom undergoes a series of sigmatropic shifts, and (2) an intermolecular pathway in which a radical abstracts a hydrogen atom from a neighboring molecule. Calculations at the UB3LYP/6-311++G(2d,p) level have been carried out to reveal the most favorable route.

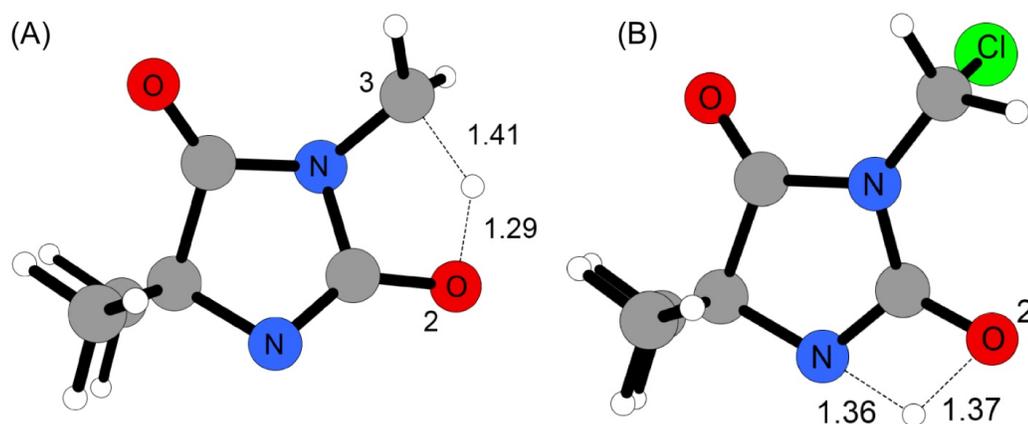
### 3.3.3.1 Intramolecular Mechanism

Following homolytic dissociation of the N–Cl bond in **1**, radical delocalization, as illustrated in **1•**, may induce the adjacent carbonyl oxygen to be the reactive atom. (Scheme 3.3)

**Scheme 3.3** Proposed Intramolecular Hydrogen Atom Transfer Mechanism for the Formation of **1a** from **1**.  $\Delta G^\ddagger$  (kcal/mol) Values Computed using UB3LYP/6-311++G(2d,p).



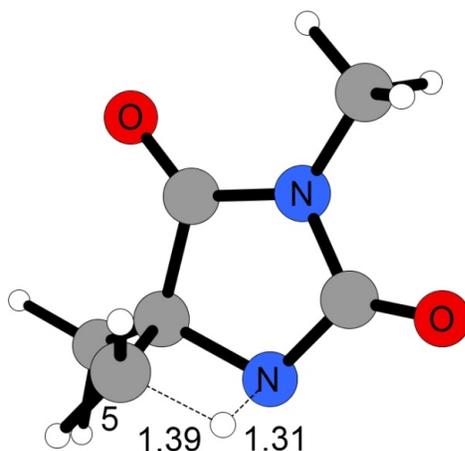
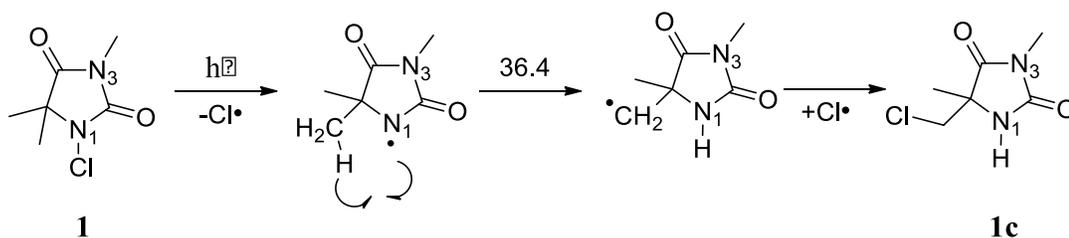
The oxygen radical could then abstract a hydrogen atom from the neighboring methyl group attached to the 3-nitrogen and generate a methylene radical; a  $\Delta G^\ddagger$  of 40.4 kcal/mol was computed for this step. The radical could then undergo a rapid chlorination of the 3-methylene group followed by tautomerization, with a calculated  $\Delta G^\ddagger$  of 36.4 kcal/mol, to yield **1a**. The DFT optimized transition structures in the reaction pathway are shown in Figure 3.6.



**Figure 3.6** UB3LYP/6-311++G(2d,p) calculated transition structures in the **1** to **1a** intramolecular reaction pathway for (A) the hydrogen atom transfer from 3-methyl to the 2-oxygen, and (B) the subsequent tautomerization between the 2-oxygen and 1-nitrogen. Optimized distances given in angstroms.

Since the amidoyl nitrogen is adjacent to the two methyl groups located at the 5 position, it is also conceivable for the HATR to occur from one of these methyl groups. (Scheme 3.4 and Figure 3.7)

**Scheme 3.4** Alternative Pathway for the Proposed Intramolecular Mechanism Given in Scheme 5 that Leads to the Formation of **1c**.  $\Delta G^\ddagger$  (kcal/mol) Calculated using UB3LYP/6-311++G(2d,p)

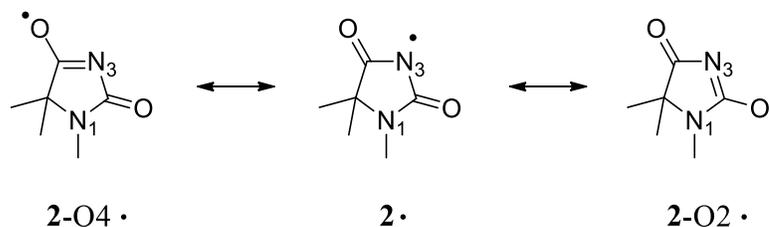


**Figure 3.7** UB3LYP/6-311++G(2d,p) calculated transition structure for the intramolecular hydrogen atom transfer from 1· to 5-methylene-1,5-dimethylhydantoin. Optimized distances given in angstroms.

While there is no NMR evidence for the formation of product **1c**, the intramolecular mechanism proposed in Scheme 3.4 could account for its existence. The calculated  $\Delta G^\ddagger$  of 34.6 kcal/mol for the formation of **1c** suggests that species should be more prevalent than **1a** (i.e.,  $\Delta G^\ddagger$  of 40.4 kcal/mol), yet there was no experimental evidence for the formation of **1c**. Therefore, this proposed intramolecular mechanism is unlikely to occur.

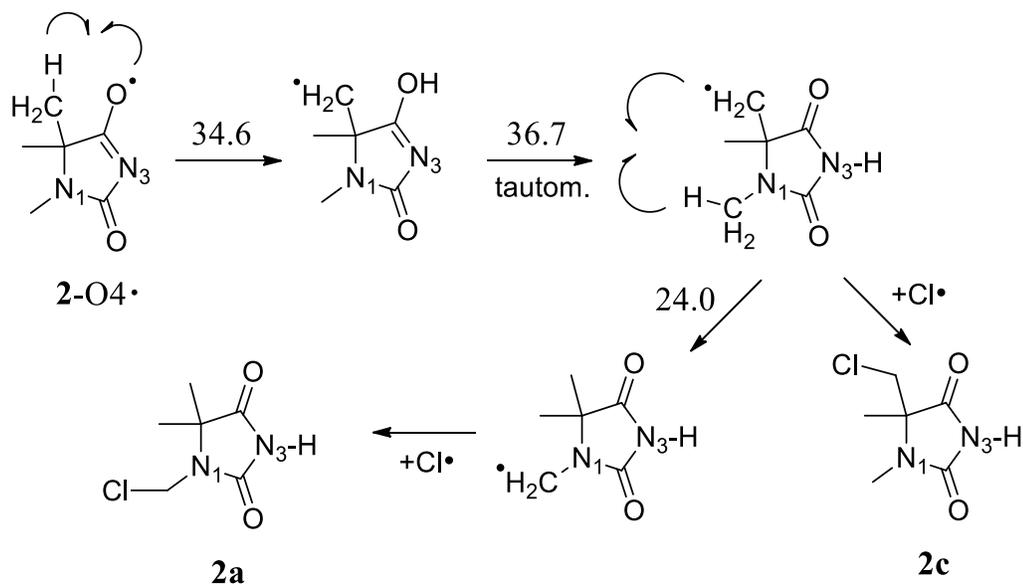
In the case of **2**, a similar intramolecular mechanism could be proposed. Following the departure of Cl• from the imide nitrogen creating an imidoyl radical, three resonance structures can be written for the radical as illustrated in Scheme 3.5.

**Scheme 3.5** Three Resonance Structures for **2**•

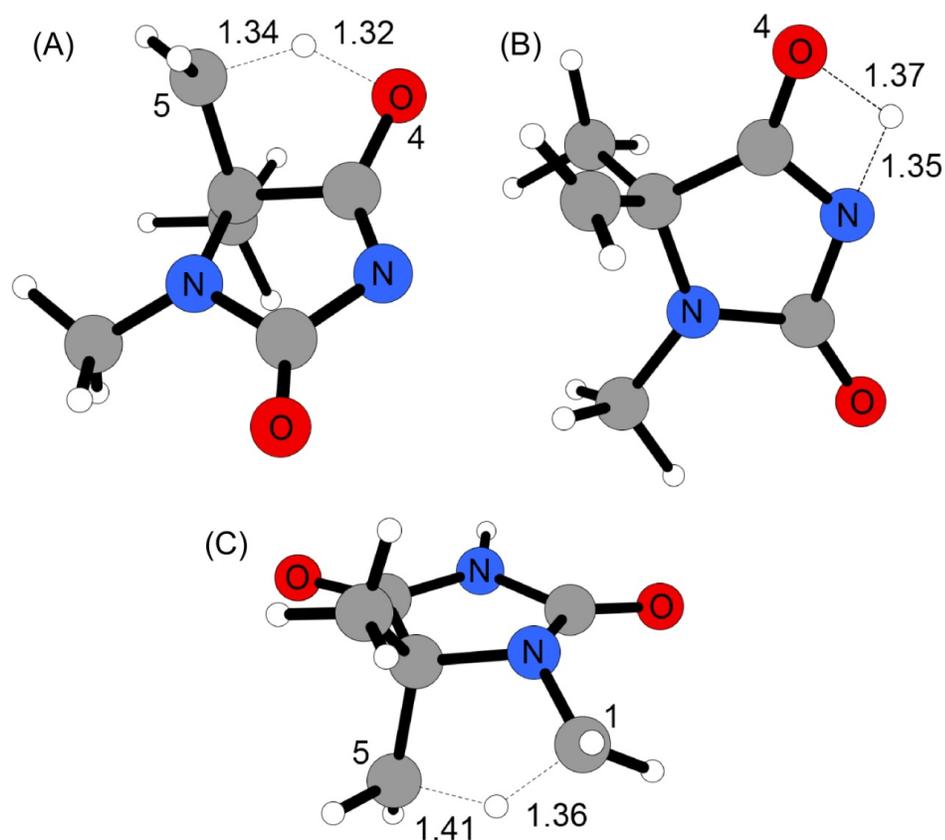


Beginning with the resonance structure **2-O4**•, the 4-oxygen radical could remove a hydrogen atom from the 5-methyl group with a computed  $\Delta G^\ddagger$  of 34.6 kcal/mol. Following the formation of the 5-methylene radical, a tautomerization could occur with a computed  $\Delta G^\ddagger$  of 36.7 kcal/mol. The 5-methylene radical may then transfer a hydrogen atom with 1-methyl, with a calculated  $\Delta G^\ddagger$  of 24.0 kcal/mol, followed by the combination of the resultant 1-methylene radical with the Cl• to create **2a** (Scheme 3.6).

**Scheme 3.6** Proposed Intramolecular Hydrogen Atom Transfer Mechanism for the Formation of **2a** and **2c** from **2-O4•**.  $\Delta G^\ddagger$  Values (kcal/mol) Computed using UB3LYP/6-311++G(2d,p).



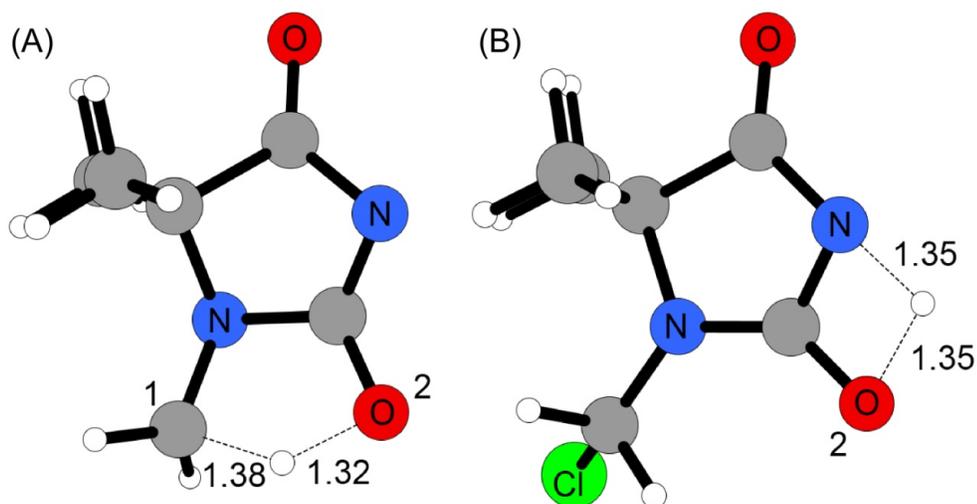
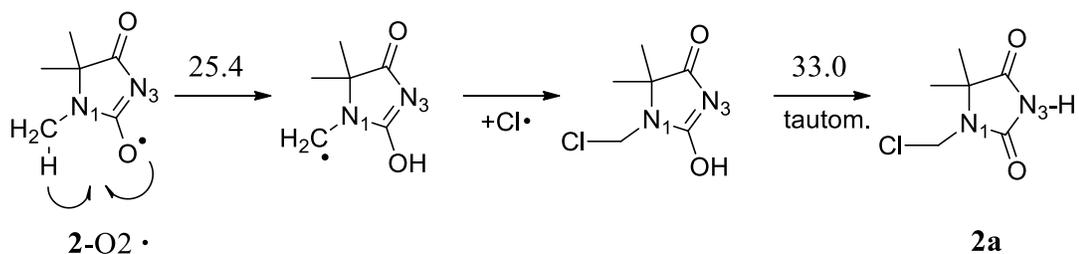
Alternatively, the  $\text{Cl}\cdot$  could combine with the 5-methylene radical facilitating the formation of **2c** with little or no activation barrier. The optimized transition structures for the **2** to **2a/c** HATRs are given in Figure 3.8. This intramolecular mechanism would largely favor the formation of **2c** over **2a**, similar to the preference for **1c** over **1a**, which contradicts experimental data from the NMR spectra indicating no formation of either **1c** or **2c**.



**Figure 3.8** Calculated UB3LYP/6-311++G(2d,p) transition structures for the HATRs of (A) 5-methyl in 2-O4· to oxygen radical, (B) the tautomerization of 2-hydroxyl-5-methylene radical, and (C) 1-methyl to the 5-methylene radical group. Distances given in angstroms.

Initiating the HATRs from the 2-O2· structure has the 2-carbonyl oxygen potentially abstracting a hydrogen atom from the 1-methyl group, with a computed  $\Delta G^\ddagger$  of 25.4 kcal/mol. The combination of Cl· and the 1-methylene radical and a subsequent tautomerization requiring a calculated  $\Delta G^\ddagger$  of 33.0 kcal/mol would then lead to **2a** (Scheme 3.7). The corresponding transition structures are given in figure 3.9.

**Scheme 3.7** Proposed Intramolecular Hydrogen Atom Transfer Mechanism for the Formation of **2a** from 2-O2 $\cdot$ .  $\Delta G^\ddagger$  Values (kcal/mol) Computed using UB3LYP/6-311++G(2d,p).



**Figure 3.9** Calculated UB3LYP/6-311++G(2d,p) intramolecular transition structures for (A) the hydrogen transfer from the 1-methyl to the 2-oxygen and (b) the tautomerization between the 2-oxygen and 3-nitrogen. Distances given in angstroms.

This intramolecular route is preferable to the 2-O4 $\cdot$  path (Scheme 3.6), as the overall rate-limiting  $\Delta G^\ddagger$  is lower, and no formation of **2c** is expected to occur. While all reasonable mechanistic possibilities for the intramolecular HATRs have been considered,

Schemes 3.3-3.7 have no direct route for producing either **1b** or **2b**. Potentially a hydrogen atom may be generated from the decomposition products, or from the walls of the container, to form **1b/2b**. However, considering that the predicted favorable formation of **1c/2c** in Schemes 3.6 and 3.7 are contrary to the experimental observations, alternative intermolecular reaction mechanisms should be considered.

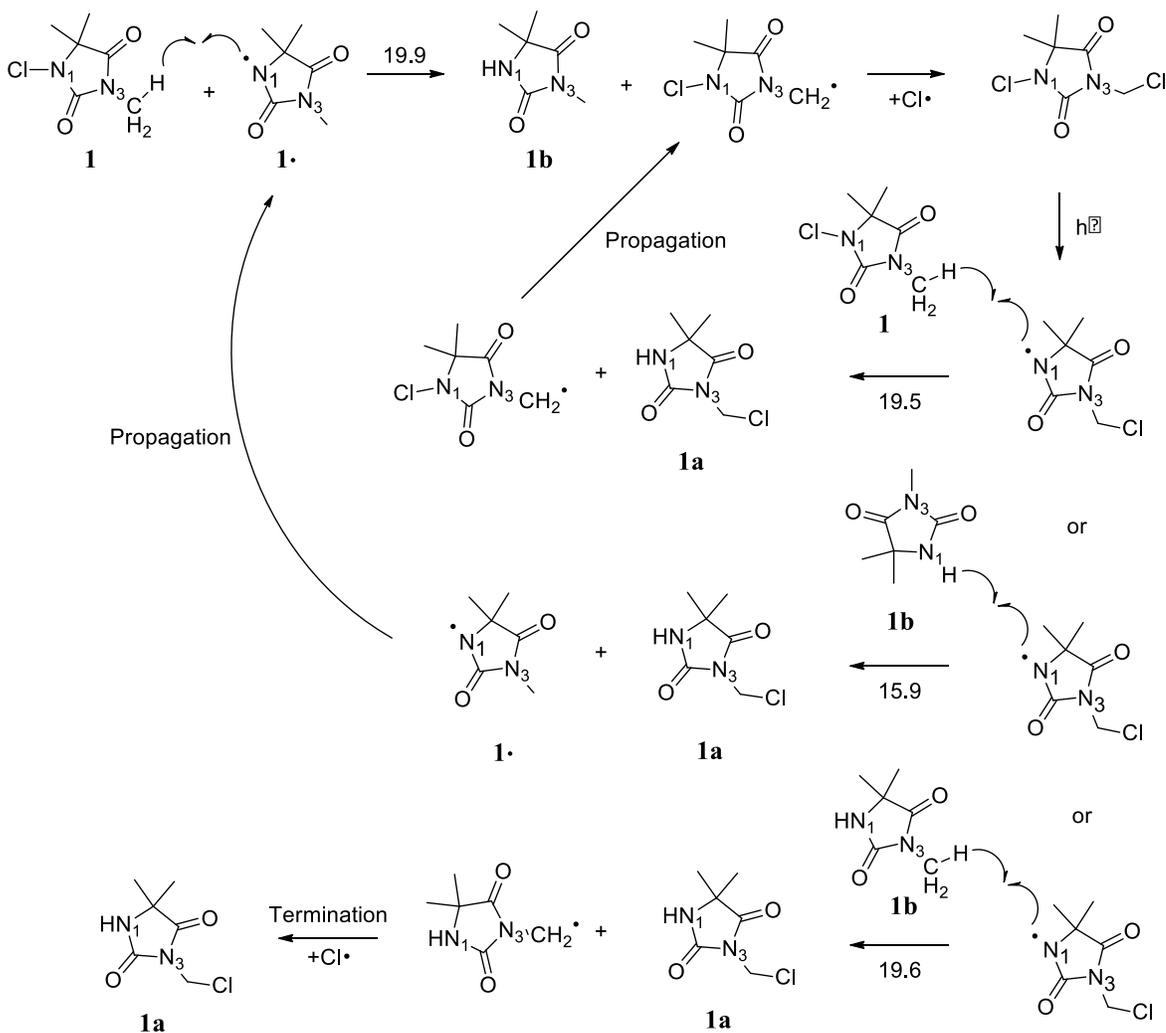
### 3.3.3.2 Intermolecular Mechanism

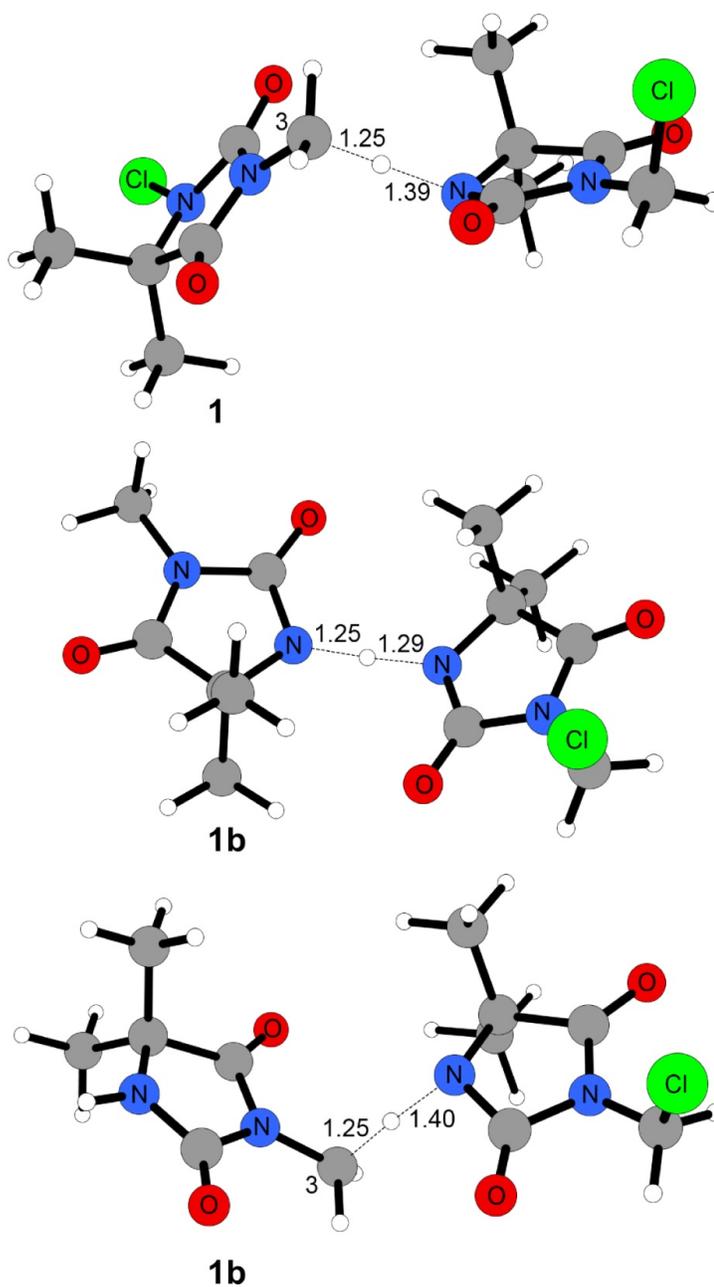
The amidoyl nitrogen in **1•** could undergo an intermolecular hydrogen transfer with a 3 or 5 position methyl group of a neighboring **1** molecule. Initiating the HATR from the methyl group on the 3-nitrogen would initially lead to the formation of **1b** and a 3-methylene radical with a computed  $\Delta G^\ddagger$  of 19.9 kcal/mol (Scheme 3.8). The radical then could combine with Cl• to produce 1-chloro-3-chloromethyl-5,5-dimethylhydantoin, an intermediate compound that could potentially account for a small amount of unidentified impurity product detected by NMR during the decomposition process before all oxidative chlorine was lost. N-Cl bond cleavage of 1-chloro-3-chloromethyl-5,5-dimethylhydantoin would produce another amidoyl radical, N1-radical-3-chloromethyl-5,5-dimethylhydantoin, that could abstract a hydrogen atom from either the starting material **1** or the **1b** produced from the previous step. Three scenarios emerge in which the hydrogen atom could be transferred from the 3-methyl of **1**, the 3-methyl of **1b**, or the 1-nitrogen of **1b** (Figure 3.10). All reaction pathways directly produce **1a** and potentially another radical that propagates the radical mechanism towards the continued production of **1a** and **1b** (Scheme 3.9). The HATR from the 3-nitrogen of **1b** is the favored route with a computed  $\Delta G^\ddagger$  of 15.9 kcal/mol as compared to 19.5 and 19.6 kcal/mol for the

hydrogen abstraction from the 3-methyl group of **1** and **1b**, respectively. Following the most energetically favored route, **1a** and **1b** are formed, but no net gain of **1b** is realized as the compound is consumed during the reaction. Overall, more **1a** should be produced than **1b** following the mechanism given in Scheme 3.9, in accord with the experimental observations.

**Scheme 3.8** Intermolecular mechanism for **1** yielding products **1a** and **1b**.

$\Delta G^\ddagger$  values (kcal/mol) computed using UB3LYP/6-311++G(2d,p).

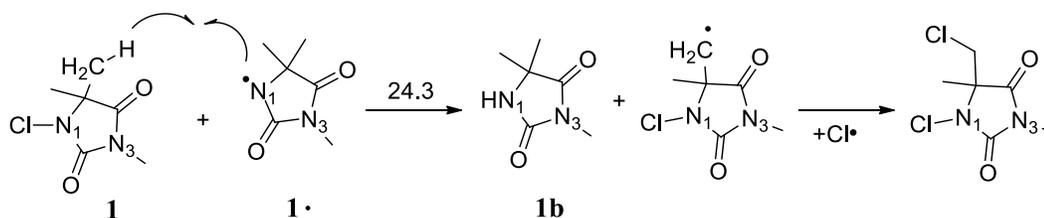




**Figure 3.10** Calculated UB3LYP/6-311++G(2d,p) intermolecular transition structures for the HATR between N1-radical-3-chloromethyl-5,5-dimethylhydantoin and either the starting material 1 or 1b. Optimized distances given in angstroms.

Alternatively, a HATR originating from the methyl group on the 5-carbon position would produce **1b** and 1-chloro-3-methylene-5,5-dimethylhydantoin radical (Scheme 3.9). The  $\Delta\Delta G^\ddagger$  for the 3-methyl-based HATR in scheme 3.8 is 4.4 kcal/mol lower than that for the 5-methyl, which corresponds to an approximately 2000-fold difference in rates assuming equal values for the pre-exponential factors and a temperature of 298 K. Given this large difference in rates and the lack of experimental evidence for a Cl atom bonded at the 5-methyl position, it is unlikely that the reaction in Scheme 3.9 would be competitive.

**Scheme 3.9** Intermolecular Mechanism of **1** Yielding Products **1b** and 1-chloro-3-methylene-5,5-dimethylhydantoin.  $\Delta G^\ddagger$  values (kcal/mol) computed using UB3LYP/6-311++G(2d,p).



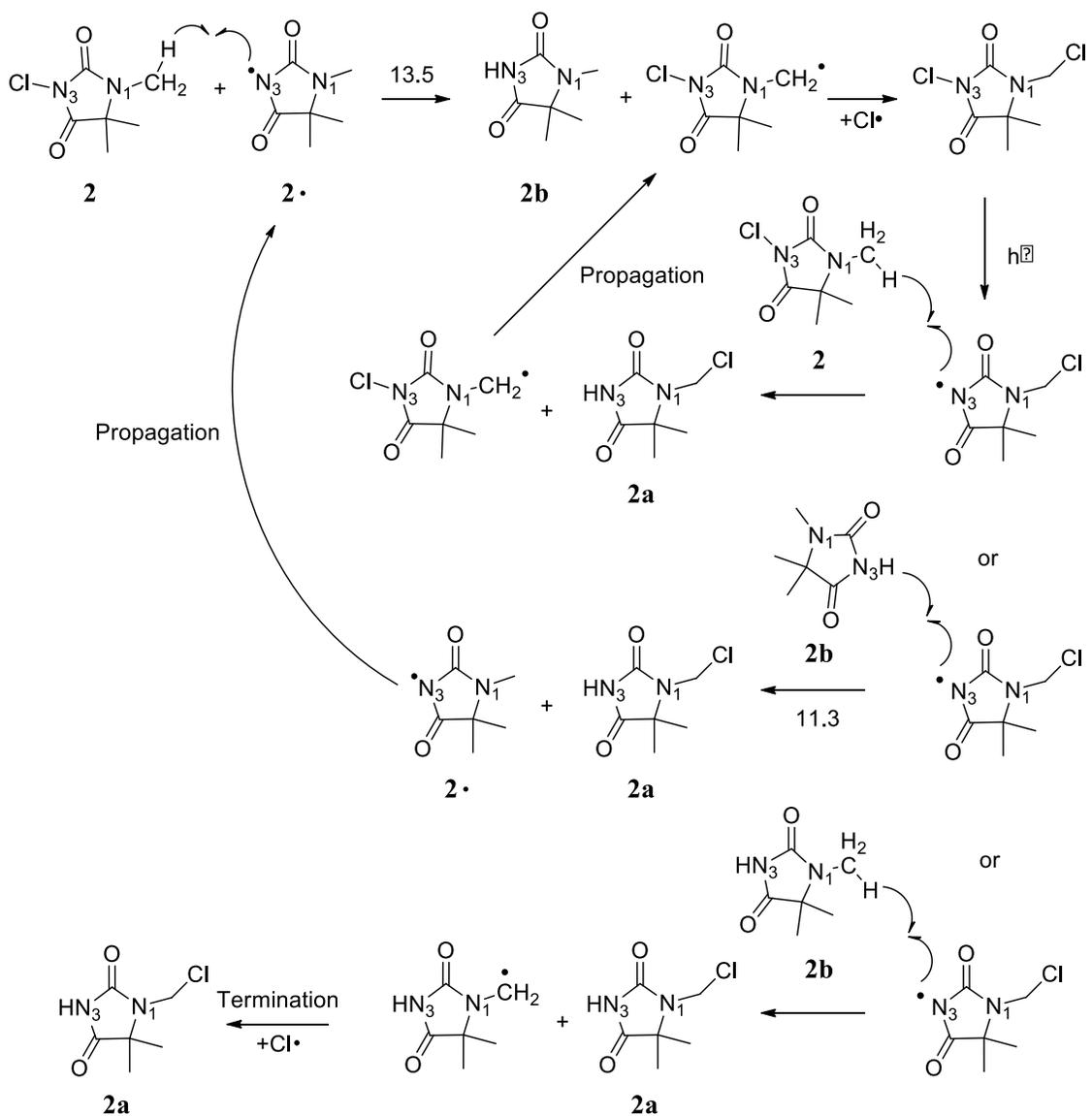
The intermolecular reaction of **2** to yield **2a** and **2b** is expected to have a mechanism very similar to that of **1**; the proposed pathways are presented in the schemes 3.10 and 3.11.

The 3-nitrogen radical in **2•** can also perform an intermolecular hydrogen atom transfer with a 1- or 5-position methyl group of a neighboring **2** molecule. Initiating the HATR from the methyl group on the 1-nitrogen would initially lead to the formation of **2b** and a 1-methylene radical with a computed  $\Delta G^\ddagger$  of 13.5 kcal/mol. The rate-limiting

$\Delta G^\ddagger$  for **2** is considerably lower than the 19.9 kcal/mol computed for **1**, which is consistent with the experimental stability tests carried out under UVA photolysis (Figure 3.3). The remainder of the mechanism should follow the same basic pathways as the mechanism proposed for **1**. However, in spite of repeated efforts to locate the transition structures for the HATR between N3-radical-1-chloromethyl-5,5-dimethylhydantoin and the 1-methyl group of **2** and **2b**, the calculations would not yield optimized structures. The HATR between the N3-radical-1-chloromethyl-5,5-dimethylhydantoin and the 3-nitrogen of **2b** was found to give a computed  $\Delta G^\ddagger$  of 11.3 kcal/mol. As the same route in the reaction pathway in Scheme 11 for **1** possessed the lowest barrier, it may be expected that the same general conclusions would hold for **2**.

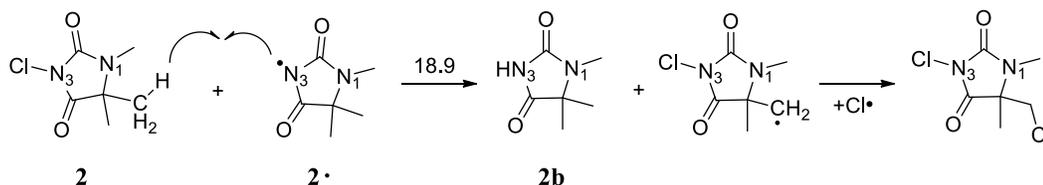
**Scheme 3.10** Intermolecular mechanism of **2** yielding products **2a** and **2b**.

$\Delta G^\ddagger$  values (kcal/mol) computed using UB3LYP/6-311++G(2d,p).



Conversely, **2•** could abstract a hydrogen atom from the 5-methyl group of a neighboring **2** molecule with a computed  $\Delta G^\ddagger$  of 18.9 kcal/mol or a  $\Delta\Delta G^\ddagger$  of 5.4 kcal/mol compared to the 1-methyl group (scheme 3.11).

**Scheme 3.11** Intermolecular hydrogen transfer between **2** and **2·** at the 5-methyl position.  $\Delta G^\ddagger$  value (kcal/mol) computed using UB3LYP/6-311++G(2d,p).



This HATR difference is equivalent to about an 11,000-fold rate increase in favor of the 1-methyl group at room temperature. Given this large rate difference and the lack of a 5-chloro product observed experimentally, this reaction pathway is unlikely to occur.

### 3.3.4 Final Product Ratios

The intermolecular mechanisms had lower absolute  $\Delta G^\ddagger$  values as compared to the intramolecular mechanisms; however, it is difficult to assess whether the molecules would possess the necessary degrees of freedom for the proposed geometries to occur in the solid state. In solution, however, the intermolecular mechanism would be expected to dominate. Experimental evidence may be found in the final product ratios of 11:6 for both **1a:1b** and **2a:2b** in  $\text{CDCl}_3$  (Table 3.1). The intermolecular mechanisms in Schemes 3.8 and 3.10 above are essentially identical, which would facilitate similar product ratios. However, the results for the solid compounds are dramatically different with **1a** dominating **1b** with a ratio of 10:1 as compared to a lower ratio of 1.5:1 for **2a:2b**. The product ratios for the solid state are more difficult to rationalize from the calculations, but

could possibly be the result of a combination of the inter- and intra-molecular mechanisms discussed herein.

### 3.4 Conclusions

The rate of loss of  $\text{Cl}^+$  for the antimicrobial 1-chloro-3,5,5-trimethylhydantoin (**1**) and 3-chloro-1,5,5-trimethylhydantoin (**2**) compounds has been measured in the presence of and in the absence of exposure to UVA irradiation. Compound **1** was determined to be more stable than **2** in both cases. Under UVA exposure the hydantoins yielded an intriguing chlorine atom migration reaction. Two novel mechanisms have been proposed: (1) an intramolecular process in which a hydrogen atom undergoes a series of sigmatropic shifts, and (2) an intermolecular pathway in which a radical abstracts a hydrogen atom from a neighboring molecule. DFT calculations at the UB3LYP/6-311G++(2d,p) level predicted **2** to have a lower  $\Delta G^\ddagger$  than **1** for both proposed HATR mechanisms in accord with the experimental chlorine stability measurements. However, the intermolecular route possessed lower absolute  $\Delta G^\ddagger$  values as compared to the intramolecular mechanism, agreed with a lack of **1c/2c** formation as found from the NMR spectra, and was consistent with the experimentally measured final product ratios of 11:6 for **1a:1b** and **2a:2b** in solution after 144 h of UVA exposure. The combination of the DFT calculations, NMR spectra, final product ratios, and stability tests support the conclusion of an intermolecular mechanism for the photodecomposition in solution. The mechanism for the solid state is more difficult to rationalize, but may utilize a combination of inter- and intra-molecular reaction pathways to reach the final products.

### 3.5 References

- (1) Becke, A. D. *Journal of Chemical Physics* **1993**, *98*, 5648–5652.
- (2) Lee, C.; Yang, W.; Parr, R. G. *Physical Review* **1988**, *37*, 785–789.
- (3) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. S.; M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. M.; G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. H.; A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. H.; M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. N.; Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. M. J.; J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. B.; K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. N.; K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. T.; M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. C.; V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. S.; O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. O.; R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. V.; P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. D.; O. Farkas, J. B. Foresman, J. V. Ortiz, J. C.; Fox, D. J. F. *Gaussian 09, Revision B.01*; Gaussian, Inc.: Wallingford, CT, 2010.
- (4) Akdag, A.; Worley, S. D.; Acevedo, O.; McKee, M. L. *Journal of Chemical Theory and Computation* **2007**, *3*, 2282–2289.
- (5) Kocer, H. B.; Akdag, A.; Worley, S. D.; Acevedo, O.; Broughton, R. M.; Wu, Y. *ACS Applied Materials & Interfaces* **2010**, *2*, 2456–64.

# Chapter 4 Pairwise Alternatives to Ewald Summation for Calculating Long-Range Electrostatics in Ionic Liquids

## 4.1 Abstract

Room temperature ionic liquid calculations require extensive sampling due to the large degree of localized structuring in the liquid phase relative to conventional solutions. Consequently, a large amount of computer time is required for the convergence of solvent properties, much of which is spent evaluating long-range electrostatics via Ewald summations. The damped Coulomb potential and cutoff-neutralized method of Wolf et al. provides the framework for an accurate, linear-scaling alternative to Ewald in the ionic liquid simulations.<sup>1</sup> The method has been the subject of multiple modifications for improved accuracy, including the damped Columbic potential of Zahn et al.,<sup>2</sup> the damped shifted force method of Fennell and Gezelter<sup>3</sup>, and the shifted force gradient of Kale and Herzfeld.<sup>4</sup> These pairwise electrostatic interaction alternatives along with the CHARMM shifted force potential<sup>5</sup> and a new method proposed herein, the shifted force 3<sup>rd</sup> derivative (SF3), have been examined on 59 unique ionic liquid combinations of 1-alkyl-3-methylimidazolium [RMIM] (R = M (methyl), E (ethyl), B (butyl), H (hexyl), and O

**Reprinted with permission from *J. Chem. Theory Comput.*, 2013, 9 (2), pp 944–950 .  
Copyright 2013 American Chemical Society.**

(octyl)) and N-alkylpyridinium [RPyr] cations, along with  $\text{Cl}^-$ ,  $\text{PF}_6^-$ ,  $\text{BF}_4^-$ ,  $\text{NO}_3^-$ ,  $\text{AlCl}_4^-$ ,  $\text{Al}_2\text{Cl}_7^-$ , and  $\text{TfO}^-$  anions. Monte Carlo simulations utilizing our custom OPLS-AA ionic liquid force field and employing the pairwise alternatives with multiple cutoff distances and electrostatic damping values are compared to the energetics from full Ewald sums. (4.2)

## 4.2 Theory

As an early alternative to the more costly Ewald schemes, a general shifting scheme was first proposed by Brooks, Pettitt, and Karplus<sup>6</sup> and implemented by Levitt et al. to treat electrostatics in biological simulations.<sup>7</sup> The method modifies the interactions energy by subtracting it from a truncated Taylor series of the original energy (Eq. 4.1), where  $V(r_{ij})$  and  $V(R_C)$  are the original potentials and equal to  $1/r_{ij}$  and  $1/R_C$ , respectively;  $dV(R_C)/dr$  is its derivative with respect to  $r_{ij}$  evaluated at the cutoff distance  $R_C$ , and  $r_{ij}$  is the distance between particles. Levitt et al. tested the  $n = 1$  and 2 cases and concluded that the former provided a better electrostatic description in weak to mildly ionic biological environments, e.g., proteins and nucleic acids in water, and had little effect on hydrogen bonding. The  $n = 1$  case is called here the *shifted force* (SF) method and given as Eq. 4.2.

$$V_{n\text{-shifted}}(r_{ij}) = \begin{cases} V(r_{ij}) - V(R_C) - \sum_{m=1}^n \frac{1}{m!} (r_{ij} - R_C)^m \frac{d^m V(R_C)}{dr^m}, & r_{ij} \leq R_C \\ 0, & r_{ij} > R_C \end{cases} \quad (4.1)$$

$$V_{SF}(r_{ij}) = \begin{cases} q_i q_j \left[ \frac{1}{r_{ij}} - \frac{1}{R_C} + \left( \frac{1}{R_C^2} \right) (r_{ij} - R_C) \right], & r_{ij} \leq R_C \\ 0, & r_{ij} > R_C \end{cases} \quad (4.2)$$

Wolf et al. proposed a similar technique to Levitt et al. where simulations of Coulombic systems were truncated spherically with a pairwise  $1/r_{ij}$  summation.<sup>1</sup> In their method, the electrostatic interaction is effectively short-ranged in the condensed phase and charge neutrality is enforced within the cutoff radius. Image charges are subtracted out through placement on the cutoff sphere. Their method is called the *shifted potential* (SP) and is specified by Eq. 4.3. A distance-dependent damping function given by the complementary error function (erfc) from the Ewald sums was subsequently included to aid in the convergence of the calculated Madelung energies to yield the *damped shifted potential* (DSP) method (Eq. 4.4), where  $\alpha$  is the damping parameter with units of  $\text{\AA}^{-1}$ . The Wolf technique has proven successful in several condensed-phase MD simulations.<sup>8-</sup>  
<sup>10</sup> However, the use of the erfc is time-consuming and methods have been proposed that reduce the amount of times the erfc calculation is required.<sup>11</sup>

$$V_{SP}(r_{ij}) = \begin{cases} q_i q_j \left( \frac{1}{r_{ij}} - \frac{1}{R_C} \right), & r_{ij} \leq R_C \\ 0, & r_{ij} > R_C \end{cases} \quad (4.3)$$

$$V_{DSP}(r_{ij}) = \begin{cases} q_i q_j \left( \frac{\text{erfc}(\alpha r_{ij})}{r_{ij}} - \frac{\text{erfc}(\alpha R_C)}{R_C} \right), & r_{ij} \leq R_C \\ 0, & r_{ij} > R_C \end{cases} \quad (4.4)$$

Truncating the interactions yields a discontinuity in the potential that corresponds to an infinite force acting between atoms that cross the discontinuity. Many MD algorithms neglect this force due to poor treatment in the soft potentials. Consequently, error is manifested by the poor energy conservation in the simulation.<sup>12</sup> Zahn et al. revised Wolf's damping method to achieve energy conservation in MD simulations of TIP3P and SPC water models.<sup>2</sup> The proposed *damped Coulomb potential* (DC) method (Eq. 4.5) was found to yield considerable improvement in dielectric properties and in static properties, including potential energy, particle density, and radial distribution functions. The method was originally intended for molecular liquids consisting of neutral molecules without internal Coulombic interaction; however, larger cutoff radii or a modification to the self-term has been speculated to satisfy ionic systems. Fennell and Gezelter further revised Zahn et al.'s DC method with two differences: (a) the  $V(R_C)$  term has been added to reduce any potential discontinuity as particles cross  $R_C$  and (b) the sign of the derivative portion is changed.<sup>3</sup> Their method is called the *damped shifted force* (DSF) and given by Eq. 4.6.

$$V_{DC}(r_{ij}) = \begin{cases} q_i q_j \left[ \frac{\text{erfc}(\alpha r_{ij})}{r_{ij}} - \left( \frac{\text{erfc}(\alpha R_C)}{R_C^2} + \frac{2\alpha}{\pi^{1/2}} \frac{\exp(-\alpha^2 R_C^2)}{R_C} \right) (r_{ij} - R_C) \right], & r_{ij} \leq R_C \\ 0, & r_{ij} > R_C \end{cases} \quad (4.5)$$

$$V_{DSF}(r_{ij}) = \begin{cases} q_i q_j \left[ \frac{\text{erfc}(\alpha r_{ij})}{r_{ij}} - \frac{\text{erfc}(\alpha R_C)}{R_C} + \left( \frac{\text{erfc}(\alpha R_C)}{R_C^2} + \frac{2\alpha}{\pi^{1/2}} \frac{\exp(-\alpha^2 R_C^2)}{R_C} \right) (r_{ij} - R_C) \right], & r_{ij} \leq R_C \\ 0, & r_{ij} > R_C \end{cases} \quad (4.6)$$

If the damping parameter  $\alpha$  equals zero, then the DSP (Eq. 4.4) reduces to the undamped SP prescription of Wolf et al. (Eq. 4.3) and the DSF (Eq. 4.6) reverts to the undamped SF method (Eq. 4.2) derived from the  $n = 1$  case in the Levitt et al. study. Extensive examination of the DSF, DSP, SF, and SP methods has been reported by Fennell and Gezelter on liquid and solid water, on crystal NaCl and high temperature (7000 K) molten NaCl melt, and on weak and strong NaCl solutions. They determined the undamped SF and moderately damped DSP methods to be ideal alternatives in the calculation of energetic and dynamic characteristics of the systems when compared to particle-mesh Ewald.<sup>3</sup> Hansen et al. independently verified the accuracy of the SF method with a reported speed-up factor of 2 – 3 relative to the Wolf et al. method for bulk molten salt and the SPC/FW water model.<sup>8</sup>

Some problems have been reported for the SF method, however, including structural artifacts in ionic liquids with charges of large magnitude, i.e.,  $\text{ThCl}_4$  at 1000K.<sup>4</sup> To address the issue, Kale and Herzfeld, extended the potential by taking the second derivative of the potential with respect to distance and called the method the *shifted force gradient* (SFG).<sup>4</sup> The method is derived from the Levitt et al. method (Eq. 4.1), where  $n = 2$  in this case and yields Eq. 4.7. Kale and Herzfeld reported that in cases of extreme ionicity in water the greater potential softening of the  $n = 2$  correction provided a reliable liquid structure in the NVT ensemble. In this work, the Levitt et al. equation is further expanded to the  $n = 3$  case and called the *shifted force 3<sup>rd</sup> derivative* (SF3) method (Eq. 4.8).

$$V_{SFG} = \begin{cases} q_i q_j \left[ \frac{1}{r_{ij}} - \frac{1}{R_C} + (r_{ij} - R_C) \left( \frac{1}{R_C^2} \right) - (r_{ij} - R_C)^2 \left( \frac{1}{R_C^3} \right) \right], & r_{ij} \leq R_C \\ 0, & r_{ij} > R_C \end{cases} \quad (4.7)$$

$$V_{SF3} = \begin{cases} q_i q_j \left[ \frac{1}{r_{ij}} - \frac{1}{R_C} + (r_{ij} - R_C) \left( \frac{1}{R_C^2} \right) - (r_{ij} - R_C)^2 \left( \frac{1}{R_C^3} \right) + (r_{ij} - R_C)^3 \left( \frac{1}{R_C^4} \right) \right], & r_{ij} \leq R_C \\ 0, & r_{ij} > R_C \end{cases} \quad (4.8)$$

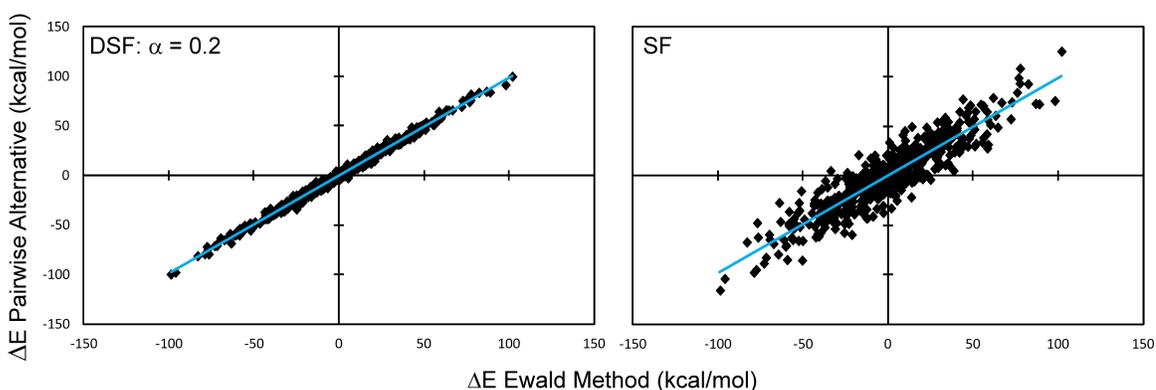
### 4.3 Computational Methods

Metropolis Monte Carlo (MC) simulations were performed beginning with fully equilibrated ionic liquid boxes from our recent OPLS-AA force field parameterization effort.<sup>13</sup> Periodic boundary conditions have been applied to the boxes containing 190 ion pairs. All cations were fully flexible, i.e. all bond stretching, angle bending, and torsional motions were sampled. Anions were simulated as rigid molecules. The use of rigid anions in OPLS-AA has been shown to provide an accurate representation of ionic liquid physical properties, including use as a reaction medium for computed QM/MM Diels-Alder<sup>14</sup> and Kemp elimination<sup>13</sup> reaction studies. Solvent-solvent intermolecular cutoff distances of 9, 12, and 15 Å were employed uniformly across all methods using the tail carbon atom of each side chain (methyl and alkyl), a midpoint carbon on the alkyl chain, and the ring carbon between both nitrogens for imidazolium. Center atoms, e.g. B in BF<sub>4</sub><sup>-</sup> and P in PF<sub>6</sub><sup>-</sup>, were used for the anions. If any distance is within the cutoff, the entire solvent-solvent interaction was included. Adjustments to the allowed ranges for rotations, translations, and dihedral angle movements led to overall acceptance rates of about 30% for new configurations. The ranges for bond stretching and angle bending were set automatically by the BOSS program on the basis of force constants and temperature. All

MC calculations were run on a Linux cluster located at Auburn University and carried out using a custom version of BOSS.<sup>15</sup>

#### 4.4 Results and Discussion

Each condensed phase ionic liquid system was simulated at 25 °C using an NVT ensemble for 20 million MC configurations saving individual coordinates every 40,000 MC steps resulting in 500 configurations. The SP and DSP methods of Wolf et al.<sup>1</sup>, the SF method of Levitt et al.,<sup>7</sup> the DSF method of Fennell and Gezelter,<sup>3</sup> the SFG method of Kale and Herzfeld,<sup>4</sup> the DC method of Zahn et al.,<sup>2</sup> the CHARMM-SF method,<sup>16</sup> and our proposed SF3 method were evaluated through comparison of the energy differences ( $\Delta E$ ) between two conformations as computed by the pairwise alternatives relative to the  $\Delta E$  derived by the full Ewald sums – analogous to the approach of Fennell and Gezelter.<sup>3</sup> A linear squares regression of the energy gap data was carried out and results closest to 1.0 for both the correlation (slope) and correlation coefficient ( $R^2$ ) indicate  $\Delta E$  values identical to those of Ewald. Figure 4.1 gives a sample correlation plot for the [EPyr][BF<sub>4</sub>] ionic liquid system when using DSF ( $\alpha = 0.2 \text{ \AA}^{-1}$ ) and the undamped SF methods. Exhaustive detail for every ionic liquid/pairwise alternative/cutoff combination is provided in the Supporting Information.



**Figure 4.1** Least squares regressions of the configurational energy differences for the [EPyr][BF<sub>4</sub>] system using a 15 Å cutoff distance.

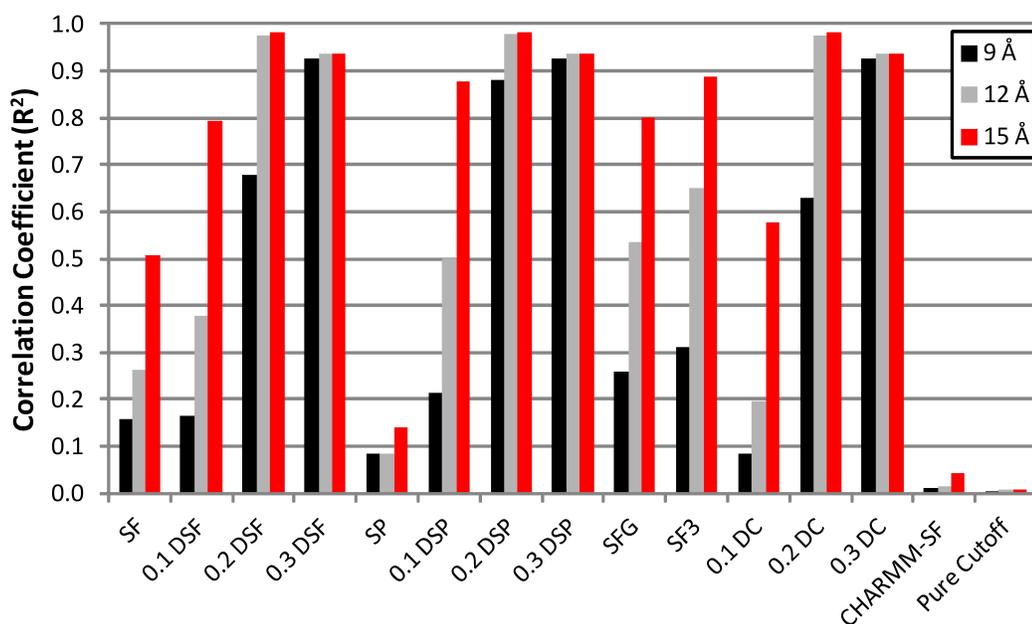
From Table 4.1 (Figure 4.2 provides a graphical representation of the data) it can be seen that the DSF, DSP, and DC methods employing a damping constant value of 0.2 Å<sup>-1</sup> and a cutoff distance of 15 Å provided the best energetic agreement with Ewald simulations. The DSP method did provide a negligibly better correlation coefficient at a cutoff distance of 12 Å and offered a slightly improved algorithm speed (discussed later) compared to DC or DSF. However, any one of the methods can be used interchangeably for the ionic liquid simulations with little difference in accuracy at a 15 Å cutoff. The ideal electrostatic damping value of 0.2 Å<sup>-1</sup> is consistent with previous benchmarking efforts for water and salt solutions. For example, Zahn et al. recommended a combination of a damping constant of 0.2 Å<sup>-1</sup> and a cutoff distance of 9 Å from SPC and TIP3P simulations when using their DC method.<sup>2</sup> Fennell and Gezelter also reported the damping constant of 0.2 Å<sup>-1</sup> to be the best value for the DSP method in Monte Carlo simulations of large chemical systems.<sup>3</sup> While the SF3 method did not perform at the same level of accuracy as the more costly damped methods, it can become a viable alternative for specific ionic liquid combinations. For instance, the [MPyr]-based ionic

liquids produced a average slope and correlation coefficient of 0.981 and 0.986, respectively, for all anions using SF3 compared to 0.987 and 0.992 for the DSP method damped by a value of  $0.2 \text{ \AA}^{-1}$  (see the Supporting Information). Specific cation and anion combinations and additional smoothing of the final  $0.5 \text{ \AA}$  before the cutoff further improved the differences between the SF3 and DSP methods (see below).

**Table 4.1** Statistical analysis of the quality of the configurational energy differences for a given electrostatic method applied across all ionic liquid simulations compared to the reference Ewald sum.<sup>a</sup>

Cutoff	9 Å		12 Å		15 Å		15 Å (smoothed) <sup>b</sup>	
	Slope	$R^2$	Slope	$R^2$	Slope	$R^2$	Slope	$R^2$
Method <sup>c</sup>								
SF	1.003	0.156	0.973	0.264	0.988	0.507	1.008	0.689
0.1 DSF	1.006	0.166	0.982	0.378	0.994	0.794	1.002	0.894
0.2 DSF	0.992	0.680	0.982	0.975	0.980	0.982	0.980	0.981
0.3 DSF	0.955	0.928	0.947	0.938	0.945	0.937	0.944	0.933
SP	1.004	0.084	1.032	0.083	1.016	0.141	0.971	0.239
0.1 DSP	1.006	0.213	1.006	0.502	1.000	0.878	0.995	0.930
0.2 DSP	0.992	0.882	0.982	0.978	0.980	0.982	0.980	0.981
0.3 DSP	0.955	0.928	0.947	0.938	0.945	0.937	0.944	0.933
SFG	0.970	0.258	0.995	0.535	0.989	0.801	0.977	0.883
SF3	0.954	0.313	0.949	0.652	0.971	0.888	0.976	0.936
0.1 DC	1.006	0.085	1.048	0.196	1.003	0.578	0.987	0.693
0.2 DC	0.992	0.631	0.983	0.975	0.980	0.982	0.980	0.981
0.3 DC	0.955	0.927	0.947	0.938	0.945	0.937	0.944	0.933
CHARMM-SF	1.108	0.012	0.855	0.014	0.971	0.044	1.102	0.096
Pure Cutoff	0.725	0.005	1.162	0.009	0.958	0.007	0.979	0.007

<sup>a</sup> Results closest to 1.0 indicate  $\Delta E$  values identical to those of Ewald. The correlation is indicated by the slope, and the correlation coefficient by  $R^2$  in the regression. <sup>b</sup> Quadratic feathering was applied to the final  $0.5 \text{ \AA}$  of the cutoff sphere. <sup>c</sup>  $\alpha$  values of 0.1, 0.2, and  $0.3 \text{ \AA}^{-1}$  used for the DSF, DSP, and DC methods.



**Figure 4.2** Graphical representation of the correlation coefficient ( $R^2$ ) of the configurational energy differences for a given electrostatic method applied across all ionic liquid simulations at three cutoff distances -- 9, 12, and 15 Å. Correlation coefficients closest to 1.0 indicate  $\Delta E$  values most similar to the reference Ewald sum.

The overall agreement between the pairwise alternatives and Ewald sums was generally poorer for ionic liquids containing hexyl and octyl substituents. Not surprisingly, removing the results for the hexyl and octyl based ionic liquids from Table 4.1 improves the average correlation with Ewald sums. For example, the DSP method ( $\alpha = 0.2$  Å) improves its average slope and  $R^2$  values to 0.985 and 0.990, respectively, at 15 Å as compared to the 0.980 and 0.982 values from Table 4.1. The SF3 method also improves considerably to 0.977 and 0.956 without the octyl or hexyl containing [RMIM] and [RPyr] cations relative to the 0.971 and 0.888 values reported in Table 4.1. Increasing the

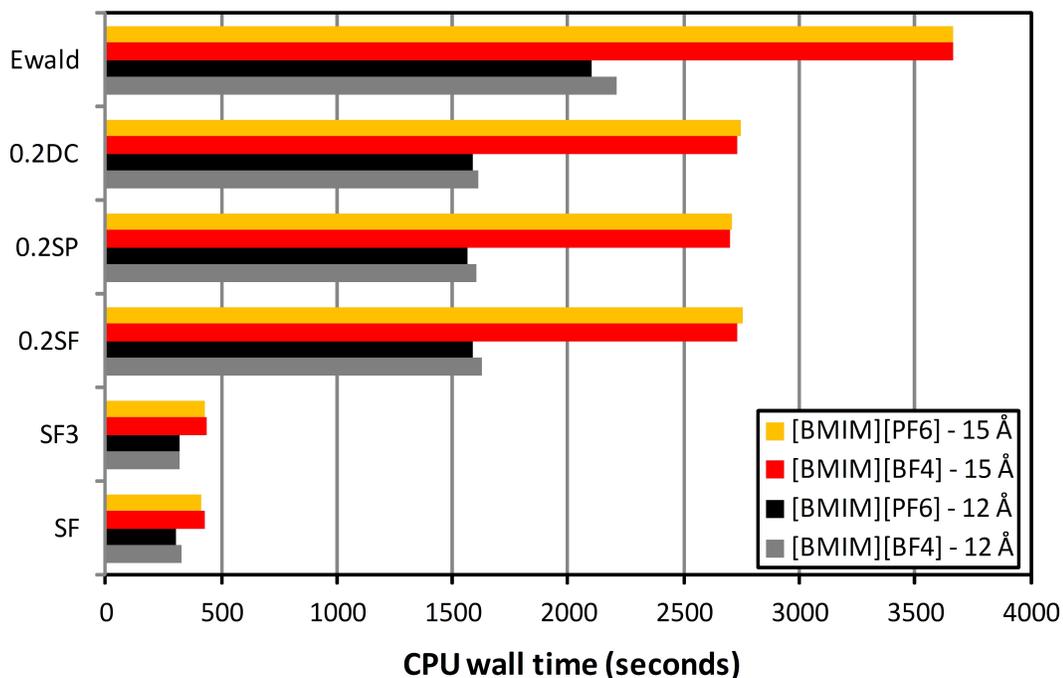
cutoff distance could potentially improve the energetic agreement for the longer alkyl-chain compounds. Detailed tables comparing the differences can be found in the Supporting Information.

## 4.5 Additional smoothing

In OPLS-AA simulations, quadratic feathering of the electrostatic interactions to zero over the last 0.5 Å before the cutoff is typically applied to soften the energy discontinuity.<sup>26</sup> Whereas this option in the BOSS program<sup>15</sup> was not applied to the pairwise alternatives reported in Figure 2.1, it was investigated and reported for 15 Å in Table 4.1. The additional smoothing effect on the average correlation (slope) and correlation coefficient ( $R^2$ ) values across all ionic liquid simulations and cutoff distances are given in the Supporting Information. For the methods with the most favorable agreement with full Ewalds sums, i.e., DSF, DSP, and DC all damped by a value of 0.2 Å<sup>-1</sup>, the feathering did not improve the results. However, for the undamped SF, SP, SFG, and SF3 methods, the agreement with Ewald were all enhanced when applying the additional smoothing. For example, smoothing the SF3 method considerably improved the average slope and  $R^2$  values to 0.976 and 0.936, respectively, compared to the previous values of 0.971 and 0.888 (Table 4.1) without the quadratic feathering. Many specific combinations of ionic liquids were particularly accurate using the SF3 with additional smoothing. For instance, SF3 with a 15 Å cutoff distance on [MMIM][BF<sub>4</sub>] gave slope and  $R^2$  values of 0.984 and 0.989, respectively, [EPyr][NO<sub>3</sub>] had values of 0.984 and 0.988, and [MPyr][TFO] produced values of 0.987 and 0.991.

## 4.6 Timings

Part of the motivation in utilizing pairwise alternatives to Ewald sums was to improve the efficiency of the ionic liquid simulations. Therefore, a basic effort has been made to measure differences in algorithm speed on the same computer: a Dell Optiplex GX760 featuring an Intel Core2Quad processor clocked at 2.83 GHz with 2 GB of RAM. A comparison was carried out for two ionic liquids, [BMIM][PF<sub>6</sub>] and [BMIM][BF<sub>4</sub>], between the damped pairwise alternatives DC, DSP, and DSF (where  $\alpha = 0.2 \text{ \AA}^{-1}$ ), and the undamped SF3 and SF methods. All pairwise interaction models used 12 and 15  $\text{\AA}$  cutoff distances to evaluate the 500 configurations. Figure 4.3 shows the absolute CPU time utilized in seconds for each technique. Relative timings find that the damped methods required ca. 72 to 75 % of the total Ewald sums time at each specific cutoff value. The undamped SF3 and SF systems utilized ca. 15 and 12 % for the 12 and 15  $\text{\AA}$  cutoff distances, respectively (see Supporting Information). The modest gain in simulation speed for the damped pairwise alternatives that employ the erfc function is consistent with the reported time-consuming nature of the erfc function.<sup>11</sup> The undamped SF3 and SF do provide a significant improvement in the calculation time relative to both full Ewald sums and the damped methods, but ultimately at a cost of reduced accuracy (Table 4.1).



**Figure 4.3** Timings for ionic liquid simulations relative to the full Ewald sums.

## 4.7 Conclusions

Monte Carlo simulations of 59 unique room temperature ionic liquid combinations of [RMIM] and [RPy] cations (where R = M (methyl), E (ethyl), B (butyl), H (hexyl), and O (octyl)) and Cl<sup>-</sup>, PF<sub>6</sub><sup>-</sup>, BF<sub>4</sub><sup>-</sup>, NO<sub>3</sub><sup>-</sup>, AlCl<sub>4</sub><sup>-</sup>, Al<sub>2</sub>Cl<sub>7</sub><sup>-</sup>, and TfO<sup>-</sup> anions have been carried out using multiple pairwise electrostatic interaction alternatives (i.e., SF, DSF, SFG, SF3, SP, DSP, DC, and CHARMM-SF) to Ewald sums. The DSP, DSF, or DC methods provided the best energetic agreement with Ewald when combined with an electrostatic damping value of 0.2 Å<sup>-1</sup> and a cutoff distance of 15 Å. The results are consistent with the recommendation by Zahn et al. of an ideal damping constant of 0.2 Å<sup>-1</sup> and a cutoff of 9 Å from SPC and TIP3P simulations when using their DC method.<sup>2</sup>

Fennell and Gezelter also reported  $\alpha = 0.2 \text{ \AA}^{-1}$  to be the best value for the DSP method in Monte Carlo simulations of liquid and solid water and NaCl solutions.<sup>3</sup> Identical agreement with the full Ewald sums energies, i.e., correlation (slope) and correlation coefficient ( $R^2$ ) values of 1.0, could potentially be achieved by fine-tuning the damping constant towards specific ionic liquid combinations. For example, Shi and Maginn reported good reproducibility of the Ewald electrostatics for the [HMIM][Tf<sub>2</sub>N] ionic liquid by using the DSF method with an  $\alpha$  value of  $0.2022 \text{ \AA}^{-1}$ .<sup>17</sup>

Timings of the individual algorithms found that the damped pairwise methods, i.e., DSP, DSF, and DC, delivered a modest 1.3 to 1.4-fold improvement in calculation speed relative to Ewald simulations. The results are consistent with the reported time-consuming nature of the erfc function used to damp the electrostatics.<sup>11</sup> The undamped methods, e.g., SF and SF3, provided a more significant 8 to 9-fold speed-up, however, the accuracy was generally reduced. Nevertheless, the SF3 method can be a viable alternative to the more costly damped methods for specific ionic liquid combinations, as the  $\Delta E$  differences computed for the methyl and ethyl substituted [RMIM] and [RPy] systems rivaled that of the damped methods. In addition, quadratic feathering of the electrostatic interactions to zero over the last  $0.5 \text{ \AA}$  before the cutoff improved the overall agreement for the undamped methods across all ionic liquids relative to the full lattice summation energies. This extensive study provides evidence that simple pairwise interaction alternatives to Ewald can be an efficient and accurate method for computing electrostatic energies in the simulation of diverse room temperature ionic liquids.

## 4.8 References

- (1) Wolf, D.; Keblinski, P.; Phillpot, S. R.; Eggebrecht, J. *Journal of Chemical Physics* **1999**, *110*, 8254–8282.
- (2) Zahn, D.; Schilling, B.; Kast, S. M. *Journal of Physical Chemistry B* **2002**, *106*.
- (3) Fennell, C. J.; Gezelter, J. D. *Journal of Chemical Physics* **2006**, *124*, 234104.
- (4) Kale, S.; Herzfeld, J. *Journal of Chemical Theory and Computation* **2011**, *7*, 215–231.
- (5) Piana, S.; Lindorff-Larsen, K.; Dirks, R. M.; Salmon, J. K.; Dror, R. O.; Shaw, D. E. *PLoS One* **2012**, *7*, e39918.
- (6) Brooks III, C. L.; Pettitt, B. M.; Karplus, M. *Journal of Chemical Physics* **1985**, *83*, 5897–5908.
- (7) Levitt, M.; Hirshberg, M.; Sharon, R.; Daggett, V. *Computational Physics Communications* **1995**, *91*, 215–291.
- (8) Hansen, J. S.; Schroder, T. B.; Dyre, J. C. *Journal of Physical Chemistry B* **2012**, *116*, 5738–5743.
- (9) Vreede, J.; Wolf, M. G.; Leeuw, S. W. d.; Bolhuis, P. G. *Journal of Physical Chemistry B* **2009**, *113*, 6484–6494.
- (10) Mahadevan, T. S.; Garofalini, S. H. *Journal of Physical Chemistry B* **2007**, *111*, 8919–8927.
- (11) Yonezawa, Y. *Journal of Chemical Physics* **2012**, *136*, 244103.
- (12) Kitchen, D. B.; Hirata, F.; Westbrook, J. D.; Levy, R.; Kofke, D.; Yarmush, M. *Journal of Computational Chemistry* **1990**, *11*, 1169–1180.
- (13) Sambavarao, S. V.; Acevedo, O. *Journal of Chemical Theory and Computation* **2007**, *3*, 1038–1050.
- (14) Acevedo, O.; Jorgensen, W. L.; Evanseck, J. D. *Journal of Chemical Theory and Computation* **2007**, *3*, 132–138.
- (15) Jorgensen, W. L.; Tirado-Rives, J. *Journal of Computational Chemistry* **2005**, *26*, 1689–1700.

- (16) Steinbach, P. J.; Brooks, B. R. *Journal of Computational Chemistry* **1994**, *15*, 667–683.
- (17) Shi, W.; Maginn, E. J. *Journal of Physical Chemistry B* **2008**, *112*.

## Chapter 5 Solution Phase Tautomerization Thermodynamics

### 5.1 Abstract

The experimental free energies of tautomerization of seven gas phase protomers have been used to benchmark the accuracy of selected composite thermochemical models, density functionals, and semi-empirical molecular orbital theories, and Hartree-Fock. The G4 composite model and the M06 density functional with a 6-31+G(d,p) basis set yielded the most accurate results, with mean unsigned errors (MUE) of 0.93 and 0.73 kcal·mol<sup>-1</sup>, respectively. Using these two levels of theory to represent the solute, 50 experimentally determined condensed phased free energies of tautomerization were used to benchmark the continuum solvation models IEFPCM, CPCM, and SMD. Differing atomic radii used in the construction of the solute cavity were also tested in conjunction with the solvation models. Depending on the solvation model and cavity chosen, ~1 kcal·mol<sup>-1</sup> MUE can be attained when the tautomers are partitioned into their respective molecular types, i.e. monocarbonyls,  $\beta$ -dicarbonyls, and heterocycles, and again into aqueous and nonaqueous solvents.

These findings are then used to study the mechanism of the enolization of 2-nitrocyclohexanone and to gauge the accuracy of the calculated free energies of activation. Mixed quantum mechanical and quantum mechanical calculations (QM/MM) coupled with free energy perturbation and Metropolis Monte Carlo statistical mechanics (MC/FEP) were employed for comparison.

## **5.2 Benchmarking Continuum Solvation Models and Cavities for the Accurate Calculation of Condensed Phase Free Energy of Tautomerization**

### **5.2.1 Introduction**

In this study, the strategy to find the best method for the calculation of condensed phase free energy of tautomerization was to first find the method that most accurately calculated the free energy of tautomerization in the gas phase as compared to experiment. The theoretical methods employed include five composite thermochemical models CBS-Q<sup>1</sup>, CBS-4M<sup>2</sup>, CBS-QB3<sup>2</sup>, CBS-APNO<sup>3</sup>, and G4<sup>4,5</sup>, semi-empirical quantum mechanical models AM1<sup>6</sup>, PDDG/PM3<sup>7,8</sup>, and PM6<sup>9</sup>, Hartree-Fock, and density functionals B3LYP<sup>10</sup>, X3LYP<sup>11</sup>, M05<sup>12</sup>, M05-2X<sup>12</sup>, M06<sup>13</sup>, M06-2X<sup>13</sup>,  $\omega$ B97X-D<sup>14</sup>, and PBEh1PBE<sup>15</sup>.

Experimental gas phase free energies of tautomerization  $\Delta G_{gas}$  include 2-nitrocyclohexanone<sup>16</sup> (2NCH), acetylacetone<sup>17</sup> (AcAc), ethyl acetoacetate<sup>17</sup> (EAA), methyl acetoacetate<sup>18</sup> (MAA), acetylacetamide<sup>19</sup> (AAM), 2-pyridone<sup>20</sup> (2Pyr), and 6-chloro-2-pyridone<sup>21</sup> (6ClPyr). In cases where the equilibrium constant  $K_T$  was given,  $\Delta G_{gas}$  was calculated by equation 5.1.1

$$K_T = \frac{[enol]}{[keto]} = \exp\left(-\Delta G/k_B T\right) \quad (5.1.1)$$

where  $K_T$  is the equilibrium constant, brackets [ ] denote the molar concentration of the species,  $\Delta G = G_{enol} - G_{keto}$ ,  $k_b$  is the Boltzmann constant, and  $T$  is the absolute temperature.

Protomers calculated in solution include the following: (1) aqueous aliphatic ketones 2-methylpropan-2-one<sup>22</sup>, 3-methylbutan-2-one<sup>22</sup>, butan-2-one<sup>22</sup>, acetaldehyde<sup>22</sup>, acetone<sup>22</sup>, and propan-1-one<sup>22</sup>; (2)  $\beta$ -dicarbonyl and  $\beta$ -dicarbonyl-like molecules acetylacetone<sup>22,23</sup>, 2-acetylcyclohexanone<sup>24</sup>, 2-nitrocyclohexanone<sup>16,25,26</sup>, ethyl acetoacetate<sup>23</sup>, acetoacetamide<sup>18,27</sup>, 2-methylcyclohexa-1,3-dione<sup>23</sup>, and tetrahydroisobenzofuran-1,7-dione<sup>23</sup>, and (3) heterocycles 2-pyridone<sup>28</sup>, 6-chloropyridone<sup>21</sup>, 2-thiopyridone<sup>21</sup>, 6-chloro-2-thiopyridone<sup>21</sup>, 2-chloro-4-pyridone<sup>21</sup>, cytosine<sup>29</sup>, 3-methylcytosine<sup>29</sup>, adenine<sup>29</sup>, and 1-methyladenine<sup>29</sup>. These protomers were calculated in solvents for which there were experimental data, including water, chloroform, dichloromethane, carbon tetrachloride, acetonitrile, and cyclohexane.

## 5.2.2 Methods

For Hartree-Fock and density functional methods, nine basis sets have been used, i.e. 6-31+G(d), 6-31+G(d,p), 6-311++G(2d,p), 6-311++G(3df,3pd), cc-pVDZ, cc-pVTZ, cc-pVQZ, aug-cc-pVDZ, and aug-cc-pVTZ. Vibrational calculations have been performed to verify that all resultant geometries were minima on the potential energy surface<sup>30</sup>, possessed no imaginary frequencies (negative force constants) and provided thermal contributions to the electronic energy. The CBS-APNO method does not

include chlorine as implemented in Gaussian 09.<sup>31</sup> Therefore its use could not be extended to the many solution phase chloro-substituted tautomers.

Solution phase calculations utilized the continuum solvation models IEFPCM<sup>32</sup>, CPCM<sup>33</sup>, and SMD<sup>34</sup> as implemented in Gaussian09. Reviews on continuum solvation models are available in the literature.<sup>35–38</sup> Molecular energies and other properties are dependent upon the choice of atomic radii used to build the cavity into which the solute is placed.<sup>32,36</sup> Therefore, all cavity types included in Gaussian09, i.e. BONDI<sup>39</sup>, PAULING<sup>40</sup>, UA0, UAHF<sup>41</sup>, UAKS, and the UFF<sup>42</sup> have been benchmarked. COULOMB atomic radii were optimized specifically for use with the SMD model.<sup>34</sup> However, the other cavities have been tested with SMD as well. Examples of radii for atoms constituting the molecules in this study are in table

**Table 5.1** Radii of atomic spheres for each cavity type used in the construction of the solute cavities. Units given in Å.

Radii of Atomic Spheres used in this Study (Å)						
	C	O	N	Cl	S	H
BONDI	1.70	1.52	1.55	1.75	1.80	1.20
PAULING	1.50	1.40	1.50	1.80	2.26	1.20
UA0	2.00	1.80	1.90	1.97	2.22	--
UAHF	1.80	1.50	1.60	1.98	2.11	--
UAKS	1.80	1.50	1.60	1.98	2.11	--
UFF	1.93	1.75	1.83	1.97	2.02	1.44
COULOMB	1.85	2.02	1.89	2.38	2.49	1.20

## 5.2.3 Results and Discussion

### 5.2.3.1 Gas Phase

The Hartree-Fock and density functional methods are averaged over eight of the nine basis sets; 6-31+G(d) consistently gave qualitatively different results than the other basis sets, predicting the predominance of opposite tautomer than all other basis sets used. This behavior demonstrates the necessity of *p* polarization functions on hydrogen in order to characterize the intramolecular hydrogen bonding present within many of the tautomers. The MUEs of the calculated gas phase  $\Delta G$  of tautomerization are tabulated in table 5.2.

**Table 5.2** Unsigned errors in calculated  $\Delta G_{gas}$  of Tautomerization. Negative  $\Delta G$  values indicate preference of the enol form. Values given in kcal·mol<sup>-1</sup>.

Unsigned Errors in Calculated $\Delta G_{gas}$ of Tautomerization								
Molecule	2NCH <sup>a</sup>	AcAc <sup>b</sup>	EAA <sup>b</sup>	MAA <sup>c</sup>	2Pyr <sup>d</sup>	6ClPyr <sup>e</sup>	AAM <sup>f</sup>	
Experimental $\Delta G_{gas}$	-3.53	-2.20	-0.08	-0.85	-0.81	-2.10	-0.37	
Method	Unsigned Errors							Mean
CBS-QB3	2.83	0.68	0.92	0.32	0.47	1.55	0.79	1.08
CBS-Q	4.09	0.94	3.23	1.74	0.57	0.25	0.77	1.66
CBS-APNO	2.52	0.15	0.77	0.04	0.35	--	0.94	0.79
CBS-4M	5.73	2.58	0.93	1.65	1.18	2.38	2.49	2.42
<b>G4</b>	3.72	0.07	0.99	0.26	0.13	1.21	0.16	<b>0.93</b>
Hartree-Fock	6.49	3.62	2.85	3.27	0.32	1.78	2.99	3.05
B3LYP	1.07	2.94	2.66	1.95	1.27	0.27	1.37	1.65
X3LYP	1.20	2.84	2.85	2.13	1.13	0.46	1.51	1.73
M05	2.41	1.44	0.90	1.09	2.02	0.41	2.23	1.50
M05-2X	0.93	3.77	3.91	3.53	1.06	2.48	2.93	2.66
<b>M06</b>	0.33	1.48	1.52	0.97	1.03	0.51	0.31	<b>0.88</b>
M06-2X	0.30	1.45	2.60	2.33	0.99	2.32	1.23	1.60
wB97XD	1.61	0.98	2.52	1.02	1.16	0.46	0.42	1.17
PBEh1PBE	2.15	3.60	3.62	2.93	0.86	0.62	2.13	2.27
AM1	2.15	3.23	1.48	2.46	0.25	1.66	0.73	1.71
PDDG/PM3	2.13	8.50	5.16	5.94	2.59	1.11	3.19	4.09
PM6	3.73	0.70	2.71	1.13	6.54	9.00	6.10	4.27

- a) Angelini, G et al. *The Journal of Organic Chemistry* **2005**, *70*, 8193–8196.  
b) Folkendt, M. M. et al. *The Journal of Physical Chemistry* **1985**, *89*, 3347–3352.  
c) Belova, N. V; Oberhammer, H.; Girichev, G. V *The Journal of Physical Chemistry A* **2004**, *108*, 3593–3597.  
d) Beak, P.; Fry, F. S. *Journal of the American Chemical Society* **1973**, *95*, 1700–1702.  
e) Beak, P.; Covington, J. B.; White, J. M. *The Journal of Organic Chemistry* **1980**, *45*, 1347–1353.  
f) Caminati, W.; Grabow, J.-U. *Journal of the American Chemical Society* **2006**, *128*, 854–7.

As can be seen in the table 5.1, CBS-APNO and M06 gave the lowest MUE's of 0.79 and 0.88 kcal·mol<sup>-1</sup>, respectively. However, since CBS-APNO cannot be applied to many of the tautomers calculated in solution, the second best thermochemical model, G4, (MUE of 0.93 kcal·mol<sup>-1</sup>) was used hereafter. Of note is the large unsigned error of the thermochemical models associated with 2NCH. Both axial and equatorial conformations of the keto tautomer were calculated, with neither achieving the kilocalorie accuracy ordinarily associated with thermochemical models. 2NCH seems to be a pathological case for these methods and is explored in detail later.

Upon finding the M06 density functional to be the superior non-composite model, the UEs of the individual basis sets were assessed. The 6-31+G(d,p) basis set was found to reproduce experimental values most accurately. (Table 5.3)

**Table 5.3** Unsigned errors of basis sets using the M06 density functional. Values given in kcal·mol<sup>-1</sup>.

MUE's of Basis Sets Using the M06 Density Functional								
Basis Set	2NCH	AcAc	EAA	MAA	2Pyr	6Cl2Pyr	AAM	MUE
<b>6-31+G(d,p)</b>	0.07	1.29	0.84	0.97	0.94	0.77	0.22	<b>0.73</b>
6-311++G(2d,p)	0.12	1.53	1.35	0.91	0.97	0.48	0.19	0.79
6-311++G(3df,3pd)	0.35	1.41	1.69	0.60	1.24	0.06	0.04	0.77
cc-pVDZ	0.56	2.14	1.47	0.90	0.68	1.06	0.48	1.04
cc-pVTZ	0.27	1.25	1.49	1.10	1.28	0.24	0.02	0.81
cc-pVQZ	0.66	0.83	1.57	0.73	1.34	0.18	0.47	0.83
aug-cc-pVDZ	0.61	1.89	1.76	1.63	0.49	1.24	0.61	1.17
aug-cc-pVTZ	0.05	1.52	1.96	0.92	1.29	0.02	0.43	0.88

The M06/6-31+G(d,p) yields a MUE of 0.73 kcal·mol<sup>-1</sup>. Hence, the G4 and M06/6-31+G(d,p) methods were used in calculating condensed phase free energies of tautomerization.

#### **5.2.3.2 Condensed Phase**

The mean unsigned errors (MUE), maximum unsigned errors (MAX), and standard deviations in the unsigned error (STDDEV) are tabulated in tables 5.4 through 5.13. The results are partitioned into all molecules and solvents considered together, aqueous monocarbonyls only,  $\beta$ -dicarbonyls only, and heterocycles only. These are then divided into all solvents, aqueous, and non-aqueous solvents. All values are given in kcal·mol<sup>-1</sup>.

**Table 5.4.** Statistical analysis of unsigned errors across all molecules and all solvents. Values given in kcal·mol<sup>-1</sup>.

<b>M06/6-31+G(d,p)</b>				<b>G4</b>			
ALL TAUTOMERS in ALL SOLVENTS							
IEFPCM	MUE	MAX	STDDEV	IEFPCM	MUE	MAX	STDDEV
BONDI	1.90	5.47	1.66	BONDI	1.74	5.80	1.26
PAULING	2.07	7.04	1.93	PAULING	1.96	5.92	1.36
UA0	1.98	6.67	1.36	<b>UA0</b>	<b>1.57</b>	<b>5.86</b>	<b>1.11</b>
UAHF	2.37	6.60	1.74	UAHF	2.61	8.31	1.96
UAKS	3.16	9.17	2.46	UAKS	2.85	8.31	2.44
UFF	1.99	5.70	1.49	UFF	1.71	5.10	1.22
CPCM				CPCM			
BONDI	1.86	5.17	1.59	BONDI	1.80	5.85	1.31
PAULING	2.08	6.15	1.90	PAULING	2.05	5.98	1.40
UA0	2.00	6.70	1.42	UA0	1.61	5.89	1.11
UAHF	2.39	6.69	1.82	UAHF	2.57	7.37	1.63
UAKS	3.41	9.94	2.76	UAKS	2.96	8.22	2.39
UFF	1.99	5.73	1.53	UFF	1.75	5.13	1.22
SMD				SMD			
BONDI	2.19	5.51	1.61	BONDI	2.07	8.72	1.64
PAULING	2.44	6.89	1.94	PAULING	2.30	7.51	1.59
UA0	2.09	5.02	1.38	UA0	1.77	5.15	1.32
UAHF	2.62	7.15	2.09	UAHF	3.01	8.46	2.16
UAKS	3.34	9.68	2.59	UAKS	3.16	9.25	2.43
UFF	2.10	5.52	1.52	UFF	1.91	6.47	1.48
COULOMB	2.05	5.11	1.58	COULOMB	1.84	7.98	1.47

**Table 5.5.** Statistical analysis of errors across all molecules in water. Values given in kcal·mol<sup>-1</sup>.

<b>M06/6-31+G(d,p)</b>				<b>G4</b>			
ALL TAUTOMERS in WATER							
IEFPCM	MUE	MAX	STDDEV	IEFPCM	MUE	MAX	STDDEV
BONDI	2.17	5.47	1.80	BONDI	2.02	5.80	1.41
PAULING	2.33	7.04	2.26	PAULING	2.20	5.92	1.49
UA0	2.31	6.67	1.46	UA0	1.95	5.86	1.33
<b>UAHF</b>	<b>1.77</b>	<b>5.26</b>	<b>1.50</b>	UAHF	2.39	4.99	1.61
UAKS	2.32	6.35	2.02	UAKS	2.18	5.13	1.66
UFF	2.68	5.70	1.66	UFF	2.48	5.10	1.27
CPCM				CPCM			
BONDI	2.04	5.17	1.68	BONDI	2.07	5.85	1.43
PAULING	2.28	6.15	2.14	PAULING	2.22	5.98	1.49
UA0	2.29	6.70	1.47	UA0	1.93	5.89	1.30
UAHF	1.80	5.33	1.54	UAHF	2.41	5.07	1.66
UAKS	2.45	6.51	2.12	UAKS	2.23	5.27	1.72
UFF	2.68	5.73	1.66	UFF	2.51	5.13	1.26
SMD				SMD			
BONDI	2.49	5.43	1.50	BONDI	2.36	8.72	2.03
PAULING	2.71	6.89	2.10	PAULING	2.59	7.51	1.82
UA0	2.35	4.81	1.31	UA0	2.18	5.15	1.62
UAHF	2.09	6.09	2.09	UAHF	2.96	6.39	2.00
UAKS	2.39	5.63	1.87	UAKS	2.42	6.05	1.50
UFF	2.77	5.52	1.51	UFF	2.69	6.47	1.75
COULOMB	2.41	5.00	1.52	COULOMB	2.17	7.98	1.78

**Table 5.6.** Statistical analysis of errors across all molecules in nonaqueous solvents. Values given in kcal·mol<sup>-1</sup>.

<b>M06/6-31+G(d,p)</b>				<b>G4</b>			
ALL TAUTOMERS in NONAQUEOUS SOLVENTS							
IEFPCM	MUE	MAX	STDDEV	IEFPCM	MUE	MAX	STDDEV
BONDI	1.69	4.97	1.50	BONDI	1.53	3.47	1.08
PAULING	1.87	5.31	1.60	PAULING	1.77	3.98	1.21
UA0	1.73	4.51	1.22	UA0	1.28	2.81	0.81
UAHF	2.81	6.60	1.78	UAHF	2.77	8.31	2.17
UAKS	3.76	9.17	2.58	UAKS	3.34	8.31	2.77
UFF	1.45	3.69	1.08	<b>UFF</b>	<b>1.11</b>	<b>2.59</b>	<b>0.77</b>
CPCM				CPCM			
BONDI	1.73	5.06	1.51	BONDI	1.59	3.54	1.17
PAULING	1.94	5.43	1.67	PAULING	1.91	4.10	1.30
UA0	1.77	4.73	1.34	UA0	1.35	3.03	0.85
UAHF	2.86	6.69	1.89	UAHF	2.68	7.37	1.60
UAKS	4.16	9.94	2.97	UAKS	3.52	8.22	2.67
UFF	1.46	3.86	1.17	UFF	1.17	2.88	0.80
SMD				SMD			
BONDI	1.96	5.51	1.65	BONDI	1.85	3.98	1.20
PAULING	2.23	5.82	1.78	PAULING	2.07	4.98	1.34
UA0	1.88	5.02	1.40	UA0	1.45	3.38	0.90
UAHF	2.99	7.15	2.00	UAHF	3.05	8.46	2.26
UAKS	4.02	9.68	2.81	UAKS	3.69	9.25	2.80
UFF	1.57	4.18	1.30	UFF	1.30	2.83	0.83
COULOMB	1.78	5.11	1.57	COULOMB	1.58	3.54	1.11

**Table 5.7.** Statistical analysis of unsigned errors for aqueous monocarbonyls.Values given in kcal·mol<sup>-1</sup>.

<b>M06/6-31+G(d,p)</b>				<b>G4</b>			
MONOCARBONYLS in WATER							
	MUE	MAX	STDDEV		MUE	MAX	STDDEV
IEFPCM				IEFPCM			
BONDI	4.51	5.47	0.78	BONDI	2.55	4.31	0.85
PAULING	5.63	7.04	0.94	PAULING	3.36	5.00	0.78
UA0	3.24	4.21	0.70	UA0	1.50	3.48	0.96
UAHF	1.01	1.85	0.65	UAHF	2.39	3.28	1.12
<b>UAKS</b>	<b>0.83</b>	<b>1.44</b>	<b>0.40</b>	UAKS	2.18	2.99	0.92
UFF	4.68	5.60	0.70	UFF	2.87	4.40	0.77
CPCM				CPCM			
BONDI	4.30	5.17	0.71	BONDI	2.56	4.32	0.86
PAULING	5.39	6.15	0.70	PAULING	3.35	4.89	0.75
UA0	3.21	4.15	0.70	UA0	1.48	3.47	0.96
UAHF	1.06	1.96	0.70	UAHF	2.46	3.38	1.18
UAKS	0.85	1.55	0.46	UAKS	2.25	3.09	0.97
UFF	4.68	5.59	0.68	UFF	2.90	4.40	0.78
SMD				SMD			
BONDI	4.49	5.43	0.79	BONDI	2.31	3.69	0.66
PAULING	5.57	6.89	0.92	PAULING	2.96	3.71	0.43
UA0	3.20	3.94	0.65	UA0	1.15	2.71	0.80
UAHF	0.90	1.93	0.74	UAHF	2.64	3.39	0.90
UAKS	0.87	1.52	0.47	UAKS	2.33	3.10	0.92
UFF	4.42	5.52	0.84	UFF	2.45	3.14	0.54
COULOMB	4.18	5.00	0.75	COULOMB	1.80	2.62	0.44

**Table 5.8.** Statistical analysis of unsigned errors for  $\beta$ -dicarbonyls in all solvents.Values given in kcal·mol<sup>-1</sup>.

<b>M06/6-31+G(d,p)</b>				<b>G4</b>			
$\beta$ -DICARBONYLS in ALL SOLVENTS							
	MUE	MAX	STDDEV		MUE	MAX	STDDEV
IEFPCM				IEFPCM			
BONDI	1.07	4.25	1.07	BONDI	1.54	4.00	1.33
PAULING	1.20	4.26	1.07	PAULING	1.87	3.98	1.29
UA0	1.69	6.67	1.37	UA0	1.58	5.86	1.26
UAHF	2.80	6.60	1.94	UAHF	3.35	8.31	2.76
UAKS	2.91	6.66	1.94	UAKS	3.27	8.31	2.87
UFF	1.44	5.70	1.22	UFF	1.40	5.10	1.17
CPCM				CPCM			
BONDI	1.06	4.31	1.04	BONDI	1.63	4.04	1.42
PAULING	1.24	4.40	1.11	PAULING	2.01	4.10	1.38
UA0	1.61	6.70	1.39	UA0	1.58	5.89	1.25
UAHF	2.50	6.69	2.07	UAHF	2.97	7.37	2.12
UAKS	2.66	6.72	2.21	UAKS	2.92	7.41	2.25
UFF	1.35	5.73	1.23	UFF	1.45	5.13	1.17
SMD				SMD			
BONDI	1.14	3.67	0.83	BONDI	1.57	3.91	1.23
PAULING	1.47	5.27	1.25	PAULING	1.91	4.98	1.35
UA0	1.43	3.27	0.94	UA0	1.37	2.53	0.70
UAHF	3.10	7.15	2.09	UAHF	3.97	8.46	2.59
UAKS	3.25	7.36	2.18	UAKS	3.81	9.25	2.64
UFF	1.21	3.16	0.96	UFF	1.33	2.83	0.70
<b>COULOMB</b>	<b>0.95</b>	<b>3.01</b>	<b>0.74</b>	<b>COULOMB</b>	1.39	3.51	1.17

**Table 5.9.** Statistical analysis of unsigned errors of  $\beta$ -dicarbonyls in water.Values given in kcal·mol<sup>-1</sup>.

<b>M06/6-31+G(d,p)</b>				<b>G4</b>			
<i><math>\beta</math>-DICARBONYLS in WATER</i>							
IEFPCM	MUE	MAX	STDDEV	IEFPCM	MUE	MAX	STDDEV
BONDI	1.42	4.25	1.43	BONDI	1.49	4.00	1.58
PAULING	1.33	3.02	1.09	PAULING	2.09	3.96	1.00
UA0	2.63	6.67	1.98	UA0	2.26	5.86	1.86
UAHF	3.32	5.26	1.93	UAHF	2.86	4.99	2.02
UAKS	3.34	4.69	1.89	UAKS	3.05	5.13	2.16
UFF	2.30	5.70	1.63	UFF	1.67	5.10	1.62
CPCM				CPCM			
BONDI	1.46	4.31	1.44	BONDI	1.73	4.04	1.65
PAULING	1.45	3.18	1.10	PAULING	2.20	4.03	0.97
UA0	2.62	6.70	1.99	UA0	2.17	5.89	1.73
UAHF	3.45	5.33	1.89	UAHF	2.87	5.07	2.11
UAKS	3.76	5.87	2.13	UAKS	3.13	5.27	2.23
UFF	2.29	5.73	1.64	UFF	1.69	5.13	1.61
SMD				SMD			
BONDI	1.43	1.87	0.40	BONDI	1.03	2.80	1.00
PAULING	1.66	3.35	1.17	PAULING	1.73	3.36	1.08
UA0	2.01	3.27	1.07	UA0	1.37	2.06	0.69
UAHF	4.83	5.28	0.40	UAHF	4.39	6.12	1.24
UAKS	5.13	5.63	0.53	UAKS	4.59	6.05	1.04
UFF	1.67	3.16	1.11	UFF	1.18	1.54	0.27
<b>COULOMB</b>	<b>0.96</b>	<b>1.67</b>	<b>0.63</b>	<b>COULOMB</b>	<b>1.03</b>	<b>2.67</b>	<b>1.07</b>

**Table 5.10.** Statistical analysis of unsigned errors for  $\beta$ -dicarbonyls in nonaqueous solvent. Values given in kcal·mol<sup>-1</sup>.

<b>M06/6-31+G(d,p)</b>				<b>G4</b>			
$\beta$ -DICARBONYLS in NONAQUEOUS SOLVENTS							
	MUE	MAX	STDDEV		MUE	MAX	STDDEV
IEFPCM				IEFPCM			
BONDI	0.93	3.08	0.85	BONDI	1.56	3.47	1.19
PAULING	1.14	4.26	1.05	PAULING	1.77	3.98	1.39
UA0	1.31	2.59	0.75	UA0	1.32	2.57	0.80
UAHF	2.67	6.60	1.92	UAHF	3.48	8.31	2.91
UAKS	2.80	6.66	1.94	UAKS	3.33	8.31	3.03
UFF	1.09	2.33	0.77	UFF	1.27	2.59	0.87
CPCM				CPCM			
<b>BONDI</b>	<b>0.90</b>	<b>3.16</b>	<b>0.77</b>	BONDI	1.59	3.54	1.32
PAULING	1.15	4.40	1.10	PAULING	1.92	4.10	1.52
UA0	1.21	2.34	0.73	UA0	1.31	2.80	0.83
UAHF	2.21	6.69	2.03	UAHF	3.00	7.37	2.12
UAKS	2.33	6.72	2.13	UAKS	2.85	7.41	2.25
UFF	0.97	2.06	0.74	UFF	1.33	2.88	0.87
SMD				SMD			
BONDI	1.03	3.67	0.92	BONDI	1.82	3.91	1.25
PAULING	1.39	5.27	1.27	PAULING	1.99	4.98	1.45
UA0	1.19	2.35	0.77	UA0	1.36	2.53	0.71
UAHF	2.67	7.15	2.11	UAHF	3.86	8.46	2.84
UAKS	2.79	7.36	2.18	UAKS	3.58	9.25	2.92
UFF	1.03	2.31	0.82	UFF	1.40	2.83	0.81
COULOMB	0.95	3.01	0.77	COULOMB	1.56	3.51	1.18

**Table 5.11.** Statistical analysis of unsigned errors for heterocycles in all solvents.Values given in kcal·mol<sup>-1</sup>.

<b>M06/6-31+G(d,p)</b>				<b>G4</b>			
HETEROCYCLES in ALL SOLVENTS							
	MUE	MAX	STDDEV		MUE	MAX	STDDEV
IEFPCM				IEFPCM			
BONDI	1.86	4.97	1.50	BONDI	1.70	5.80	1.21
PAULING	1.92	5.31	1.59	PAULING	1.66	5.92	1.31
UA0	1.82	4.51	1.28	<b>UA0</b>	<b>1.58</b>	<b>3.82</b>	<b>1.01</b>
UAHF	2.43	5.59	1.59	UAHF	2.22	4.79	1.32
UAKS	3.92	9.17	2.61	UAKS	2.78	7.51	2.38
UFF	1.66	3.69	1.06	UFF	1.67	4.79	1.18
CPCM				CPCM			
BONDI	1.95	5.06	1.54	BONDI	1.73	5.85	1.25
PAULING	2.00	5.43	1.68	PAULING	1.74	5.98	1.35
UA0	1.94	4.73	1.38	UA0	1.66	3.77	1.01
UAHF	2.59	5.90	1.74	UAHF	2.38	4.80	1.39
UAKS	4.36	9.94	2.88	UAKS	3.16	8.22	2.68
UFF	1.74	3.86	1.14	UFF	1.71	4.74	1.19
SMD				SMD			
BONDI	2.43	5.51	1.55	BONDI	2.27	8.72	1.82
PAULING	2.46	5.82	1.70	PAULING	2.28	7.51	1.62
UA0	2.27	5.02	1.60	UA0	2.10	5.15	1.54
UAHF	2.72	6.11	2.06	UAHF	2.65	6.39	1.84
UAKS	3.99	9.68	2.72	UAKS	3.11	8.08	2.44
UFF	2.17	4.27	1.36	UFF	2.09	6.05	1.69
COULOMB	2.34	5.11	1.53	COULOMB	2.18	7.98	1.74

**Table 5.12.** Statistical analysis of unsigned errors for heterocycles in water.Values given in kcal·mol<sup>-1</sup>.

<b>M06/6-31+G(d,p)</b>				<b>G4</b>			
HETEROCYCLES in WATER							
IEFPCM	MUE	MAX	STDDEV	IEFPCM	MUE	MAX	STDDEV
BONDI	1.10	2.42	0.82	BONDI	2.01	5.80	1.46
PAULING	0.79	1.54	0.49	PAULING	1.62	5.92	1.64
UA0	1.47	2.78	0.84	UA0	2.44	3.82	1.08
UAHF	1.75	4.29	1.34	UAHF	2.41	4.79	1.70
UAKS	2.98	6.35	2.14	UAKS	1.40	5.05	1.75
UFF	1.61	2.23	0.64	UFF	3.04	4.79	1.00
CPCM				CPCM			
BONDI	1.17	2.89	0.91	BONDI	1.94	5.85	1.52
<b>PAULING</b>	<b>0.75</b>	<b>1.57</b>	<b>0.57</b>	PAULING	1.48	5.98	1.68
UA0	1.46	2.78	0.85	UA0	2.07	3.77	1.08
UAHF	1.74	4.32	1.39	UAHF	2.23	4.80	1.73
UAKS	3.08	6.51	2.17	UAKS	1.91	5.15	1.81
UFF	1.60	2.20	0.63	UFF	2.79	4.74	0.98
SMD				SMD			
BONDI	1.86	3.05	0.93	BONDI	2.88	8.72	2.40
PAULING	1.50	3.07	1.07	PAULING	2.63	7.51	2.07
UA0	2.01	4.81	1.52	UA0	3.46	5.15	1.75
UAHF	1.98	6.09	2.21	UAHF	3.44	6.39	2.35
UAKS	2.50	5.30	1.66	UAKS	1.53	4.56	1.18
UFF	2.41	4.27	1.12	UFF	3.95	6.05	1.77
COULOMB	2.04	3.33	1.05	COULOMB	3.22	7.98	2.16

**Table 5.13.** Statistical analysis of unsigned errors for heterocycles in nonaqueous solvents. Values given in kcal·mol<sup>-1</sup>.

<b>M06/6-31+G(d,p)</b>				<b>G4</b>			
HETEROCYCLES in NONAQUEOUS SOLVENTS							
IEFPCM	MUE	MAX	STDDEV	IEFPCM	MUE	MAX	STDDEV
BONDI	2.35	4.97	1.63	BONDI	1.51	3.39	0.98
PAULING	2.64	5.31	1.64	PAULING	1.77	3.78	1.02
UA0	2.04	4.51	1.46	UA0	1.25	2.81	0.81
UAHF	2.86	5.59	1.59	UAHF	2.21	4.03	0.99
UAKS	4.52	9.17	2.72	UAKS	3.35	7.51	2.55
UFF	1.70	3.69	1.26	<b>UFF</b>	<b>0.97</b>	<b>2.07</b>	<b>0.64</b>
CPCM				CPCM			
BONDI	2.45	5.06	1.65	BONDI	1.60	3.50	1.02
PAULING	2.80	5.43	1.67	PAULING	1.90	3.93	1.05
UA0	2.25	4.73	1.56	UA0	1.40	3.03	0.87
UAHF	3.13	5.90	1.72	UAHF	2.48	4.33	1.10
UAKS	5.17	9.94	2.99	UAKS	3.96	8.22	2.83
UFF	1.83	3.86	1.36	UFF	1.02	2.24	0.69
SMD				SMD			
BONDI	2.79	5.51	1.74	BONDI	1.88	3.98	1.16
PAULING	3.08	5.82	1.73	PAULING	2.14	4.35	1.23
UA0	2.43	5.02	1.62	UA0	1.53	3.38	1.04
UAHF	3.20	6.11	1.80	UAHF	2.42	4.60	1.36
UAKS	4.95	9.68	2.84	UAKS	3.78	8.08	2.71
UFF	2.02	4.18	1.47	UFF	1.20	2.63	0.84
COULOMB	2.53	5.11	1.75	COULOMB	1.60	3.54	1.05

It is evident from tables 5.4 through 5.13 that a smaller mean unsigned error is attained upon partitioning the tautomers into their respective types and then again into aqueous and nonaqueous solvents. The most accurate quantum mechanical method, solvation model, and cavity type combination are tabulated in table 5.14.

**Table 5.14.** Summary of statistical analysis of unsigned errors for the most accurate method, solvation model and cavity type. Values given in kcal·mol<sup>-1</sup>.

ALL TAUTOMERS				
<i>All solvents</i>		MUE	MAX	STDDEV
G4	IEFPCM/UA0	1.57	5.86	1.11
<i>Water</i>				
M06/6-31+G(d,p)	IEFPCM/UAHF	1.77	5.26	1.50
<i>Nonaqueous</i>				
G4	IEFPCM/UFF	1.11	2.59	0.77
AQUEOUS MONOCARBONYLS				
<i>Water</i>		MUE	MAX	STDDEV
M06/6-31+G(d,p)	IEFPCM/UAKS	0.83	1.44	0.40
$\beta$ -DICARBONYLS				
<i>All solvents</i>		MUE	MAX	STDDEV
M06/6-31+G(d,p)	SMD/COULOMB	0.95	3.01	0.74
<i>Water</i>				
M06/6-31+G(d,p)	SMD/COULOMB	0.96	1.67	0.63
<i>Nonaqueous</i>				
M06/6-31+G(d,p)	CPCM/BONDI	0.90	3.16	0.77
HETEROCYCLES				
<i>All solvents</i>		MUE	MAX	STDDEV
G4	IEFPCM/UA0	1.58	3.82	1.01
<i>Water</i>				
M06/6-31+G(d,p)	CPCM/PAULING	0.75	1.57	0.57
<i>Nonaqueous</i>				
G4	IEFPCM/UFF	0.97	2.07	0.64

Partitioning the tautomers into their respective molecular class and then again into aqueous or nonaqueous solvents improves the MUE, MAX, and STDDEV, with the exception of nonaqueous  $\beta$ -dicarbonyls whose MAX increases approximately half a kcal·mol<sup>-1</sup>. For example, the most accurate model when averaged across all aqueous

tautomers, SMD/COULOMB/M06/6-31+G(d,p), has a MUE, MAX, and STDDEV of 1.77, 5.26, and 1.50 kcal·mol<sup>-1</sup> respectively. However, when partitioned into aqueous monocarbonyls,  $\beta$ -dicarbonyls, and heterocycles, the MUEs of the best methods reduce to 0.83, 0.96, and 0.75 kcal·mol<sup>-1</sup> respectively. The MAXs and STDDEVs decrease by approximately 3.5 kcal·mol<sup>-1</sup> and 1.0 kcal·mol<sup>-1</sup>, respectively.

A peculiar result of this investigation was that many dicarbonyls in some solvents spontaneously deprotonate when using the UAKS and UAHF cavities. These cavities give the dicarbonyls the smallest solute volumes within the united atom class. An extra sphere was added to the hydrogen on the  $\alpha$ -carbon using the *addsphere* keyword. However, deprotonation still occurred. Calculating the force constants at every step in the geometry optimization using the *calcall* keyword did not correct the erroneous deprotonation either. This effect is not related to the  $pK_a$  value of the solute, as was determined by a calculation on trifluoroacetic acid using the UAKS and UAHF cavities, wherein deprotonation did not occur. This phenomenon is not observed for the monocarbonyls or heterocycles.

Since the relation of cavity size to solute physical properties has been widely recognized, an effort to correlate cavity and surface area percent differences between the tautomers to the accuracy of the method was attempted. However, the analysis leads to no insight. For example, while UAKS/IEFPCM/M06/6-31+G(d,p) yields the most accurate results for aqueous monocarbonyls, the average surface area percent difference is  $2.07 \pm 1.41\%$  while the average volume percent difference is  $2.58 \pm 2.58\%$ . Such large standard deviations preclude meaningful insight into a correlation of cavity volume or surface area to accuracy due to the large variance in the values.

### 5.3 Application to an Enolization Reaction: 2-nitrocyclohexanone

A second goal of this research was to determine the accuracy of the calculated free energies of activation using the most accurate solvation model and cavity determined in the above analysis as well determining the mechanism of enolization. The base-catalyzed keto-enol interconversion of 2-nitrocyclohexanone (2NCH) has been studied in a variety of solvents.<sup>16,25,26,43,44</sup> Energies of activation were determined from reaction rates using the Eyring equation<sup>44,45</sup>

$$\Delta G^\ddagger(T) = RT \left( \ln \frac{T}{k^B} + \ln \frac{k_b}{h} \right) \quad (5.3.1)$$

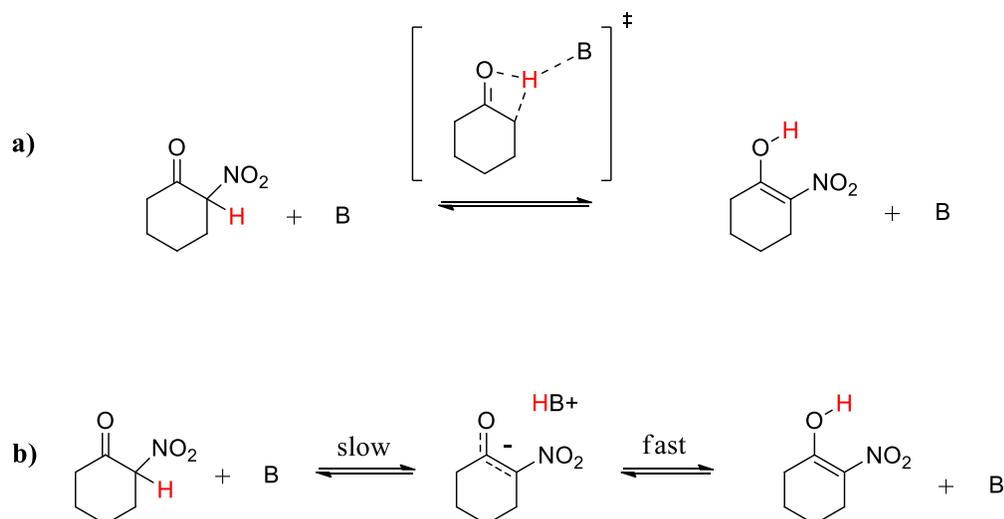
where  $\Delta G^\ddagger(T)$  is the free energy of activation at absolute temperature  $T$ ,  $R$  is the universal gas constant,  $k^B$  is the rate of reaction,  $k_b$  is the Boltzmann constant, and  $h$  is the Planck constant. The experimentally determined free energies of activation for the enolization of 2NCH catalyzed by pyridine (Pyr), 3-methylpyridine (3MePyr), 3-methoxypyridine (3MeOPyr), 4-methylpyridine (4MePyr), 4-methoxypyridine (4MeOPyr), and triethylamine (Et<sub>3</sub>N), in cyclohexane, carbontetrachloride, chloroform, dichloromethane, and acetonitrile are tabulated in table 5.15.

**Table 5.15** Experimental Free Energies of Activation  $\Delta G^\ddagger$  (kcal·mol<sup>-1</sup> at 298.15 K) for the Base-Catalyzed Enolization of 2-NCH in Various Solvents<sup>26,44</sup> All values given in kcal·mol<sup>-1</sup>.

$\Delta G^\ddagger$ of Enolization for 2-nitrocyclohexanone					
Base	C <sub>6</sub> H <sub>12</sub>	CCl <sub>4</sub>	CHCl <sub>3</sub>	CH <sub>2</sub> Cl <sub>2</sub>	CH <sub>3</sub> CN
Et <sub>3</sub> N	18.5	17.0	15.6	15.3	15.1
Pyr	22.4	21.9	20.3	20.3	20.2
4MeOPyr	--	--	18.5	18.5	18.6
4MePyr	--	--	19.1	19.1	19.2
3MeOPyr	--	--	20.0	20.3	20.4
3MePyr	--	--	19.5	19.5	--

Siani et al. proposed a shift in mechanism from concerted to stepwise upon going from less polar to more polar solvents, illustrated in scheme 5.1.<sup>44</sup> This mechanistic shift is also examined via the methods detailed below.

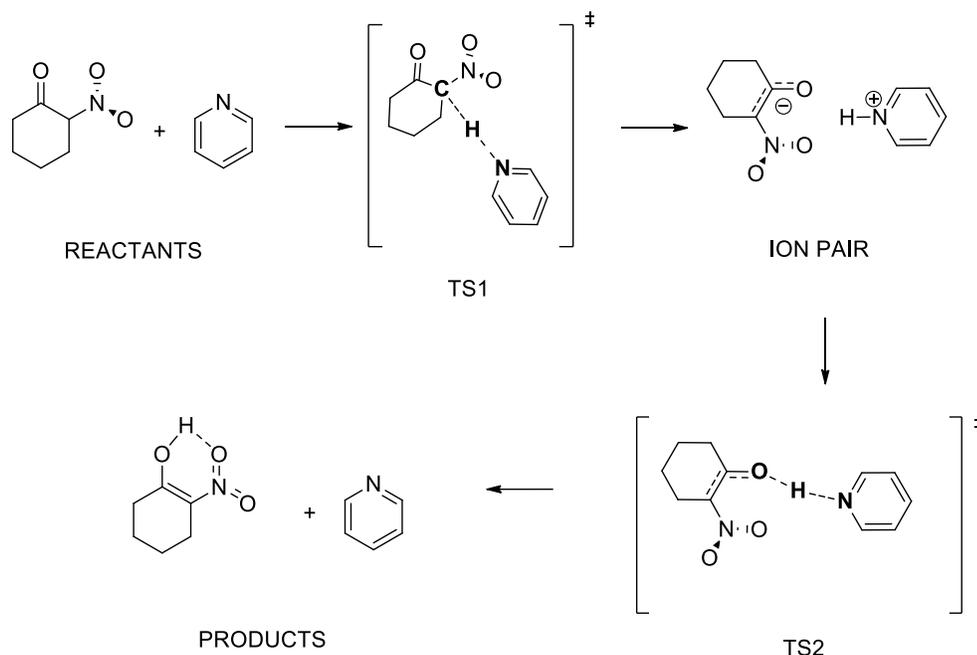
**Scheme 5.1.** Proposed shift in the enolization mechanism of 2NCH from (a) concerted in less polar solvents to (b) stepwise in more polar solvents.



### 5.3.1 Methods

Two computational methods have been used in determining the  $\Delta G^\ddagger$  and mechanism of enolization of 2NCH. First, M06/6-31+G(d,p) was used in combination with continuum solvation models and cavities PAULING/CPCM and COULOMB/SMD. While the latter is the most accurate for dicarbonyls in all solvents, the former is most accurate for 2NCH specifically, with a MUE, MAX, and STDDEV of 0.26, 0.34, and 0.09 kcal·mol<sup>-1</sup>, respectively. Assessing both solvation models and cavities will determine the accuracy of calculated free energies of activation when using the best model and cavity for a class of tautomer verses using the best model and cavity for a specific tautomer. Stationary points on the potential energy surface associated with the stepwise mechanism (scheme 5.1.b), illustrated in scheme 5.2, were calculated. Vibrational calculations were performed to verify geometries as either minima or transition structures and to provide thermal corrections to the electronic energy.

**Scheme 5.2** Stationary points characterizing the stepwise enolization of 2NCH.



Initial conformation analysis was performed by first optimizing at the AM1 level and then using a combination of stochastic and internal coordinate sampling. From the initial AM1 geometry, 200 trial geometries were created via random changes to the bond lengths, angles, or dihedrals in the case of internal sampling or Cartesian coordinates in the case of stochastic sampling. The energies of all initial trial geometries were then calculated using the OPLS-AA force field and optimized using the quasi-Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm as implemented in BOSS 4.7.<sup>46</sup> To eliminate identical local minima, three criteria were applied: an maximum energy difference tolerance of 0.070 kcal/mol; a maximum root mean square difference in atom positions between two overlaid conformers of 0.200 Å; and a maximum difference in the

sum of all of the internuclear distances squared of 4.000 Å. This “fine” differentiation of local minima is able to distinguish between very similar conformations, e.g. gauche and trans ethanol.

Because of the rigidity of the pyridine-derived bases, alternative conformations were not explored. The enol form of 2-nitrocyclohexanone (2NCH) optimized to only two conformers, with the *cis* conformer having the ability to intramolecularly hydrogen-bond. This intramolecular stabilization lead to a drastic difference in energy and hence the *trans* conformer was not included in later, higher level conformational analyses.

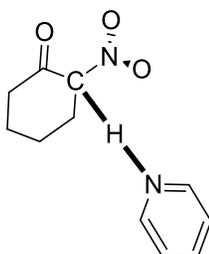
The many local minima which were found for 2NCH and triethylamine were then optimized using the M06 density functional with a 6-31+G(d,p) basis set and the SMD and IEFPCM solvation models with the COULOMB and PAULING cavities, respectively, in cyclohexane, carbon tetrachloride, acetonitrile, dichloromethane, and water. The most stable conformers were then used as starting geometries for subsequent calculations. In all solvents, 2NCH favored the conformer in which the nitro group is in the equatorial position. Many conformers were produced for triethylamine, some of which appear to be sterically unfavorable when forming the activated complex, though they are the most stable in isolation.

Second, a mixed quantum mechanical/molecular mechanical (QM/MM) analysis was performed. The solute and solvent energies were calculated using an AM1 Hamiltonian and the OPLS-AA force field, respectively. Solute-solvent interaction energies were calculated using equation 2.6.2. The solute point charges used in equation 2.6.2 were determined using the CM1 charge model and scaled by a factor of 1.14 to reflect the increased dipole moment in solvent (see section 2.6). Free energy perturbation

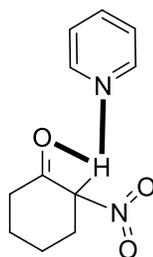
(FEP) was used in combination with Metropolis Monte Carlo (MC) statistical mechanics to equilibrate configuration space and to determine the change in free energy at each point on the potential energy surface. Five million Monte Carlo configurations were used for the initial equilibration of the system followed by 10 million configurations of averaging.

For the stepwise mechanism (scheme 5.1a), the two dimensional free energy map was created by perturbing both the carbon-hydrogen  $R_{CH}$  and nitrogen-hydrogen  $R_{NH}$  bond distances (scheme 5.3) from approximately 1.00 to 5.00 Å. For the concerted mechanism (scheme 5.1b), the oxygen-hydrogen  $R_{OH}$  and nitrogen-hydrogen and  $R_{NH}$  distances were perturbed from 0.94 to 3.06 and from 1.00 to 2.00 Å, respectively. (Scheme 5.4) Bond distances were initially partitioned into windows of 0.05 Å to characterize the potential energy surface, while a window size of 0.01 Å was used to locate and determine the relative free energies of critical points. Mechanisms were simulated in acetonitrile, dichloromethane, carbon tetrachloride, and cyclohexane. For acetonitrile, dichloromethane, and carbon tetrachloride, the initial deprotonation of the stepwise mechanism was explored. For cyclohexane, the solvent with the lowest dielectric constant, the concerted mechanism was performed.

**Scheme 5.3** Reaction coordinates (bold) used in creating the 2D free energy map for the stepwise enolization of 2NCH.



**Scheme 5.4** Reaction coordinates (bold) used in creating the 2D free energy map for the concerted enolization of 2NCH.



## 5.3.2 Results and Discussion

### 5.3.2.1 M06/6-31+G(d,p) with Continuum Solvation Model

The calculated and experimental free energies of activation  $\Delta G^\ddagger$  for rate-determining, first transition state (TS1) of the stepwise mechanism calculated using M06/6-31+G(d,p) with the PAULING/CPCM and COULOMB/SMD cavities and models are tabulated in tables 5.16 and 5.17, respectively. All energies are given in kcal·mol<sup>-1</sup>.

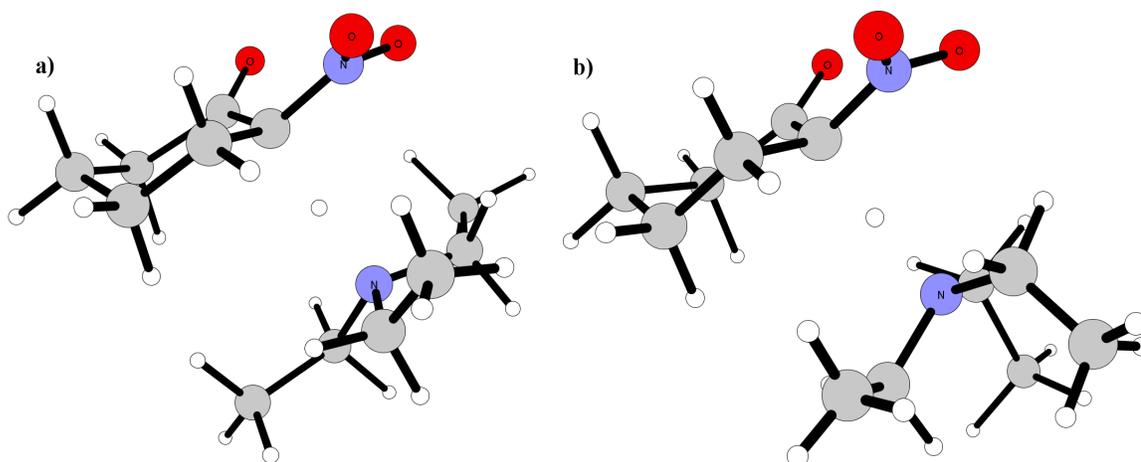
**Table 5.16** Experimental and calculated  $\Delta G^\ddagger$  for the stepwise TS1 of enolization of 2NCH. Values calculated using the PAULING/CPCM/M06/6-31+G(d,p) method and given in kcal·mol<sup>-1</sup>.

Experimental and PAULING/CPCM/M06/6-31+G(d,p) Calculated $\Delta G^\ddagger$										
Base	C <sub>6</sub> H <sub>12</sub>		CCl <sub>4</sub>		CHCl <sub>3</sub>		CH <sub>2</sub> Cl <sub>2</sub>		CH <sub>3</sub> CN	
	Calc	Exp	Calc	Exp	Calc	Exp	Calc	Exp	Calc	Exp
Et <sub>3</sub> N	16.4	18.5	16.2	17.0	13.5	15.6	13.7	15.3	13.1	15.1
Pyr	22.1	22.4	21.6	21.9	19.1	20.3	18.2	20.3	16.9	20.2
4MeOPyr	--	--	--	--	18.0	18.5	17.1	18.5	16.5	18.6
4MePyr	--	--	--	--	19.3	19.1	18.5	19.1	--	19.2
3MeOPyr	--	--	--	--	19.5	20.0	18.4	20.3	17.7	20.4
3MePyr	--	--	--	--	--	19.5	17.3	19.5	--	--

**Table 5.17** Experimental and calculated  $\Delta G^\ddagger$  for the stepwise TS1 of enolization of 2NCH. Values calculated using the COULOMB/SMD/M06/6-31+G(d,p) method and given in kcal·mol<sup>-1</sup>.

Experimental and COULOMB/SMD/M06/6-31+G(d,p) Calculated $\Delta G^\ddagger$										
Base	C <sub>6</sub> H <sub>12</sub>		CCl <sub>4</sub>		CHCl <sub>3</sub>		CH <sub>2</sub> Cl <sub>2</sub>		CH <sub>3</sub> CN	
	Calc	Exp	Calc	Exp	Calc	Exp	Calc	Exp	Calc	Exp
Et <sub>3</sub> N	21.0	18.5	21.2	17.0	18.9	15.6	20.5	15.3	17.2	15.1
Pyr	23.2	22.4	22.9	21.9	22.0	20.3	21.6	20.3	21.0	20.2
4MeOPyr	--	--	--	--	20.4	18.5	20.6	18.5	19.9	18.6
4MePyr	--	--	--	--	--	19.1	19.5	19.1	19.0	19.2
3MeOPyr	--	--	--	--	--	20.0	22.2	20.3	20.3	20.4
3MePyr	--	--	--	--	21.4	19.5	21.6	19.5	--	--

As may be expected, the solvation model and cavity that is ideal for 2NCH, PAULING/CPCM, yielded the most accurate results, having a MUE, MAX, and STDDEV of 1.5, 3.3, and 0.9 kcal·mol<sup>-1</sup>. The COULOMB/SMD model, which is most accurate for the dicarbonyl class of tautomers all considered together, yielded a MUE, MAX, and STDDEV of 1.8, 5.2, and 1.3 kcal·mol<sup>-1</sup>. The highest errors for COULOMB/SMD are associated with Et<sub>3</sub>N. For example, in dichloromethane, COULOMB/SMD has an error of 5.2 kcal·mol<sup>-1</sup> while the error for PAULING/CPCM is 1.6 kcal·mol<sup>-1</sup>. Most likely, this is caused by steric hindrance, due to the orientation of the methyl groups, within the activated complex, as shown in figure 5.1a. Removing the triethylamine results lowers the MUE, MAX, and STDDEV of COULOMB/SMD to 1.2, 2.1, and 0.7 kcal·mol<sup>-1</sup>. Removal of triethylamine results for PAULING/CPCM lowers the MUE and STDDEV by only 0.01 kcal·mol<sup>-1</sup>. This issue with triethylamine may reflect the accuracy in the conformational ranking when using the SMD solvation model.



**Figure 5.1** Geometries for the first transition structures in the enolization of 2NCH in dichloromethane as calculated by (a) COULOMB/SMD and (b) PAULING/CPCM. The orientation of the methyl groups of (a) causes significant steric hindrance.

Regarding the mechanism of enolization, the stepwise first transition structure, having a carbonyl carbon,  $\alpha$ -carbon, hydrogen bond angle of approximately  $100^\circ$ , was found for all bases in solvents of both low and high dielectric constants. This bond angle is consistent with a stepwise deprotonation, while an acute angle would be expected for a concerted mechanism. All transition structures are essentially identical regardless of solvent dielectric constant, with only minor geometric variations. No evidence was found of the hypothesized concerted mechanism. In the attempts to locate the concerted transition structure, the methodology was not restricted to M06, but also CBS-QB3, MP2, B3LYP, and M06-2X methods with augmented, triple zeta basis sets.

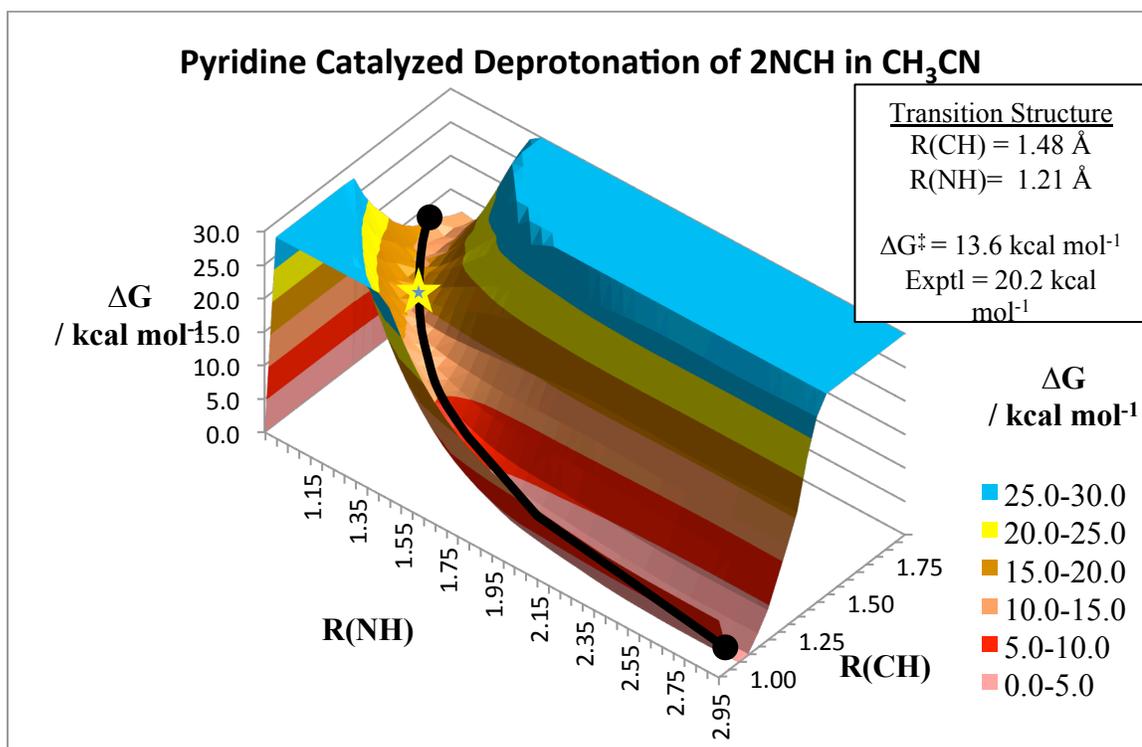
Additionally, the fully separated ion pair is a stationary point on the potential energy surface for all solvents. Should the hypothesized concerted mechanism occur for

low dielectric solvents, the carbonyl oxygen of 2NCH would spontaneously abstract the proton from the protonated base. These findings are strong evidence that the stepwise mechanism occurs in all solvents.

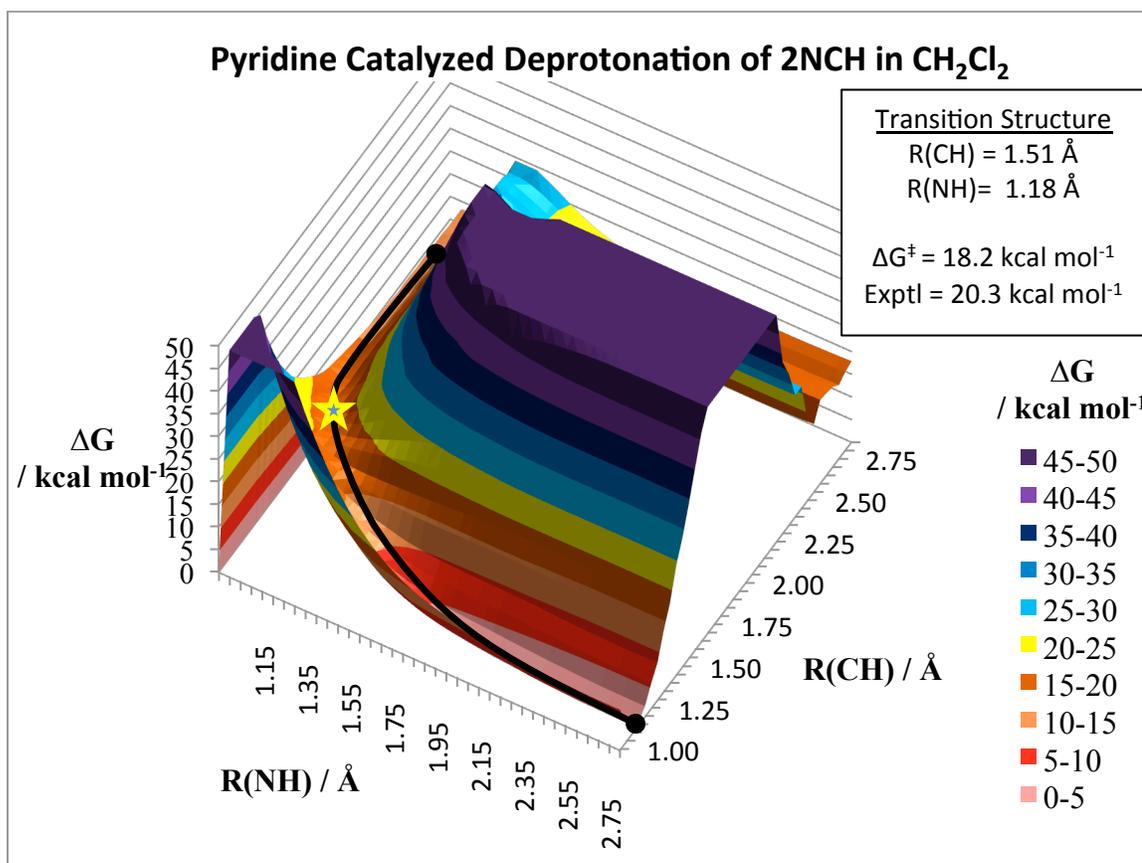
Since continuum solvation models represent the solvent as a grid of polarizable charges and lack specific site-to-site interactions between solute and solvent, explicit solvation was then used to investigate the reaction mechanism.

### 5.3.2.2 QM/MM/FEP/MC Results

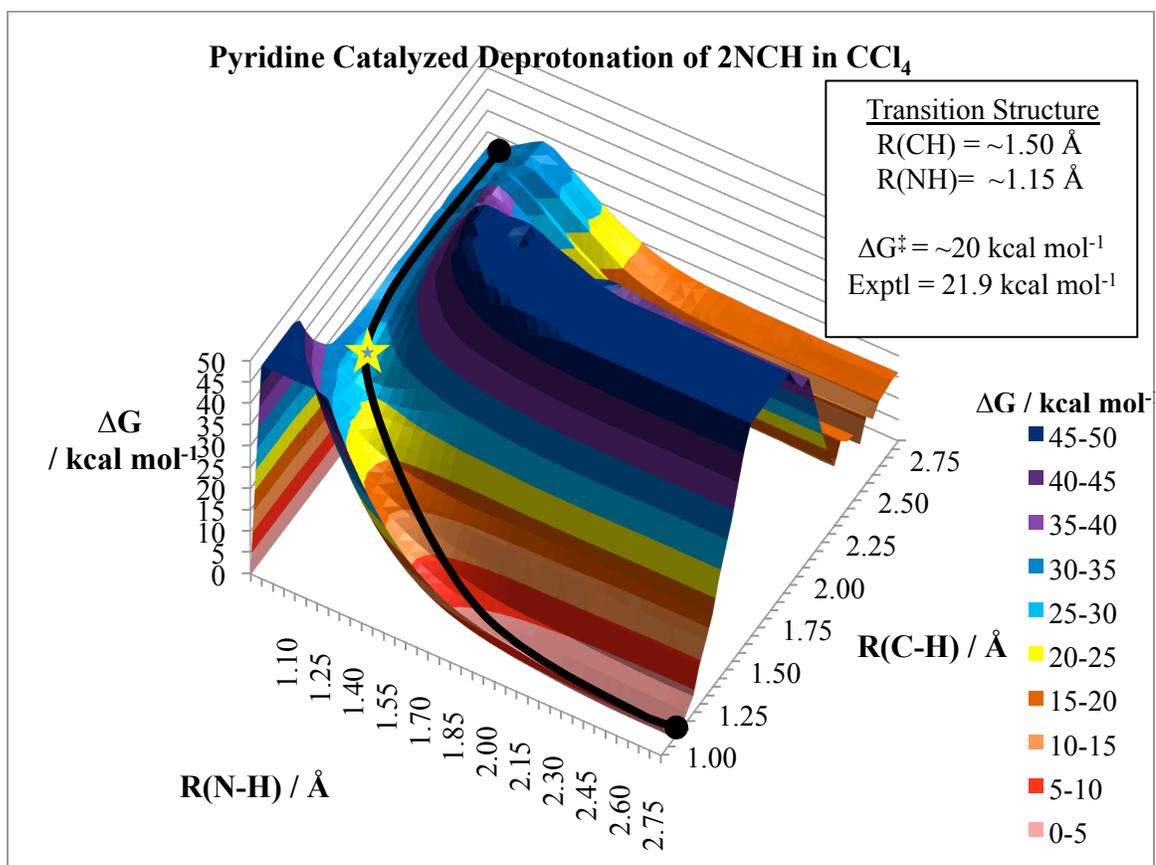
Figures 5.2 through 5.4 illustrate the potential energy surfaces for the base-catalyzed deprotonation of 2NCH in solvents of decreasing dielectric constant, i.e. acetonitrile ( $\epsilon = 37.5$ ), dichloromethane ( $\epsilon = 8.9$ ), and carbon tetrachloride ( $\epsilon = 2.2$ ). Figure 5.5 shows the potential energy surface for hypothesized concerted enolization in the solvent of the lowest dielectric constant, cyclohexane ( $\epsilon = 2.01$ ). The reactions in acetonitrile and dichloromethane have well defined saddle points (gold stars in figures) at bond distances of approximately  $R_{\text{CH}} = 1.5$  and  $R_{\text{NH}} = 1.2 \text{ \AA}$ . In carbon tetrachloride, however, the potential energy surface remains flat as the ion pair separate, maintaining a free energy approximately equienergetic relative to the transition structure. This indicates a more intimately bound ion pair for the case of carbon tetrachloride. This finding provides qualitative support for the concerted mechanism in this solvent. However, when the concerted mechanism is explicitly calculated in cyclohexane, the free energy of activation is found to be  $64.0 \text{ kcal}\cdot\text{mol}^{-1}$ , which is far too high to occur at room temperature.



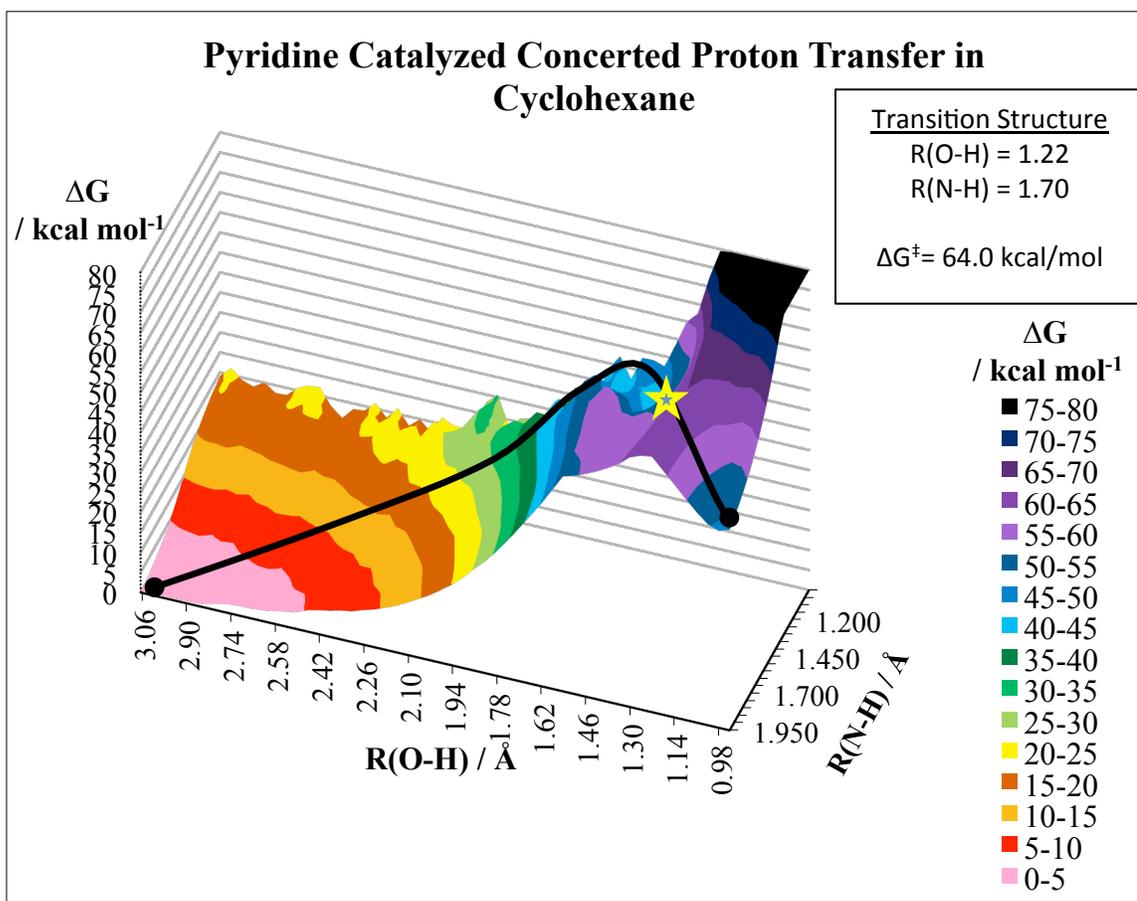
**Figure 5.2** Potential energy surface for TS1 of the stepwise pyridine catalyzed deprotonation of 2NCH in acetonitrile.



**Figure 5.3** Potential energy surface for TS1 of the stepwise pyridine catalyzed deprotonation of 2NCH in dichloromethane.



**Figure 5.4** Potential energy surface for TS1 of the stepwise pyridine catalyzed deprotonation of 2NCH in carbontetrachloride.



**Figure 5.5** Potential energy surface for the concerted enolization of 2NCH catalyzed by pyridine in cyclohexane.

## 5.4 Conclusions

Benchmarking seven gas phase free energies of tautomerization reveals that the composite G4 model and M06 density functional with a 6-31+G(d,p) basis set yield the lowest mean unsigned error, 0.93 and 0.73 kcal·mol<sup>-1</sup>, respectively. These two levels of theory were chosen to represent the solute monocarbonyls, dicarbonyls, and heterocycles for the 50 solution phase free energy of tautomerization calculations. The solvent was characterized using solvation models IEFPCM, CPCM, and SMD, along with cavities PAULING, BONDING, UA0, UAHF, UAKS, and COULOMB. This study finds that no one single solvation model and cavity is best for all molecules. Partitioning the results into classes of tautomers and then again into aqueous and nonaqueous solvents greatly improved results. IEFPCM/UAKS solvation model and cavity calculates aqueous monocarbonyls with a mean unsigned error of  $0.83 \pm 0.40$  kcal·mol<sup>-1</sup>.  $\beta$ -dicarbonyls in aqueous and nonaqueous solvents are calculated with the SMD/COULOMB and CPCM/BONDI combinations with mean unsigned errors of  $0.96 \pm 0.63$  kcal·mol<sup>-1</sup> and  $0.90 \pm 0.77$  kcal·mol<sup>-1</sup>, respectively. Heterocycles in aqueous and nonaqueous solvents are calculated with the CPCM/PAULING and IEFPCM/UFF combinations with a mean unsigned error of  $0.75 \pm 0.57$  kcal·mol<sup>-1</sup> and  $0.97 \pm 0.64$  kcal·mol<sup>-1</sup>, respectively.

The mechanism and free energies of activation of enolization of the  $\beta$ -dicarbonyl 2-nitrocyclohexanone with various bases has been explored. The SMD/COULOMB and CPCM/PAULING models and cavities, the former most being most accurate for  $\beta$ -dicarbonyls in general and the latter being most accurate for 2-nitrocyclohexanone specifically, were used. CPCM/PAULING and SMD/COULOMB reproduce the free

energies of activation with a mean unsigned error of  $1.5 \pm 0.9$  kcal·mol<sup>-1</sup> and  $1.8 \pm 1.3$  kcal·mol<sup>-1</sup>, respectively. The above methods along with QM/MM/FEP/MC calculations all provide strong evidence that the mechanism is stepwise in solvents of all dielectric constants.

## 5.5 References

- (1) Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *The Journal of Chemical Physics* **1999**, *110*, 2822.
- (2) Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *The Journal of Chemical Physics* **2000**, *112*, 6532.
- (3) Ochterski, J. W.; Petersson, G. A.; Montgomery, J. A. *The Journal of Chemical Physics* **1996**, *104*, 2598.
- (4) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *The Journal of Chemical Physics* **2007**, *126*, 084108.
- (5) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 810–825.
- (6) Dewar, M.; Zoebisch, E.; Healy, E. *Journal of the American Society* **1985**, 3902–3909.
- (7) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *Journal of Computational Chemistry* **2002**, *23*, 1601–22.
- (8) Tubert-Brohman, I.; Guimarães, C. R. W.; Jorgensen, W. L. *Journal of Chemical Theory and Computation* **2005**, *1*, 817–823.
- (9) Stewart, J. J. P. *Journal of Molecular Modeling* **2007**, *13*, 1173–213.
- (10) Becke, A. D. *The Journal of Chemical Physics* **1993**, *98*, 1372.
- (11) Xu, X.; Goddard, W. a *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 2673–7.
- (12) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2006**, *2*, 364–382.

- (13) Zhao, Y.; Truhlar, D. G. *Theoretical Chemistry Accounts* **2007**, *120*, 215–241.
- (14) Chai, J.-D.; Head-Gordon, M. *Physical Chemistry Chemical Physics : PCCP* **2008**, *10*, 6615–20.
- (15) Ernzerhof, M.; Perdew, J. P. *The Journal of Chemical Physics* **1998**, *109*, 3313.
- (16) Angelini, G.; Chiappe, C.; De Maria, P.; Fontana, A.; Gasparrini, F.; Pieraccini, D.; Pierini, M.; Siani, G.; Maria, P. De *The Journal of Organic Chemistry* **2005**, *70*, 8193–8196.
- (17) Folkendt, M. M.; Weiss-Lopez, B. E.; Chauvel, J. P.; True, N. S. *The Journal of Physical Chemistry* **1985**, *89*, 3347–3352.
- (18) Belova, N. V; Oberhammer, H.; Girichev, G. V *The Journal of Physical Chemistry A* **2004**, *108*, 3593–3597.
- (19) Caminati, W.; Grabow, J.-U. *Journal of the American Chemical Society* **2006**, *128*, 854–7.
- (20) Beak, P.; Fry, F. S. *Journal of the American Chemical Society* **1973**, *95*, 1700–1702.
- (21) Beak, P.; Covington, J. B.; White, J. M. *The Journal of Organic Chemistry* **1980**, *45*, 1347–1353.
- (22) Toullec, J. In *The chemistry of enols*; Rappoport, Z., Ed.; John Wiley and Sons Ltd.: West Sussex, England, 1990; pp. 323–398.
- (23) Mills, S. G.; Beak, P. *The Journal of Organic Chemistry* **1985**, *50*, 1216–1224.
- (24) Iglesias, E. *The Journal of Organic Chemistry* **2003**, *68*, 2680–8.
- (25) Angelini, G.; Maria, P. De; Fontana, A.; Pierini, M.; Siani, G. *The Journal of Organic Chemistry* **2007**, *72*, 4039–47.
- (26) Angelini, G.; De Maria, P.; Chiappe, C.; Fontana, A.; Gasbarri, C.; Siani, G. *The Journal of Organic Chemistry* **2009**, *74*, 6572–6.
- (27) Bunting, J. W.; Kanter, J. P. *Journal of the American Chemical Society* **1993**, *115*, 11705–11715.
- (28) Beak, P. *Accounts of Chemical Research* **1977**, *10*, 186–192.

- (29) Kwiatkowski, J. S.; Zielinski, T. J.; Rein, R. In *Advances in Quantum Chemistry*; Lowdin, P.-O., Ed.; Academic Press: Orlando, 1986; pp. 85–130.
- (30) Schlegel, H. B. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 790–809.
- (31) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. S.; M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. M.; G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. H.; A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. H.; M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. N.; Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. M. J.; J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. B.; K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. N.; K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. T.; M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. C.; V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. S.; O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. O.; R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. V.; P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. D.; O. Farkas, J. B. Foresman, J. V. Ortiz, J. C.; Fox, D. J. F. *Gaussian 09, Revision B.01*; Gaussian, Inc.: Wallingford, CT, 2010.
- (32) Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. *The Journal of Chemical Physics* **2002**, *117*, 43.
- (33) Barone, V.; Cossi, M. *The Journal of Physical Chemistry A* **1998**, *102*, 1995–2001.
- (34) Marenich, A. V; Cramer, C. J.; Truhlar, D. G. *The Journal of Physical Chemistry. B* **2009**, *113*, 6378–96.
- (35) Mennucci, B. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 386–404.
- (36) Cramer, C. J.; Truhlar, D. G. *Chemical Reviews* **1999**, *99*, 2161–2200.
- (37) Tomasi, J.; Persico, M. *Chemical Reviews* **1994**, *94*, 2027–2094.
- (38) Tomasi, J.; Mennucci, B.; Cammi, R. *Chemical Reviews* **2005**, *105*, 2999–3093.
- (39) Bondi, A. *Journal of Physical Chemistry* **1964**, *68*, 441–451.
- (40) Pauling, L. *The Nature of the Chemical Bond*; 3rd ed.; Cornell University Press: Ithaca, NY, 1960.
- (41) Barone, V.; Cossi, M.; Tomasi, J. *The Journal of Chemical Physics* **1997**, *107*, 3210.

- (42) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *Journal of the American Chemical Society* **1992**, *114*, 10024–10035.
- (43) Angelini, G.; Chiappe, C.; De Maria, P.; Fontana, A.; Gasparrini, F.; Pieraccini, D.; Pierini, M.; Siani, G. *The Journal of Organic Chemistry* **2005**, *70*, 8193–6.
- (44) Siani, G.; Angelini, G.; De Maria, P.; Fontana, A.; Pierini, M. *Organic & Biomolecular Chemistry* **2008**, *6*, 4236–41.
- (45) Eyring, H. *The Journal of Chemical Physics* **1935**, *3*, 107.
- (46) Jorgensen, W. L.; Tirado-Rives, J. *Journal of Computational Chemistry* **2005**, *26*, 1689–700.