Charge-dependent non-bonded interaction methods for use in quantum mechanical modeling of condensed phase reactions

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Erich R. Kuechler

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Darrin M. York

October, 2015

Acknowledgements

I would like to express my sincere gratitude to my adviser Professor Darrin M. York; for providing consistent research funding while being patient with my curiosity, encouraging with my scientific endeavors and supportive of my professional aspirations. I would also like to thank all current and past members of the York research group for their guidance and camaraderie over the years. Special mention must be given to Timothy J. Giese, as his keen insight and constant support in my research have proven themselves to be indispensable.

Abstract

Molecular modeling and computer simulation techniques can provide detailed insight into biochemical phenomena. This dissertation describes the development, implementation and parameterization of two methods for the accurate modeling of chemical reactions in aqueous environments, with a concerted scientific effort towards the inclusion of charge-dependent non-bonded non-electrostatic interactions into currently used computational frameworks. The first of these models, QXD, modifies interactions in a hybrid quantum mechanical/molecular (QM/MM) mechanical framework to overcome the current limitations of 'atom typing' QM atoms; an inaccurate and non-intuitive practice for chemically active species as these static atom types are dictated by the local bonding and electrostatic environment of the atoms they represent, which will change over the course of the simulation. The efficacy QXD model is demonstrated using a specific reaction parameterization (SRP) of the Austin Model 1 (AM1) Hamiltonian by simultaneously capturing the reaction barrier for chloride ion attack on methylchloride in solution and the solvation free energies of a series of compounds including the reagents of the reaction. The second, VRSCOSMO, is an implicit solvation model for use with the DFTB3/3OB Hamiltonian for biochemical reactions; allowing for accurate modeling of ionic compound solvation properties while overcoming the discontinuous nature of conventional PCM models when chemical reaction coordinates. The VRSCOSMO model is shown to accurately model the solvation properties of over 200 chemical compounds while also providing smooth, continuous reaction surfaces for a series of biologically motivated phosphoryl transesterification reactions. Both of these methods incorporate charge-dependent behavior into the non-bonded interactions variationally, allowing the 'size' of atoms to change in meaningful ways with respect to changes in local charge state, as to provide an accurate, predictive and transferable models for the interactions between the quantum mechanical system and their solvated surroundings.

Contents

List of Tables

List of Figures

Forward

The content of this dissertation is concerned with the creation of computational tools for the simulation of chemical reactions occurring in condensed phase environments. More specifically, it describes the development, implementation and application of charge-dependent interaction models for use in quantum chemical calculations and simulations. The driving motivation of the research contained herein is to further the field computational sciences by providing a framework for more accurate, transferable and predictive models for use in molecular simulation. While these models are universally applicable, the Darrin York research group (York Group) has a keen focus on the elucidation and characterization of the mechanisms and free energy landscapes involved in catalytic Riboneucleic Acid (RNA) enzymes; commonly known as ribozymes. As such, much of the research presented in this manuscript will be framed from the perspective of biomolecular simulation and in the study of the mechanistic pathways for ribozyme catalysis.

The York Group takes a multiscale approach to addressing the questions surrounding ribozymes. Meaning that various levels of molecular modeling, from high-level quantum mechanical methods to fast approximate semiempirical models, will be levied in an effort to gain an atomistic understanding of the reaction pathways in RNA catalysis. Additionally, the York Group endeavors to develop of new theoretical models, tested against experimental data, to increase the accuracy of computational data obtained. Shown in Figure 1 is a general flowchart of the research paradigm used by the York Group, with indications of how the results and models presented in this manuscript fit into the overall scheme of deciphering the mechanistic details of RNA catalysis.

The research presented in this dissertation will proceed as follows: Chapter 1 will

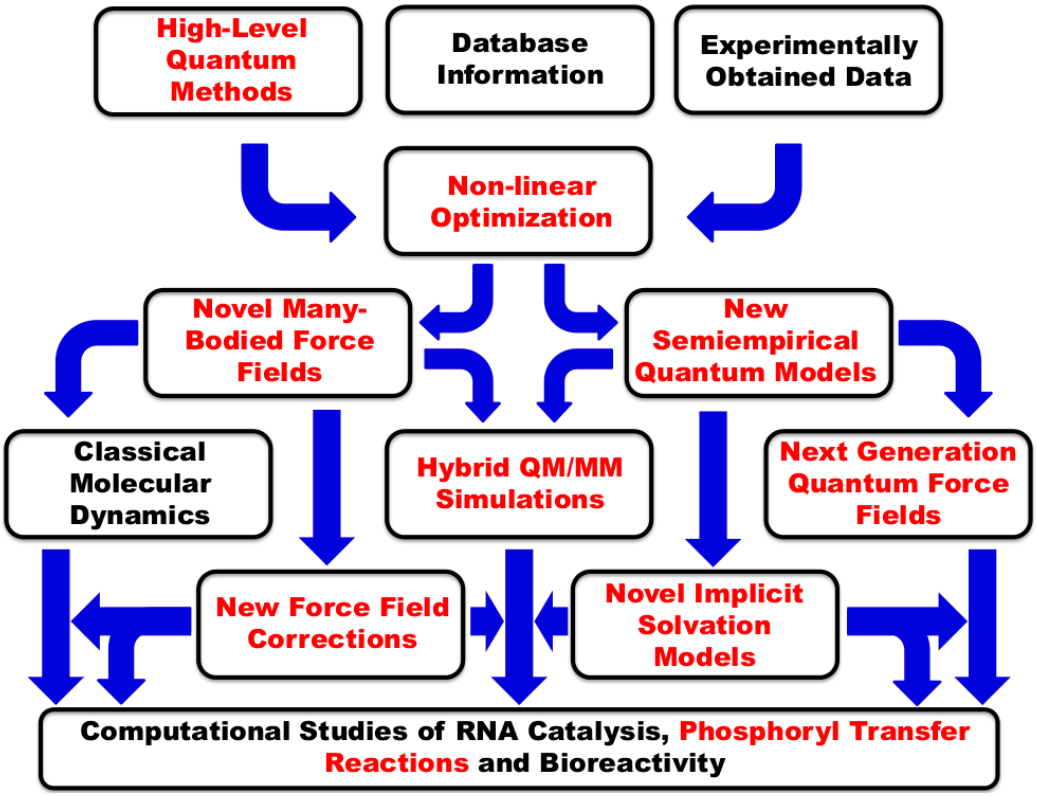


Figure 1: **General schematic of the multiscale research paradigm utilized by the York Group.** Starting with high-level computational data, data from well-established databases and/or directly from experimental data computational models are developed through non-linear optimization and then utilized in molecular calculation and simulation to probe catalytic RNA pathways and model systems. Information gained from all levels of detail, from highly accurate quantum calculation to approximate classical models, are then compiled in the attempt to construct a complete chemical picture of RNA catalysis. Research completed by the author is highlighted in the scheme and the work presented in Chapters 2, 3 and 4 relates to several of these individual areas, while Chapter 1 will provide general information as to better understand the basic terminology and concepts needed to better appreciate the research presented therein for a more general audience.

provide background information, outlining several different concepts and detailing nomenclature for which an basic understanding is required to appreciate the research contained herein. Subsections contained within Chapter 1 will begin by briefly discussing the biochemical arena which provides the motivation for the models that are discussed later in the work. While no large-scale systems have been explicitly examined in this work, several systems modeling key aspects of the mechanisms seen in the catalyzed RNA self-cleavage reaction, a phosphoryl transesterification reaction, have been studied and used in the testing and validation of the newly developed computational models. Then the chapter briefly discusses the field of computational chemistry and provides a brief review of pertinent literature to the research discussed in this dissertation. The subsections contained therein delve more deeply into computational and statistical methods which are utilized in this work, providing a general background to some concepts which may not be necessarily well known, but whose concepts are important to the greater understanding of the research discussed.

The manuscript will then continue into Chapters 2, 3 and 4 where newly developed computational models are presented and discussed, with the goal of addressing some of the underlying assumptions made for computational convenience and circumventing them with physically meaningful, more accurate computational models. Chapter 4 will also detail the mechanistic pathways of non-enzymatic phosphoryl transfer reactions. These studies serve to examine model systems for catalytic RNA, providing insights into the mechanistic details of the transesterification reaction without the complication of the enzymatic environment. These calculations provide a groundwork which can then be used as a launchpad for further studies into ribozymes to better understand the catalytic effect of the RNA framework on chemical reactions taking place.

As a whole, this work represents a broad approach to including next-generation interaction models into molecular modeling, with a focus on biocatalysis. The models from published works described have been well received in the peer-reviewed chemical literature and are transferable into a wide range of avenues in computational chemistry; representing a modest but firm step towards a deeper understanding of reactive chemical systems in general.

Chapter 1

Background

1.1 RNA Catalysis and Model Phosphoryl Transfer Reactions

Discovered in the mid-1980s[1, 2, 3] riboneucleic acid (RNA) enzymes, also known as "ribozymes," had a profound impact on molecular biology and biochemisty; eventually leading to Tomas Cech being awarded the Nobel Prize in 1989 for his work in uncovering the catalytic properties RNA.[4, 5, 6, 7, 8, 9] It was traditionally thought that RNA served mainly as a messenger molecule; its primary function being to fulfill its role in the central dogma of biology. Proteins were thought to act as the "heavy lifters" in biochemical processes, serving as the machinery which manipulated the fundamental building blocks of life processes while RNA, on the other hand, simply relayed information between dioxyribosenuecleic acid (DNA) and proteins. However, with the surprising discovery that RNA could undergo catalyzed self-cleavage reactions, the paradigm of modern molecular biochemistry shifted leading to the quick discovery of more ribozyme and ribozyme/protein complex systems; such as the ribosome[10, 11, 12, 13, 14] and spliceosome.[15, 16, 17] From these new discoveries came the "RNA World hypothesis" which posits that the central dogma of modern biology may, at some point in the past, have consisted entirely of RNA.[18, 19, 20, 21]

Key to the discussion of this work is the central mechanism by which RNA catalysis is thought to unfold. In catalytic RNA, one proposed mechanism is that, the phosphate

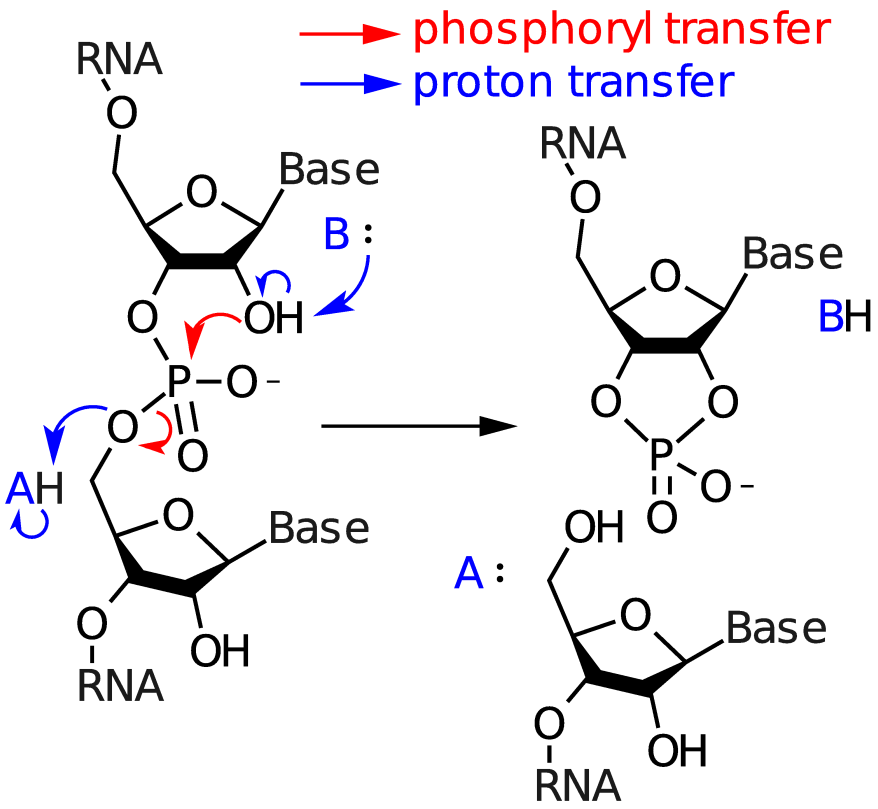


Figure 1.1: **Example of the proposed mechanism for general acid, general base catalysis seen in many RNA enzymes, known as ribozymes.** In this O2'-transphosphorylation reaction the O2' oxygen acts as a nucleophile after being activated by some general base (B) then attacks the phosphorus center of the RNA backbone. The O5' of that nucleotide then acts as a leaving group, which is then stabilized by some general acid (A).

backbone is cleaved by an 2'-O-transphosphorylation reaction. This reaction is thought to be fairly ubiquitous throughout ribozyme chemistry, and while details may vary slightly from system to system, an understanding of this class of reactions as a whole grants further insight into RNA catalysis. Figure 1.1 shows the proposed general-acid general-base mechanism for the self-cleavage reaction for the Hairpin ribozyme.[22, 23, 24] The reaction is initiated when some general base abstracts a proton from the O2' oxygen, causing a localization of electronic density on the oxygen and activating it as a nucleophile. Then, this oxygen attacks the net negatively charged phosphate center, forming a chemical bond and causing the bond to break with the opposing O5' oxygen, making it the leaving group and localizing excess electronic density upon it. Finally, the O5' oxygen is then stabilized by some general acid. Currently, most of the pressing questions involving ribozyme catalysis frequently ask of the nature of the involvement and identities of the general acid/base and whether there is the possibility of metal ion stabilization during any of the key steps in the catalytic pathway. One aspect of this reaction which poses a particular computational challenge, and that is most relevant to the research contained herein, is the highly charged nature of the active site. Carrying a local net charge of $-2e$, correctly capturing the motions of the electrons over the coarse of the reaction and accurately modeling the interaction of these highly charged systems with their surroundings can be a significant source of error for traditional theoretical models. It is easy to see that when attempting to study these, or similar, reactions computationally that having a model which can meaningfully respond to changes in local charge dynamically would greatly increase the model's accuracy and predictive power.

1.2 Computational Methods for Modeling Biochemical Reactions

Computational biochemistry is the melding of several different core areas of research; blending elements of computer science, mathematics, physics, chemistry and biology. With it, computers are used to study a system of chemical or biological interest using some physical or mathematical model in order to determine a property or physical observable of that system. With the ever increasing amount of readily available

high-performance computing platforms, chemical modeling is similarly becoming a more wide-spread and powerful tool for obtaining molecular level data. Additionally, when compared to experimental techniques, computational modeling provides a relatively low-cost and hazard-free environment to probe challenging chemical questions while frequently complimenting experimental data to help re-enforce *in vivo* findings. The interplay between experiment and theoretical modeling cannot be understated; models must be validated against reliable experimental data. Then after proper parameterization and benchmarking these models can be used to help culminate experimental findings into a complete chemical picture while eliciting experiment to explore avenues which theory predicts to be of importance.

Biochemical processes posse a particular challenge to theoretical models. The sheer magnitude of the systems, the biomolecule and the surrounding physiological environment, involved in these phenomena require special treatment in order to adequately examine. The size of these systems push traditional computational methods to their limit, requiring long-time simulation to allow both the biomolecule and the solvent to appropriately explore conformation states. Additionally, when studying these systems and especially those involving nucleic acids and RNA, special care must be taken to accurately model the long-range electrostatic behavior as well as the underlying electronic structure of residues involved in the reactions carried out by these biomolecules as they are particularly sensitive to their surrounding electrostatic environment.

The remainder of this chapter provides the reader with a general overview into some of the concepts, theoretical methods and computational models used in computational chemistry and later in this work. A fundamental understanding of the techniques reviewed here will be needed to appreciate the research contained later within this manuscript. The purpose of this chapter is not to re-derive any of these methods, but rather to orient the reader toward a better understanding of the research discussed later.

The remaining sections of this chapter will unfold as follows: In subsection 1.2.1 a brief description of classical molecular mechanical techniques and force fields is given. In subsection 1.2.2 some techniques and models used for electronic structure calculations are discussed. Subsection 1.2.3 continues to detail hybrid quantum mechanical/molecular mechanical simulations; a key technique used throughout the research in

this manuscript. Subsection 1.2.4 covers implicit solvation models, an attractive alternative for reducing computational cost for modeling chemical reactions occuring in the condensed phase as compared to QM/MM simulations. Finally, section 1.3 introduces two enhanced sampling techniques used in this work to calculate physical observables from the chemical simulations completed in this manuscript.

1.2.1 Molecular Mechanical Models

The central conceit of molecular dynamics (MD) is the idea that the motions of the modeled atoms, as dictated by the underlying potential, if given enough time will explore all possible configurations available to the system as according to their Boltzmann probability. This *ergodic hypothesis*, stated more concisely as

$$\langle A \rangle_{ensemble} = \langle A \rangle_{time} \tag{1.1}$$

where A is some chosen observable, $\langle \dots \rangle_X$ is the average over the condition X. In this example the left hand expression is commonly referred to as "ensemble average of A". Along with the *Gibb's postulate*, which posits that the macroscopic state of a system is determined by a distribution of microscopic states, connects the statistical sampling of microscopic configurations and momenta gained though molecular simulation and relate them to macroscopic measurements. From these simple statements an important realization about dynamic molecular models can be gained. For accurate data, one must allow their simulations an adequately long amount of time for the system to have a chance to visit all areas of *phase space*, unique combinations of configurations and momenta, in order to generate a representative sample upon which meaningful results can be obtained.

One method for sampling phase space is to utilize classical mechanics and to propagate molecules using a classical *force field*. A force field is a series of parameters and equations which determine the interactions and forces between the atoms being modeled. Utilizing a classical force field to explore phase space is commonly known as molecular mechanics (MM) and one can generally envision the interaction between atoms in these cases as a number of differently charged balls interacting through a series of springs.[25] More exactly, one could express a MM force field as [26]

$$V_{MM} = \sum_{bonds} k_b(x - x_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} k_\phi \left[1 + \cos(n\phi - \phi_0)\right]^2 + \sum_{non-bonded} \left\{ 4\varepsilon_{\mathrm{ij}} \left[\left(\frac{\sigma_{\mathrm{ij}}}{r_{\mathrm{ij}}} \right)^{12} - \left(\frac{\sigma_{\mathrm{ij}}}{r_{\mathrm{ij}}} \right)^{6} \right] + \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{\mathrm{ij}}} \right\} \quad (1.2)$$

Where V_{MM} is the total MM potential energy. The first term in the expression is a Hooke's Law-like expression representing chemical bonds between atoms, each with a 'bond strength,' k_b, to exert a restoring force to return to a set 'bond length' of x_0. The second and third terms are similar 3- and 4-body expressions for angle and dihedral terms, respectively. The final summation in the above expression is representative of the of the interactions between all non-bonded atom pairs (between, generally, between atoms i and j). The first two terms together are known as the Lennard-Jones (LJ) potential, seen in Figure 1.2, which models the non-classical interactions of atoms, and their electrons, with an empirical potential. The LJ interaction is dictated by two pair-wise terms, σ_{ij} and $\varepsilon_{\mathrm{ij}}$, which are the contact distance and the magnitude of the interaction energy at its lowest point. The contact distance is the distance at which the LJ potential intersects the zero of energy. There is an explicit connection between the contact distance and the distance where these minimum of energy resides, known as $\mathrm{R_{min,ij}}$

$$R_{\mathrm{min,ij}} = 2^{\frac{1}{6}} \sigma_{\mathrm{ij}} \quad (1.3)$$

and as such the LJ expression can be rewritten as

$$V_{\mathrm{ij}}^{LJ} = \varepsilon_{\mathrm{ij}} \left[\left(\frac{R_{\mathrm{min,ij}}}{r_{\mathrm{ij}}} \right)^{12} - 2 \left(\frac{R_{\mathrm{min,ij}}}{r_{\mathrm{ij}}} \right)^{6} \right]. \quad (1.4)$$

This is the form which will be referenced later in the manuscript and is central to the methods presented in Chapter 3. The r^{-6} term arises from the leading term of the multipole expansion, having the form of the dipole/induced dipole interaction, while the r^{-12} term largely is used for computational convenience. The final term in Equation 1.4 is the Coulombic repulsion/attraction energy term, as frequently seen in classical mechanics, where ε_0 in this case is the electronic permutivity of the space being modeled.

Further, more rigorous and physically meaningful terms such as polarization or viscous damping,[27] can be included into the force field to improve the accuracy of desired properties, but generally at the cost of computational efficiency.

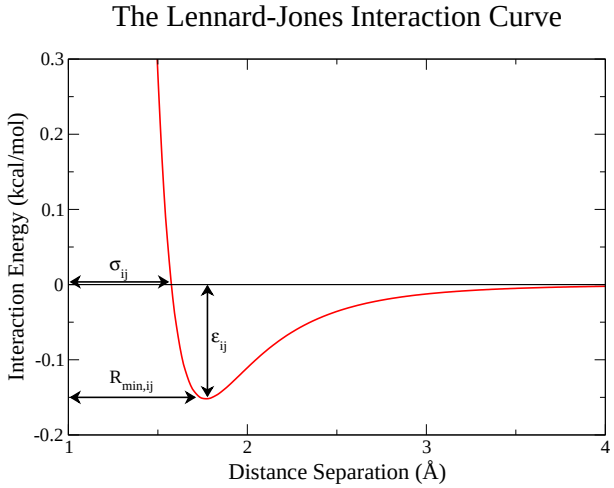


Figure 1.2: **Diagram of the Lennard-Jones interaction potential used for modeling exchange-repulsion and dispersion in molecular mechanical force fields.** This empirical interaction potential is commonly used in molecular mechanics force fields to model the non-bonded non-electrostatic interactions.

It should be noted that there are a significant number of parameters which go into making a force field; for instance in the simplistic force field expression provided previously: k_b, x_0, k_a, θ_0, k_d, ϕ, $\varepsilon_{\mathrm{ij}}$, σ_{ij}, ε_0, q_i and so on are all parameters which require optimization. These parameters can take years to refine[28, 29, 30, 31, 32, 33] and must be determined for each different *atom type* being modeled. An atom type is a representation of an atom in a specific electronic and bonding environment. Meaning that an oxygen in an ester and an oxygen in an alcohol, while being the same element, are different atom types and would therefore be assigned different force field parameters. This type of specific parameterization can yield very accurate behavior for interactions which are within the training set of the model but can lead to some issues, such as a lack flexibility when modeling systems which do not fit easily within the current parameterization. Also due to this framework, and a lack of any explicit knowledge of the underlying electronic structure of the systems being modeled, traditional MM methods cannot be used accurately for studying most chemical reactions; those which break or

form chemical bonds.

This does not mean MM techniques are not without merit. Given that these methods are generally propagated using classical Newtonian Laws or Monte Carlo sampling and the relative brevity of force field expressions, MM techniques are some of the most efficient ways to explore phase space for biomolecules. Being orders of magnitude faster than non-classical quantum mechanical (QM) techniques allows for much greater promise for representative sampling within a specific time frame; leading to a higher likelihood of better converged simulations and a more representative picture of accurate data from the model. MM techniques can readily be used to observe many different structural phenomena; such as protein folding, binding pocket stability and water networking.

1.2.2 Electronic Structure Methods

If one desires to model the creation or destruction of chemical bonds, or the motions of electrons in general, then quantum mechanical (QM) methods can be utilized. Generally much slower than MM modeling, QM methods strive to achieve highly accurate data about chemical reactivity and structure. However, the expressions used for such techniques are computationally challenging to evaluate and QM methods must be chosen carefully to balance the accuracy of the model with its associated computational cost.

While the research contained within this dissertation does not modify or expand upon any electronic structure methods, it will prove useful to have a brief review of some of the concepts used in electronic structure calculations. These types of calculations were used throughout the coming research and a semiempirical quantum model as parameterized. As such, a detailed explanation of these methods will not provided but rather an overview will be given as to layout the general landscape of the applied quantum methods and to detail how the variational corrections to the quantum mechanical self-consistent field procedure were applied.

In a majority of electronic structure calculations the goal is to gain wavefunction, or density distribution in *density functional methods*, corresponding energy eigenvalue

for the time independent Srchödinger equation

$$\hat{H}|\Psi\rangle = E|\Psi\rangle \tag{1.5}$$

where $\hat{H}$ is the Hamiltonian operator, $|\Psi\rangle$ is the many-electron wavefunction written in Dirac notation and E is the energy.[34] For this work all relativistic effects are ignored and all calculations are performed working under the *Born-Oppenheimer approximation* to separate electronic and nuclear degrees of freedom. In *wave mechanics*, the wavefunction itself is sought after by trying to solve for a set of auxiliary, molecular orbital (MO) basis set coefficients, $C_{\mu i}$, which can be used to obtain the total wavefunction which corresponds to the ground state,

$$|\Psi\rangle = |\psi_i \psi_j \dots \psi_k\rangle \,. \tag{1.6}$$

where

$$\begin{aligned} \psi_i =& C_{\mu i}\chi_\mu + C_{\nu i}\chi_\nu + \cdots + C_{ni}\chi_n \\ =& \sum_\mu C_{\mu i}\chi_\mu \end{aligned} \tag{1.7}$$

where χ_μ represent an *atomic orbital* (AO) basis for the MOs.

To approach obtaining the ground state energy, it can be stated that, generally, the energy depends on the spin-resolved electronic density matrices, e.g.

$$E = E[\mathbf{P}^\alpha, \mathbf{P}^\beta] \tag{1.8}$$

where $\mathbf{P}^\alpha$ and $\mathbf{P}^\beta$ can be defined as

$$P^\sigma_{\mu\nu} = \sum_i n_i^\sigma C^\sigma_{\mu i} C^\sigma_{\nu i} \tag{1.9}$$

where σ is either α/β (up or down spin), n_i^σ is the spin-resolved molecular orbital (MO) occupation number for the i^{th} MO (and has the potential values of either 0 or 1) and $C^\sigma_{\mu i}$ is the μ^{th} spin-resolved AO basis set coefficient for the i^{th} MO. This energy must

be minimized under the constraint that the MO basis remain orthonormal

$$\mathbf{C^T} \cdot \mathbf{S} \cdot \mathbf{C} = \mathbf{I} \tag{1.10}$$

where $\mathbf{S}$ is known as the overlap matrix,

$$S_{\mu\nu} \equiv \langle \phi_\mu | \phi_\nu \rangle \,, \tag{1.11}$$

where ϕ_μ is an arbitrary basis set function. In order to obtain the lowest, or ground state, energy one must then solve the generalized eigenvalue problem commonly known as the *Roothann-Hall equation*

$$\mathbf{F}^\sigma \cdot \mathbf{C}^\sigma = \mathbf{S}^\sigma \cdot \mathbf{C}^\sigma \cdot \mathbf{E}^\sigma \tag{1.12}$$

where $\mathbf{E}^\sigma$ is a diagonal matrix of spin-resolved orbital eigenvalues and $\mathbf{F}^\sigma$ is the Fock matrix, defined as

$$F^\sigma_{\mu\nu} = \left. \frac{\partial E}{\partial P^\sigma_{\mu\nu}} \right|_{\mathbf{R}} \tag{1.13}$$

To solve equation for the lowest energy eigenvalue, an iterative approach is taken where:[35]

1.) Define the system you are acting upon by defining the nuclear coordinates and charges, the number of electrons in the system and a basis set to operate within, ϕ_i, and the appropriate integrals, such as $S_{\mu\nu}$.

2.) Given a trial set of MO, and corresponding trail set of MO coefficients ($C^\sigma_{\mu i}$), calculate the density matrix shown in Equation 1.9. Alternatively, one could provide an initial guess to the atomic density directly.

3.) Given the previously supplied density matrix build the Fock matrix from Equation 1.13.

4.) Evaluate the energy based on the current density matrix.

5.) If the change in energy from previous iterations is small, below a set threshold, then you have reach a converged ground state energy and can stop. Otherwise, use this energy and construct a new set of MO using Equation 1.12 to obtain new guesses for $C^\sigma_{\mu i}$.

6.) With the new set of MO coefficients repeat the cycle starting from step 2.

In wave mechanics this procedure a modulation of the variational constraint condition that

$$\delta\left[\left\langle\psi|\hat{H}|\psi\right\rangle - E_0\right] = 0 \tag{1.14}$$

where, in this case, $|\psi\rangle$ is some trial wavefunction which would depend on the MO basis set coefficients $C_{\mu i}$, E_0 is the ground state energy, and $\hat{H}$ is the Hamiltonian operator with defines the interactions of the system. In DFT this constraint is slightly modified to

$$\delta\left[E[\rho] - \mu\left(\int \rho(\mathbf{r})d^3\mathbf{r} - N\right)\right] = 0 \tag{1.15}$$

where $E[\rho]$is the energy functional, μ is a Lagrange multiplier and N is the number of electrons in the system.

Hartree-Fock Energy

As an example the energy of a single determinate, Hartree-Fock-like, method as determined by acting the Hamiltonian operator upon the wavefunction can be written as

$$\begin{aligned} E = &\sum_{a>b}\frac{Z_a Z_b}{|R_a - R_b|} + \sum_{\mu\nu} P_{\mu\nu}H_{\mu\nu,\text{core}} + \frac{1}{2}\sum_{\mu\nu}\sum_{\kappa\lambda} P_{\mu\nu}P_{\kappa\lambda}(\mu\nu|\kappa\lambda) \\ &- \frac{1}{2}\sum_{\mu\kappa}\sum_{\nu\lambda} P^{\alpha}_{\mu\nu}P^{\alpha}_{\kappa\lambda}(\mu\kappa|\nu\lambda) - \frac{1}{2}\sum_{\mu\kappa}\sum_{\nu\lambda} P^{\beta}_{\mu\nu}P^{\beta}_{\kappa\lambda}(\mu\kappa|\nu\lambda) \end{aligned} \tag{1.16}$$

where Z_a is the charge of nuclei a and $|R_a - R_b|$ is the magnitude of the distance between the atom centers of a and b. $P_{\mu\nu}$ is defined as

$$P_{\mu\nu} = P^{\alpha}_{\mu\nu} + P^{\beta}_{\mu\nu} \tag{1.17}$$

and the *core Hamiltonian* is

$$H_{\mu\nu,\text{core}} = (\mu| - \frac{1}{2}\nabla^2|\nu) - (\mu|\sum_a \frac{Z_a}{|r - R_a|}|\nu) \tag{1.18}$$

where ∇^2 is the Laplacian operator and the two electron integrals

$$\begin{aligned}(\mu\nu|\kappa\lambda) &\equiv \langle\chi_\mu\chi\nu|\,\frac{1}{r_{12}}\,|\chi_\kappa\chi_\lambda\rangle \\ &= \int\int \frac{\chi_\mu(r)\chi_\nu(r)\chi_\kappa(r')\chi_\lambda(r')}{|r-r'|}d^3rd^3r'.\end{aligned} \tag{1.19}$$

In this case the Fock matrix would be

$$F_{\mu\nu} = H_{\mu\nu,\text{core}} + \sum_\kappa\sum_\lambda P_{\kappa\lambda}\left\{(\mu\nu|\kappa\lambda) - \frac{1}{2}(\mu\kappa|\nu\lambda)\right\} \tag{1.20}$$

Density Functional Energy

In the case of DFT methods the, instead of framing the energy in terms of the wave-function, the energy directly depends on the electron density, $\rho(\mathbf{r})$. Here the electronic density depends on the $3N$ coordinates of the electrons in the system. Based off of work by Hohenberg and Kohn,[36] the DFT energy can be written as

$$E[\rho] = F[\rho] + \int \rho(\mathbf{r}\nu(\mathbf{r})d^3\mathbf{r} \tag{1.21}$$

where $\nu(\mathbf{r})$ is the potential which is generated by the nuclei and any additional external potential and

$$F[\rho] = T_s[\rho] + U[\rho] + E_{XC}[\rho] \tag{1.22}$$

where $T_s[\rho]$ is the kinetic energy functional for non-interaction electrons and $U[\rho]$ is the electron-electron Coulomb interaction. The $E_{XC}[\rho]$ is the exchange-correlation energy term, which has received much examination[37, 38, 39, 40, 41, 42] in the literature and frequently is granted increased accuracy by the incorporation of HF-like expressions in hybrid functionals.

Semiempirical Approaches

Advancements in quantum chemical theory frequently strive to achieve the highest level of accuracy for the depiction of electron-electron interactions, however these developments come at a steep computational cost and are not applicable to large systems

and/or to dynamic simulations where nuclei are allowed to move. Single geometry optimiazations of moderately sized molecules, on the order of 20-30 atoms, can take several hundreds of computing hours to complete with *ab initio* QM techniques and applying these methods to dynamics simulations, where this type of evaluation would need to be completed hundreds of thousands of times, would cause the timescale of the simulation to grow much too long. An alternative approach to alleviate this computational bottleneck is to, over Hartree-Fock and density functional type theories, use 'semiempirical' methods quantum models. In these methods, instead of trying to more accurately describe all subatomic interactions, experimental and higher-level quantum mechanical results are reproduced through parameterization of various subatomic interactions. As one would expect, semiempirical Hamiltonians are limited by the fact that they provide parameterized model, and venturing outside of the designed parameter space of the Hamiltonian can result in erratic behavior. However, semiempirical approximations attempt to address the issue of slow computational speed and poor scaling while maintaining a high level of accuracy. Through careful parameterization and use of the Hamiltonian, semiempirical methods can be made to recapitulate experimental results at chemical accuracy over a restricted range of applications.

Most modern semiempirical methods are an enhanced implementation of the NDDO (neglect of diatomic differential overlap) method. Within this method's approximation, atomic orbitals on different atomic centers do not overlap. Consequently, the overlap matrix is a diagonal, unity matrix. This approximation, in itself, greatly increases computational efficiency by greatly reducing, if not entirely eliminating, the number of 2-electron integrals in the calculation while not directly introducing any extra parameters needed for the calculation. Also, as part of the approximation the 1-center, 1-electron and 1-center, 2-electron integrals quantities are treated as parameters to further reduce computational demand. The oldest, and yet still one of the most widely used, methods using the NDDO approximation is the MNDO, or Modified Neglect of Differential Overlap model. Using the same infrastructure as the NDDO, MNDO methods treat the 2-center, 1-electron integrals. From here, additional parameterizations and implementations of the MNDO framework extended into several other models; MNDO/d, AM1, PM3, and so on.

Of the most relevance to this manuscript is the AM1 method, as a new specific

reaction parameterization of this Hamiltonian is developed in Chapter 2. MNDO, PM3 and AM1, methods all operate under the assumption that the core electrons, those not in the valence level, are assumed to be non-reactive and subsumed into the *core charge* of the nuclear center. This *minimal valence atomic orbital basis* only treats the valence electrons explicitly, reducing computational cost. However, due to this assumption additional potentials were required to accurately model the repulsive nature of atoms experience at short to very-short ranges common in chemical bonding. Assuming that the nuclei and immediately surrounding electronic cloud can be collapsed into a point charge results in too strong of an attractive force when in at distances where the non-valence electronic density could theoretically extend past. Additionally, these potentials subsume some parts to the exchange-repulsion interactions among other interactions and have largely been parameterized to reproduce chemical data for sets of molecules rather than specific atomic or electronic phenomena. The PM3 and AM1 models differ from the MNDO model in that additional repulsive off-center potentials have been added into the formalism to aid the model in obtaining such properties as non-bonded interactions like hydrogen bonding. The PM3 and AM1 models differ themselves through the values of their parameters, however they share the same functional form.

1.2.3 Hybrid Quantum Mechanical/Molecular Mechanical Potentials

When modeling chemical reactions a formalism must be adopted which has some knowledge about the underlying electronic structure of the system. It must know how to model the movement of electronic density in order to be able to model the cleavage and formation of chemical bonds. As discussed in Subsection 1.2.2, one can utilize QM methods to accomplish this task. However when operating within a condensed phase environment, especially when trying to study large biomolecular processes, the system size can be on the order of tens if not hundreds of thousands of atoms and even the most efficient semiempirical methods, which are orders of magnitude faster than contemporary *ab initio* and density functional methods, are still too slow to perform calculations on a system of that size. This issue is compounded when trying to simulate reactions and one must adequately sample phase space to fulfill the requirement of the ergotic principle. Similarly if one were to adopt a purely classical MM force field in these cases, as discussed in Subsection 1.2.1, then while the scaling of the computational cost with

system size would likely allow more sampling during the simulation, the MM model would lack of any formal treatment of the electronic wave function or density making accurately modeling a reaction not possible. This impasse was circumvented by the introduction of hybrid quantum mechanical/molecular mechanical potentials (QM/MM), seen in Figure 1.3 .

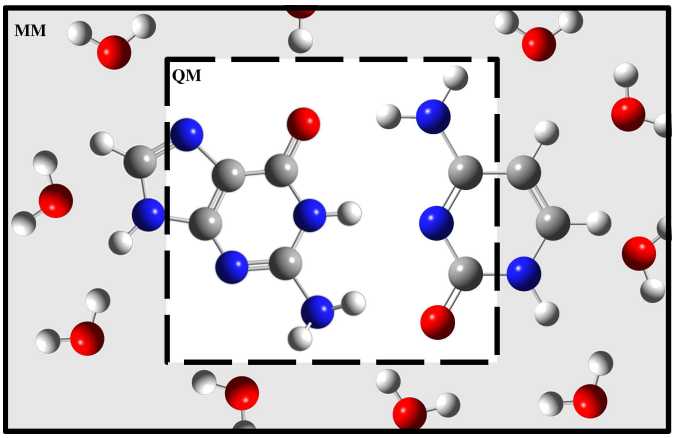


Figure 1.3: **Simple example of the division of the QM and MM subsystems in a QM/MM simulation.** The QM region encompasses the atoms and their immediate surrounding which are involved in the interaction one wished to study with a higher level method. The rest of the system is then segregated into the MM region (shown as gray) and treated with a classical force field. As an example, in the above schematic the hydrogen bond interaction between the base-pairing of guanine and cytosine is enveloped within the QM region while the rest of the system, including solvent, is in the MM region. When studying large biomolecules only a small portion is in the QM region while the rest of the molecule flows into the MM region. The QM and MM regions are coupled through a QM/MM boundary.

In QM/MM simulations, a small active region is modeled through the use of a QM Hamiltonian. This region is kept, generally, as small as possible to maintain computational efficiency and commonly contains only the reactive atoms and their immediate surroundings. The rest of the system is modeled through the use of a classical MM potential. The treatment of how these two regions interact is paramount to the QM/MM formalism. One approach is a "subtactive" scheme in which when adopted the total QM/MM energy determined by calculating the entire system at the MM level, then calculating the QM and MM energies of the active region, then by adding in the QM

energy to the total MM energy while removing the MM energy for the active region. A schematic of this formalism is shown in Figure 1.4. Another approach is to adopt an

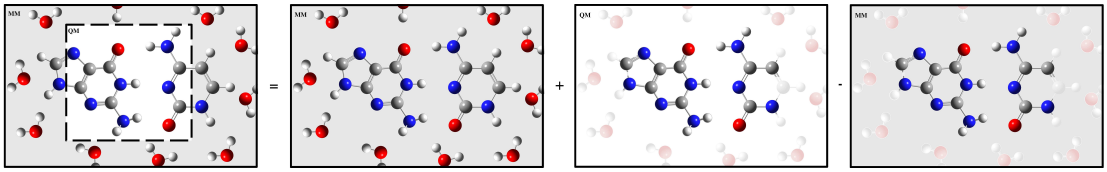


Figure 1.4: **Schematic of the subtractive QM/MM scheme for obtaining the total QM/MM energy.** In this protocol the total QM/MM energy is obtained by calculating the entire system at the MM level, then the reactive region is calculated at the QM level. The MM energy for just the QM atoms is then subtracted from the combination of the first two energies. Special treatment for the interactions which cross the QM/MM boundary may be included if they are desired to be different than the MM force field.

"additive" scheme in which the total QM/MM energy can be written as

$$E_{QM/MM}^{Tot} = E_{\mathrm{QM}} + E_{\mathrm{MM}} + E_{\mathrm{QM/MM}}^{interaction} \tag{1.23}$$

where E_{QM} is the energy of the QM region, E_{MM} is the energy of the MM region and $E_{QM/MM}^{interaction}$ is the interaction energy that arises between the two other regions. A schematic of this formalism can be seen in Figure 1.5. Special attention must be given to the QM/MM interaction energy term. Most generally,

$$E_{\mathrm{QM/MM}}^{interaction} = E_{\mathrm{QM/MM}}^{electrostatic} + E_{\mathrm{QM/MM}}^{non-electrostatic} \tag{1.24}$$

where $E_{\mathrm{QM/MM}}^{electrostatic}$ is an energy term which arises from the QM electrostatic potential interacting with the field of MM charges. This term, once again most generally, can be handled in three different ways. The first is mechanical embedding. In this scheme, QM atoms are assigned point charges which interact with the MM charges in a purely classical manner. While being the most computationally efficient means of treating this interaction, it also leaves much to be desired as there is no meaningful polarization response from either the QM or MM atoms with respect to one another which can lead to significant error. Next the QM/MM electrostatic interaction could be handled with electrostatic embedding, which is generally the most wide-spread method for this

interaction. In it, the QM electrostatic potential is allowed to respond and polarize to the surrounding MM point charges by including them into the QM Hamiltonian. This method allows residues involved in the reaction being modeled to polarize in a meaningful why to their surroundings while still striking a balance in computational effort. Finally, one could employ polarization embedding where both the QM and MM regions are allowed to polarize in some fashion to each other. The details by which this technique are carried out can largely depend on the treatment of polarization in the MM region. This method, while being the most physically realistic treatment of the electrostatic QM/MM interaction, is the most computationally intensive. This work utilizes the electrostatic embedding scheme and all work will be presented as such. Continuing, the QM/MM non-electrostatic term, $E_{\mathrm{QM/MM}}^{non-electrostatic}$, can be divided further as

$$E_{\mathrm{QM/MM}}^{non-electrostatic} = E_{\mathrm{QM/MM}}^{bonded} + E_{\mathrm{QM/MM}}^{non-bonded} \tag{1.25}$$

where $E_{\mathrm{QM/MM}}^{bonded}$ occurs when a QM atom is directly bound to an MM atom, causing a covalent bond to transverse the QM/MM boundary. This term is an active field of study and can be treated a number of ways. The simplest cases is to avoid this term entirely by choosing the QM/MM boundary in such a way to avoid dividing the system across covalent bonds. However, when studying large biomolecules, such a division of space is not always possible and crossing a chemical bond with the QM/MM boundary becomes unavoidable. Link-atom,[43, 44, 45, 46] boundary-atom[47, 48] and localized-orbital[49] schemes are all viable techniques that could be used to treating this interaction.[26] The form of each of these methodologies varies in its implementation, and as the research contained within present work does not encounter this interaction in any of its studies no further discussion is needed.

The final term in the non-electrostatic energy is the $E_{\mathrm{QM/MM}}^{non-bonded}$ energy which represents the non-classical exchange repulsion and correlation dispersion interactions. While these interactions are intrinsically linked to the motions of the electrons of atoms, for simplification of the model these interactions are commonly handled by assigning atom types to the QM atoms in the system and giving them empirical Lennard-Jones parameters, as one would for the MM atom of the system. This treatment is fast and efficient, however it is non-ideal, and research to address the shortcomings of this treatment are

detail in Chapter 3.

The combination of utilizing both QM and MM techniques simultaneously allows for the simulation of condensed phase reactions. This technique has the benefit that it is orders of magnitudes faster than fully QM techniques while still affording some knowledge of the electronic structure of chemically important atoms to the end user. However, this technique has several widely known limitations. First off, the QM Hamiltonian used frequently need to still be an approximate one, as evaluating a complete wavefunction is still too computationally intensive to complete in a timely manner. Secondly if the chemical event being modeled is non-local in nature, either having molecules move in and out of the active site region freely over the coarse of the reaction or having an active site which itself is very large while including several disparate residues, then QM/MM techniques will either be inapplicable or further, more specialized approaches need to be considered.

1.2.4 Implicit Solvation Models

The effects of solvent stabilization has an immense impact on biological and chemical reactions. As such, the treatment of these effects must be considered carefully and taken into account while trying to model condensed phase reactions. However, the computational cost of modeling solvent is not insubstantial. In fact, in standard MD simulations of reactions happening at physiological concentrations (typically on the over of μM concentrations of solute) the number of water molecules surrounding can be on the order of millions of times greater than the number of solute molecules. To alleviate this computational burden explicit water simulations are commonly run under periodic boundary conditions, which simulate an infinitely repeating images of a smaller cavity of the total solvated system whose interaction with the modeled system can be handled with computationally efficient means such as fast multipole and Particle Mesh Ewald methods.[50, 51, 52] Additionally, explicit waters in the modeled system are generally treated with fixed hydrogen atoms which drastically reduces the number of degrees of freedom per water molecule that needs to be propagated.[53, 54, 55, 56, 57] However, even with these advances, explicit water simulation can still represent an insurmountable computational bottleneck when higher-level quantum mechanical methods are required or if the size of the solute being modeled is quite large.

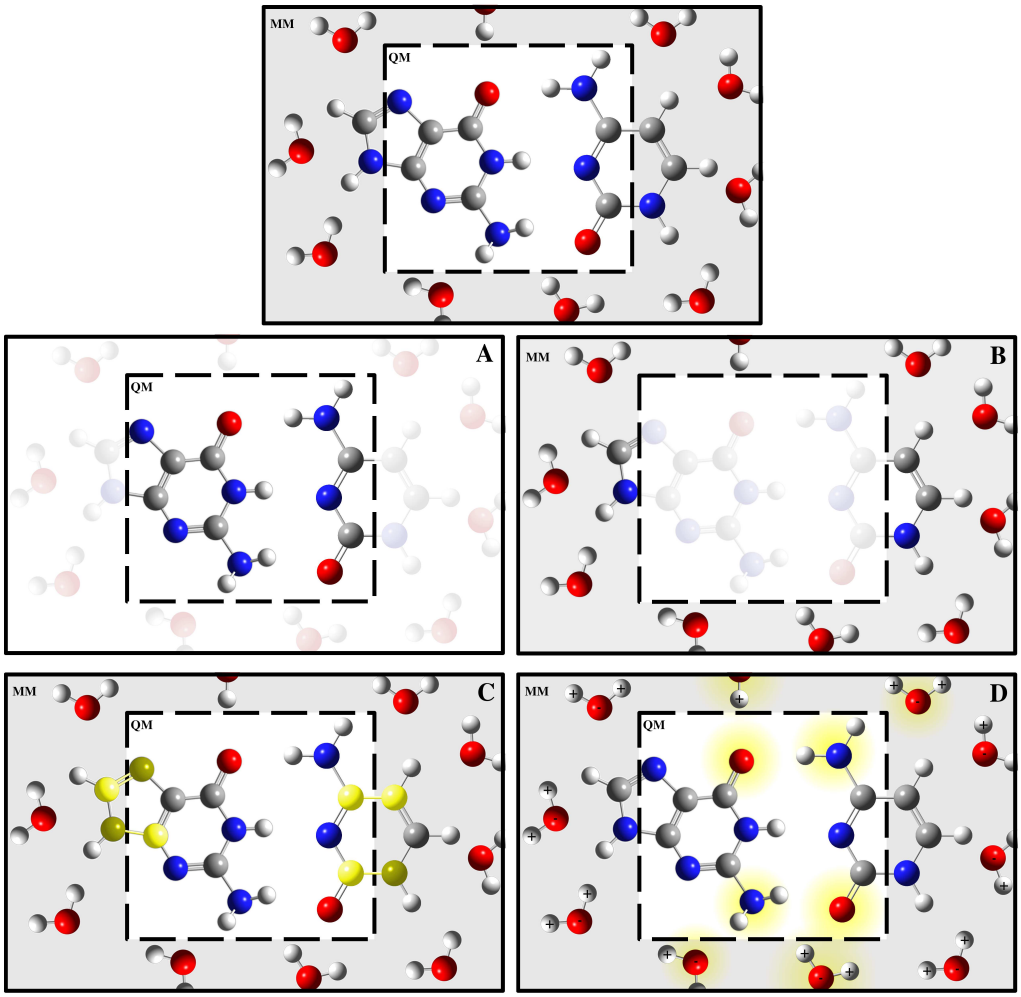


Figure 1.5: **Schematic of the interactions required for calculating the total QM/MM energy in the additive QM/MM scheme.** For this protocol the total QM/MM energy (top) is equal to the total added energy of the four lower diagrams. The first (A) is the total QM energy of the atoms contained within the active region. The next (B) is the total MM energy of all atoms that are not contained within the active region. Then the QM/MM interaction energy is shown in the bottom who pictures (C and D). The QM/MM bonded term (C) can be treated by a variety of specialized techniques while the non-bonded term (D) consists of allowing the QM charges to interact with the QM Hamiltonian and the Lennard-Jones potential to treat exchange repulsion and dispersion.

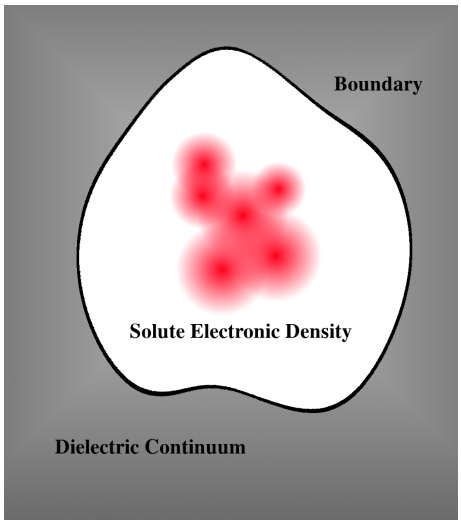


Figure 1.6: **Cartoon of an electronic distribution contained within an implicit solvent dielectric continuum.** Implicit solvent models replace explicitly modeled solvent waters with a dielectric continuum. This drastically reduces the number of degrees of freedom which need to be modeled in condensed phase simulation. Solute molecules interact with the continuum through the implicit solvent boundary, which generally serves as a bulk screening agent and responds to the electronic density within the solvation cavity.

To this end, one could employ the use of an implicit solvation model to eliminate the need for any explicit treatment of any water. These methods represent the solvent in an average way by modeling it through the use of some sort of dielectric continuum, see Figure 1.6. The fundamental idea behind implicit solvation is that the dielectric continuum serves as the average representation of the bulk solvent surrounding the solute which is placed inside of a cavity within the continuum. The cavity is determined by generally by some radial term about atoms which define the continuum boundary. This boundary can be determined in a number of different ways, some of which are outlined in Figure 1.7. The solute is then placed inside this cavity and the electronic density of the solute interacts with the continuum boundary to model solvation effects. Generally, the goal of these models is to predict the solvation free energy (ΔG_{solv}) of a system, the energy associated with moving that system from the gas phase into solution.

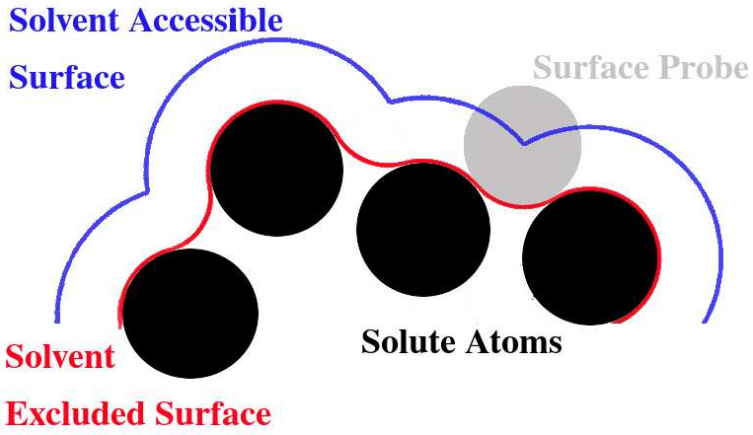


Figure 1.7: **Cartoon of different implicit solvation model boundary types.** There are many different interpretations of where the boundary of an implicit solvent model should belong. Two of the most common are the Solvent Excluded Surface (SES), or van der Waals surface, and the Solvent Accessible surface (SAS). Generally, each can be determined by 'rolling' a probe that is representative of the size of a solvent molecule along the canvitation radii of the atoms of the solute. The minimum contact distance will trace the SES while the SAS will be traced by the center of the solvent probe.

Numerous different continuum models have been developed; Poisson-Boltzmann (PB),[58] Minnesota solvation (SMx),[59, 60, 61] conductor-like screening (COSMO),[62, 63, 64] general Borne (GB),[65, 66] polarizable continuum (PCM)[67, 68, 69] models as

a small set of examples. All of these models approach the question of implicit solvation from a slightly different perspectives, each with their own strengths and weaknesses. One of the major drawbacks to implicit solvation models is inherent with their strengths; because of the average representation of the solvent environment, which provides the gained computational efficiency, specific solute/solvent interactions cannot be accurately modeled (unless these interactions are known *a priori*, are long lived and included within the solute cavity). For instance, if transient water or salt bridges are important to the overall stability to a protein being modeled then an implicit solvation model my completely neglect these interactions. Additionally, solvation effects are highly dependent on and sensitive to the definition of the solute cavity. Care must be taken when selecting an implicit solvation model to make sure it is well parameterized for the given application However, even within these concerns in mind implicit solvation models provide a useful tool for representing solvent properties in a cost-effective manner, average manner.

1.3 Techniques to Aid Computational Sampling

Frequently, simulations are run not to obtain the motions of molecules directly (which in itself would be a tenuous practice as the precision needed to propagate motion accurately at a molecular level is very computationally demanding) but rather to calculate some other physical observable, like a free energy or the density of water, by sampling the configurations generated by a model. Then, using the frequency that the model visits particular configurations data can be obtained. Using a strategy first developed by Metropolis *et al.*[70], probability in which configurations are visited by the modeled system will correspond to ensemble averages, which in turn can be related to physical observables. However, while it is theoretically possible to perform conventional molecular dynamics to sample the whole of phase space, in practice molecular dynamics is too inefficient to address exploring chemical reaction energy landscapes in this manner. If attempted, the model would likely remain in a local energy basin, repeatedly sampling a small (and likely generally uninteresting) area of phase space.

To overcome this issue and to sample phase space in a more efficient manner, specialized techniques have been developed which allow for sampling specifically selected,

narrow but important, regions of phase space. These areas are generally involved in chemical interactions, rearrangements or reactions and involve crossing over relatively energetically unfavorable areas of phase space directly. It is important to obtain a representative distribution of states in these particular areas of phase space so that information can be gained about how the relative energy and probably of visiting of these states compares to that of the energetically favorable minima basins in the surrounding neighborhood. These *enhanced sampling* techniques have proven themselves to be powerful tools, despite all of them requiring some amount of knowledge about the system *a priori* to properly implement. If care is not taken with utilizing these techniques, then the area of phase space which is sampled may be artificially limited and will result in and incomplete picture, and likely an altered value of the desired observable.

The following subsections will discuss two of the many enhanced sampling techniques available in modern molecular modeling; umbrella sampling and thermodynamic integration. Each of these techniques were used extensively in the research presented in this work and, while no fundamental developments where made to either of these techniques, an understanding of why each technique was chosen and will provide some insight into the findings and results of the research herein.

1.3.1 Umbrella Sampling

A term first coined by Torrie and Valleau[71], umbrella sampling, conventionally refers to a technique used in molecular modeling in conjunction with molecular dynamics simulations to more expediently sample a specific area of phase space. This task is accomplished through the inclusion of a fictitious 'biasing force.' This force, whose potential can be of any form, changes the underlying potential felt by the modeled atoms and causes the simulation to preferentially favor one particular neighborhood of states, as defined by the user. This approach proves particularly useful, as one can bias their simulations to favor energetically unstable or high-energy states as compared to meta-stable or lower-energy states, allowing for more complete sampling of these high-energy configurations which then in turn provides a more complete picture of these states and more accurate estimations on the physical observables correlated with the free energy landscape.

The most widespread use of this technique is by defining the bias potential to be a

harmonic restraint along some user-defined *progression coordinate*,

$$U_{bias}(\xi) = k(\xi - \xi_0)^2, \tag{1.26}$$

where $U_{bias}(\xi)$ is the biasing potential, ξ_0 is the location progression coordinate (sometimes also referred to as the anchor position), and k is the a parameter which determined the strength of the restoring force. A brief schematic on how umbrella sampling can enhance local sampling can be found in Figure 1.8. The progression coordinate, ξ, is arbitrary and can be defined in any number of ways. However, it is common to construct this coordinate in terms of geometric properties, such as bond distances or angles, of the particles undergoing some chemical event. For instance, later in the manuscript biasing potentials are used in a bimolecular nucleophilic substitution (S_N2) reaction and ξ is defined as the difference between the distances of the nucleophile and the leaving group with respect to the central atom chirality inversion. In cases like this, it is also common to refer to the progression coordinate as a *reaction coordinate* as it is indicative of progress in a chemical reaction.

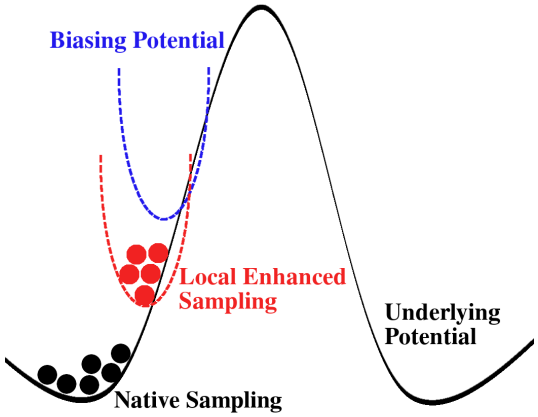


Figure 1.8: **Diagram of how the use of an umbrella potential in umbrella sampling allows for locally enhanced sampling of a potential energy surface** Without the use of an umbrella potential simulations would not be able to surmount large barriers in the underlying potential, and would be restricted in the area of native sampling. However, when a biasing potential is applied, simulations can sample these higher energy states.

After defining a progression coordinate, ξ, then several different *windows* must be simulated at different positions along that coordinate, ξ_0. Afterwards, specialized techniques such as the weighted histogram method (WHAM),[72, 73, 74] the multistate Bennett acceptance ratio estimator (MBAR),[75, 76, 77] or the variational free energy profile method (vFEP)[78, 79] can be used to remove effects of the biasing potential from the simulated results and reconstruct a free energy profile based off of the probability density of states across all windows.

There are a few shortcomings which can arise due to improper use or application of this technique. If the progression coordinate is ill-defined, either because it restricts the model to not sample important areas of phase space or there does not exist a coordinate (which can even be a combination of several different independent coordinates) that captures the necessary motions for meaningful simulation of the system, then the use of umbrella sampling will yield poor results. Additionally if when defining the progression coordinate the values of either ξ_0, by being too far apart, or k, by being too strong or too loose, do not allow for adequate sampling in and between the different windows on the simulation then the resulting estimations of the free energy landscape will be erroneous.

1.3.2 Alchemical Perturbation and Thermodynamic Integration

Another method to generate free energy data with molecular simulation is to utilize free energy perturbation (FEP) to sample along an "alchemical" pathway[80, 81, 82, 83] and then to connect the different states with Thermodynamic Integration (TI). Originally set forth by Kirkwood,[84] and then further refined by others,[85, 86, 87] TI can utilize a non-physical pathway, which may be include intermediate steps in which atoms partially exist or have fractions of electrons assigned to them, to obtain experimentally relevant free energy data as long as the two end states of this "transmutation," have physical meaning. Intuitively, concept has merit as free energy is a state function and, as all state functions, the path taken to obtain this value does not matter. Rather, the relative difference between the each state dictates the value of this observable. More rigorously, one can state a free energy as

$$A(N, V, T) = -\beta^{-1} \ln(Q) \tag{1.27}$$

where

$$\beta = \frac{1}{k_b T} \tag{1.28}$$

where k_b is the Boltzmann constant, T is the temperature in Kelvin and Q is the partition function

$$Q(N,V,T) = \frac{1}{N!\Lambda^{3N}} \int e^{-\beta U(p_1 \ldots p_N, x_1 \ldots x_N)} d^3p_1 \ldots d^3p_N d^3x_1 \ldots d^3x_N \tag{1.29}$$

where N is the number of particles in the system and x_1 and p_1 are the position and momenta of particle 1 respectively, and so on. In this example the Helmholtz free energy was used, however this type of derivation can be extended into other ensembles[88] and the current selection is arbitrary. The total energy of a given combination of positions and momenta is represented as U. If there are two different systems which experience two different potentials, U_1 and U_2 then it is possible to connect these two different potentials through one "metapotential" which utilizes a perturbation progression variable, λ, to dictate how far between each potential the simulation is located. This new, combined potential would take the form[89, 90]

$$U(\mathbf{x}, \mathbf{p}, f(\lambda)) = (1 - f(\lambda))U_1(\mathbf{x}, \mathbf{p}) + f(\lambda)U_2(\mathbf{q}, \mathbf{p}). \tag{1.30}$$

where $f(\lambda)$ is some switching function such that

$$U(\mathbf{x}, \mathbf{p}, f(\lambda)) = \begin{cases} U_1(\mathbf{x}, \mathbf{p}) & \text{if } \lambda = 0 \\ U_2(\mathbf{x}, \mathbf{p}) & \text{if } \lambda = 1 \end{cases} \tag{1.31}$$

but other than these constraints $f(\lambda)$ had arbitrary form. If one were to then take the derivative of that free energy, A, with respect to the perturbation progression variable,

λ, one would get

$$\begin{aligned}\frac{\partial A(\lambda)}{\partial \lambda} &= -\frac{1}{\beta Q}\frac{\partial Q}{\partial \lambda} \\ &= \frac{k_b T}{Q}\frac{\partial}{\partial \lambda}\int e^{-\beta U(\mathbf{x},\mathbf{p},\lambda)} d^3\mathbf{x} d^3\mathbf{p} \\ &= \frac{k_b T}{Q}\int \left(-\beta \frac{\partial U(\mathbf{x},\mathbf{p},\lambda)}{\partial \lambda}\right) e^{-\beta U(\mathbf{x},\mathbf{p},\lambda)} d^3\mathbf{x} d^3\mathbf{p} \\ &= -\left\langle \frac{\partial U(\mathbf{x},\mathbf{p}\lambda)}{\partial \lambda} \right\rangle\end{aligned} \tag{1.32}$$

Then by stating the relationship that the free energy of moving a system from U_1 to U_2 with absolute free energies A_1 and A_2, respectively,

$$\Delta A = A_2 - A_1 \tag{1.33}$$

by following the progression variable λ is equal to integrating over all possible values of λ

$$\Delta A(\lambda) = \int_0^1 \frac{\partial A}{\partial \lambda} d\lambda \tag{1.34}$$

which would lead to the conclusion that the fee energy difference between two states can also be obtained by

$$\Delta A(\lambda) = \int_0^1 \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda d\lambda \tag{1.35}$$

where $\langle \dots \rangle_\lambda$ is the ensemble average at a given value of λ. Practically, this means that when running a simulation if one is able to calculate the value of $\frac{\partial U}{\partial \lambda}$ while constraining a system at a particular value of λ then, no matter the path between U_1 and U_2 which is dictated by λ, it should be possible to determine the relative free energy difference between those two states. Therefore, transformations can proceed along an alchemical pathway which are not limited to 'physical' changes and can, for example mutate one atom of a particular element into a different element or pathways could lead functional groups to be perturbed into nothing.[91]

For all intents and purposes direct computation of Equation 1.35 is not feasible, as it would require an infinite number of λ windows to be simulated over with infinitesimal spacing. However, within reasonable approximation the integral within that equation

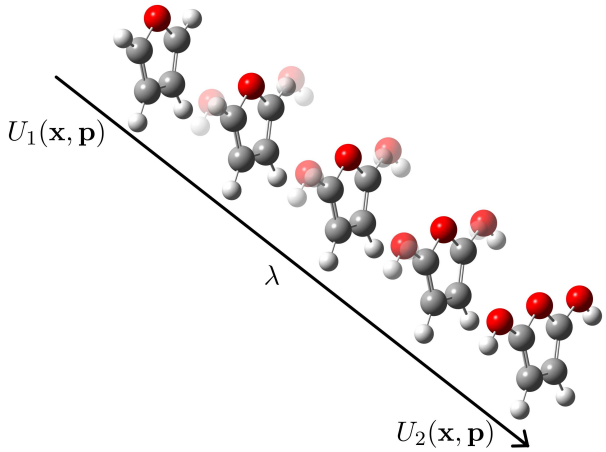


Figure 1.9: **Example of an alchemical pathway in which one compound is transmuted into another via free energy perturbation (FEP) methods.** FEP methods can link two different compounds with an "alchemical pathway" to allow for efficient simulation of traditionally challenging to calculate free energies. For instance if the pathway shown above, where the selected hydrogen atoms are directly transformed into hydroxyl groups, was followed in both solution and gas phases one could calculate the relative free energy of solvation , $\Delta\Delta G_{solv}$, between furan and 2,5-hydroxyfuran.

can be replaced with a Riemann sum;

$$\Delta A = \sum_{i=1}^{n} \left\langle \frac{\partial U(\mathbf{x}, \mathbf{p}, f(\lambda))}{\partial \lambda} \right\rangle_{\lambda_i} \Delta\lambda_i. \tag{1.36}$$

where n is the number of λ windows, each run at a specific value of λ_i to determine the value of $\left\langle \frac{\partial U(\mathbf{x},\mathbf{p},f(\lambda))}{\partial \lambda} \right\rangle_{\lambda_i}$. The severity of the errors due to this approximation is directly related to the number of λ windows simulated (n) and the form of the metapotential coupling the two different states, $U(\mathbf{x}, \mathbf{p}, f(\lambda))$. Commonly a linear coupling scheme is used to connect the two end-state potentials and only a few, somewhere around 3 to 5, λ windows are required to obtain reasonably accurate data, assuming smooth convergence of $\left\langle \frac{\partial U(\mathbf{x},\mathbf{p},f(\lambda))}{\partial \lambda} \right\rangle_{\lambda_i}$ during the simulation. Then simple extrapolation protocols can be used to fit the obtained data and an accurate result can be obtained. However if a more complex scheme for $f(\lambda)$ is used, generally in order to avoid computational pitfalls associated with a linear coupling scheme, then more λ windows may be needed to reduce the error due to replacing the integral with a Riemann sum.

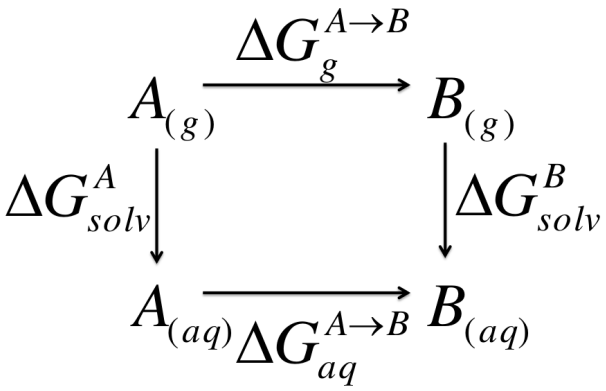


Figure 1.10: **Example thermodynamic cycle which could be used for calculating $\Delta\Delta G$ of solvation between two different compounds.** In order to calculate the relative solvation free energy, $\Delta\Delta G^{AB}_{\text{solv}}$ between two compounds, A and B, the above thermodynamic cycle could be used where $\Delta\Delta G^{AB}_{\text{solv}} = \Delta G^{B}_{\text{solv}} - \Delta G^{A}_{\text{solv}}$ and $\Delta G^{A}_{\text{solv}}$ is the free energy of moving A form the gas phase into solution and $\Delta G^{A\to B}_{\text{g}}$ is the free energy for transmuting A directly into B while in the gas phase, and so on. Keep in mind that due to the cyclical nature of this calculation only three of the free energy transformations shown above need to be simulated, as the fourth can be calculated directly from the other three. This is useful if one of the transformations is particularly challenging, or this information could be used to help verify the validity and accuracy of the simulations performed by completing the full cycle and checking whether the simulated results from each transformation agree with the sum of the other three.

FEP with TI is a very powerful tool for computational biochemistry as it allows for the calculation of the relative free energy between two different system states. Direct calculation of pK_a values[92], solvation free energies[93, 91, 82] and binding free energies[94] can be accomplished in a straightforward manner; commonly through the use of a thermodynamic cycle, see Figure 1.10.[95, 96, 97] This capability provides a distinct advantage to these methods over umbrella sampling-type methods as no specific reaction coordinate must be determined *a priori*. For instance, when using umbrella sampling if one were to try to determine the free energy of binding of a small organic compound into the active sight of a protein then a specific reaction coordinate would have to be determined which could move the small molecule from bulk solution into the protein scaffold. With subtle rearrangements and structural changes associated with ligand binding, this reduction in system degrees of freedom frequently not possible. One could try to adopt a procedure in which several different reaction coordinates are probed simultaneously, however the computational cost of these simulations scale exponentially with the number of additional reaction coordinates needed. FEP on the other hand, can probe this event directly by simply transmuting the bound ligand into the bulk solution.

Chapter 2

Modeling the S_N2 reaction of Cl^- on CH_3Cl with a Specific Reaction Parameterization of Semiempirical AM1 Hamiltonian

In order to parameterize the subtle interactions which are modified by the model described in Chapter 3, a base-line reaction had to be established. Ideally, this reaction would involve biologically active atoms; such as oxygen, nitrogen, phosphorus and hydrogen; would have a large amount of electronic transfer from among reagents, would have a small number of degrees of freedom; such as different elements involved in the reaction; and would have a large amount of experimentally known solvation free energies of compounds in or related to the reagents of the modeled reaction. Clearly, this list of criteria is complex and slightly restrictive.

Initially, due to the key roll oxygen plays in the phosphoryl transfer reaction outlined in Section 1.1, the different protonation states of water were considered as a test system. However, the exact nature of how hydroxide, hydronium or even a water molecule behave

[1] This chapter compiles material reprinted with permission from Kuechler, Erich R., and Darrin M. York. "Quantum mechanical study of solvent effects in a prototype S_N2 reaction in solution: Cl attack on CH_3Cl." The Journal of chemical physics 140.5 (2014): 054109. Copyright 2014, AIP Publishing LLC.

in bulk water is still not entirely clear[98, 99, 100, 101, 102] combined with the common treatments of hydrogen atoms in molecular simulation, frequently using them as point charges with no non-bonded non-electrostatic interactions or as 'fudge factors' to these interactions of other atoms, this set of systems did not appear to be an ideal choice for a launching-off point for the model and future parameterizations of it. After some consideration, the system of the attack of the chloride anion on methylchloride was thought to be sound. This reaction has been extensively studied both computationally and experimentally studied, involves a significant amount of charge transfer over the coarse of the reaction, involves only three different elements, and the reagents of the reaction had experimentally known solvation free energies. While this reaction did not fulfill the criteria of being particularly biologically motivated, once a reasonable set of parameters for the reaction were obtained then they could be used in similar reactions to acquire parameters for other elements.

Once the chloride/methylchloride symmetric reaction was chosen, models were assessed to determine whether an existing semiempirical Hamiltonian could accurately model this reaction. While some did reasonably reproduce the solution phase barrier none were able to capture both the gas phase and solution phase barriers simultaneously, as discussed and shown later in this chapter. This result was not surprising as most QM/MM simulations in the literature today do not involve reactive chlorine, rather chloride anions are mainly used in the MM region for their bulk properties. With the need for an accurate QM method to handle reactive chlorine, it was decided that a specific reaction parameterization of the AM1 Hamiltonian was needed. The remainder of this chapter will detail the development and performance of this new model.

2.1 General Background

Ab initio and density functional quantum chemical calculations are powerful tools which can be used for highly accurate predictions of gas phase geometries and reaction energetics.[103, 104, 105] However, when systems are in a condensed phase environment extensive conformational sampling is required and the computational cost of these quantum mechanical (QM) calculations can pose a practical bottleneck. One

strategy to overcome this limitation and obtain QM level accuracy in molecular dynamics simulations at a reduced computational cost is to use a hybrid quantum mechanical/molecular mechanical (QM/MM) potential [106, 107, 108, 109] where the system is divided into two different regions; a small active site region containing any chemically active atoms with their immediate neighbors to be modeled with QM techniques and then the rest of the system is modeled with classical molecular mechanics (MM). Even within a QM/MM framework, however, when interrogating reaction pathways of large biological systems, where QM regions can be one hundred atoms or more and the total system size can be on the order of hundreds of thousands of atoms, traditional QM techniques may still be prohibitive. As a result, the last decade has seen a resurgence of interest in the development of fast, approximate "semiempirical Hamiltonian" quantum models for use in QM/MM simulations. Specifically parameterized semiempirical models have shown themselves to be invaluable at exploring different classes of highly specialized reactions. [110, 111, 112, 113]

This chapter will outline a procedure for creating an Specific Reaction Parameterization (SRP) quantum model for chlorine within the Austin Model 1 (AM1)[114] Hamiltonian for the symmetric bimolecular nucleophilic reaction (S_N2) of a chlorine anion attacking methyl chloride in both the gas phase and in solution. Simple S_N2 reactions, such as the focus reaction of this study, are some of the most fundamental reactions in organic chemistry and have provided keen insights into gas phase reactivity and aqueous, solvation properties, reaction energetics and kinetics over the past several decades. [115, 116, 117, 118, 119] It is well known that S_N2 reactions, when in the gas phase, have an double-well potential energy surface with a single transition state (TS), as seen in Figure 2.1, bracketed on either side by low wells generated from the formation of a more energetically stable Ion-Dipole Complex (IDC). Then, when the reaction is moved into solution, the reaction profile is likely to be dominated by an enhanced central barrier. This change in the general shape of the reaction energy surface is due to destabilization of the TS and the IDC. Relative to the reactants/products, both of these complexes have an increased solvent-excluded space because they are generally much more diffuse than either of the separated reagents. As a result, solvation strongly favors the reactant state over the TS and IDC, leading to a pronounced single barrier and less prominent IDC stabilization. The solvation stabilization of ionic species and

another, possibly dipolar, molecule at infinite separation is much greater than for the TS. Although the TS bears a charge, the solvation cavity it creates is considerably larger than those made from either of the reactants. Also, if the reaction is symmetric, the TS will not have a net dipole moment. Similarly, solvation destabilizes the IDC relative to the reactants, although to a smaller extent than the TS, causing a drastic reduction of the depth of the corresponding minima in the free energy profile.

The study of chloromethane and other small halogenated hydrocarbons are not only of central interest to understanding a wide range of fundamental chemistry, but are of health and environmental interest in themselves.[120, 121, 122, 123, 124] Having been so extensively studied, by both experimental [125, 126, 127, 128, 129] and computational[130, 131, 132, 133, 134, 135, 136, 137, 138] means, lends this class of reactions to be an outstanding starting position for benchmarking and refining new computational methods and techniques. In particular, the chloride/methyl chloride reaction is a strong candidate as a prototype reaction as it has be quite extensively studied in order to understand the exact nature of the transition state in the gas and condensed phases.[139, 140, 141, 142] Furthermore, by starting out with a small system which has inherently less degrees of freedom in interaction and parameter space, one can refine parameters more easily and then proceed to build on the system from that strong foundation.

Very recently, a similar set of reactions has been the focus of semiempirical model development and application in QM/MM simulations by Liang *et al.*[143] This work performed critical examinations of the S_N2 reactions of chloride and fluoride attacking methanol and develops a new AM1 parameterization using the internal reaction coordinates of each respective gas phase reaction to include reaction energetics and force-matching to MP2 reference data. This strategy has been demonstrated to be a successful one in subsequent QM/MM applications. In the present chapter, we examine a different reaction and utilize a different approach in parameterization strategy that includes sets of constrained geometry optimizations at different stages along the reaction coordinate, as well as properties of the isolated reactant species, and demonstrate that it is robust and transferable in application to QM/MM simulations in solution. Results of the simulations provide insight into the origins of solvation effects on the reaction.

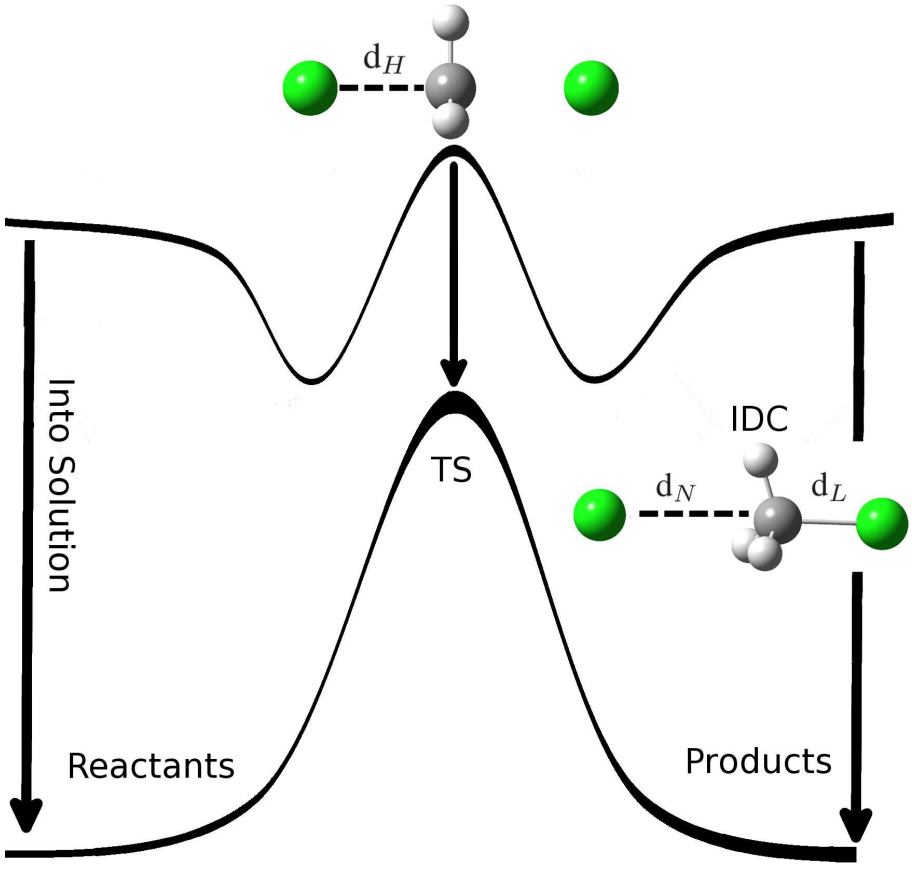


Figure 2.1: **Schematic representation of the effects of solvation upon the potential energy of a symmetric S_N2 reaction.** Specifically, the attack of a chloride anion on methylchloride is shown. Moving from the gas phase into solution frequently significantly reduces the energy associated with the presence of an ion-dipole complex (IDC) and generally increases the attack barrier as charge is concentrated into a larger intermediate system in the transition state (TS). The nucleophilic attack distance (d_N), the leaving group distance (d_L) for the IDC and the halide distance in the TS (d_H) are labeled.

2.2 Computational Methodologies

The bimolecular identity reaction of a chloride anion attacking methyl chloride via an S_N2 mechanism both in the gas and solution phases are examined for the purposes of creating a SRP semiempirical Hamiltonian. This chapter explores the use of a stepwise approach to the parameterization and testing of new models to study chemical reactions in solution. The first step is to perform benchmark gas phase quantum chemical calculations to obtain a target reaction profile, as well as key molecular properties of the reactants and products including geometries, electron affinities, dipole moments. All values are validated the data against available literature values where possible. The second step is to use the most accurate high-level QM results, in conjunction with known experimental values where appropriate, from the methods tested as a training set to parametrize an accurate semiempirical Hamiltonian model that is orders of magnitude faster than any of the benchmark-level models allowing it to be utilized in molecular dynamics simulations. The parametrization will only focus on recapitulating the gas phase quantum chemical data, with no other modulation of the parameters to better performance in solution. In the third step, the new semiempirical model will be tested in QM/MM simulations using standard molecular force field van der Waals parameters to capture free energy profiles for the target reaction. Results will be scrutinized against experimental and other semiempirical data to determine the accuracy of the SRP Hamiltonian for solution phase reactions. The following subsections detail the computational methods used for each of these steps.

2.2.1 Benchmark Quantum Chemical Calculation Protocol

In order to establish a baseline onto which the SRP Hamiltonian will be optimized a number of different quantum models were tested as to determine which would best reproduce experimental gas phase reaction data. The models chosen for this task were Hartree Fock (HF) and the Second-order Møller-Plesset (MP2)[144] methods. Additionally these *ab initio* methods were compared along side a series of density-functionals including BH&HLYP,[145] B3LYP,[146] ωB97x-D[147] and M06-2X.[148] Each of the tested QM methods perform differently depending largely on their general formulations, therefore it is important to test each to determine which best captures the physics of

the reaction of interest.

HF has several known issues with chemical bonding[149] however it does provide the exact exchange while being computationally less demanding than several other *ab initio* and hybrid density functional methods. MP2 is the only perturbative method chosen for our bench-marking calculations; being fairly widely used and having been shown to be highly accurate, even when given small to moderately sized basis sets, for small molecules and their reactions, however computational cost can become an issue if system size becomes too large. [150] Several different hybrid density functionals were examined; BH&HLYP, B3LYP, ωB97x-D and M06-2X, each has different strengths, related to each of their formulations and treatment of the exact (HF) exchange. BH&HLYP uses equal parts of the Kohn-Sham exchange energy and of Becke's[145] correction to the local spin density approximation (LSDA) while the B3LYP functional contains each of these, with different weightings, while also having three other fit mixing parameters relating to the exchange or correlation energies. M06-2X weights the exact HF exchange more highly than either of the two functionals at 54% in conjunction with several other empirical functions. The ωB97x-D is a long-range corrected (LC) hybrid density functional, which uses 100% of the HF exchange to describe distant electron-electron interactions and then uses parametrized functionals and general gradient approximations (GGA) for shorter, close-range, interactions. To enhance short-range behavior, both the ωB97x-D and M06-2X functionals also use Grimme's[151] DFT dispersion correction with weighting factors of 1.0 and 0.06 respectively. Finally, the M06-2X functional also includes meta-GGA character to further differentiate itself from the other functionals in this test set.

As for why each of these functionals were chosen initially; B3LYP was an obvious choice to test for a model function, as a majority of density functional computational studies today use this method due to its generally good performance and low computational cost. The BH&HLYP was included because other similar studies noted that this functional performs quite well for capturing energy barriers for small reactions, however it has also been noted that BH&HLYP does not correctly predict reaction thermodynamics.[134] The ωB97x-D functional has shown superior performance for measuring both bonded and non-bonded interactions in the LC hybrid class of functionals and finally M06-2X is the recommended functional of the M06 suite for main group chemistry, which has been shown to accurately predict thermochemistry, kinetics

and non-covalent interactions for non-multi-reference systems.

To calculate gas phase reaction energies, first the stationary points of the reaction were determined; including the IDC and the TS. The TS was verified as a first order saddle point through normal mode analysis which reported a single negative eigenvalue in the Hessian, whose mode amplitude and motion were along the predicted reaction pathway. Furthermore, intrinsic reaction coordinate profiles were calculated by following the steepest gradients down from the TS to ensure that the TS correctly connected the reactant and product minima. Reactants and products minima were calculated at 'infinite separation' and then added to calculate their respective energy basins. All gas phase quantum benchmark calculations performed were carried out using the Gaussian09 software suite[152] in conjunction with the GaussView tool.[153] Each of the tested high-level quantum methods were geometry optimized and their energies evaluated at the 6-311++g(3df,2p) level of theory with tight optimization and tight SCF convergence criteria while using an ultrafine integration grid. This work was carried out using hardware and software provided by the University of Minnesota Supercomputing Institute and the supercomputing facilities at Rutgers, the State University of New Jersey.

After selecting a quantum method to act as the model for the new semiempirical Hamiltonian, it is used to create gas phase reaction profiles. To create the reaction profiles, constrained optimizations were performed where, departing from the TS, the nucleophile was incrementally pulled away from the carbon which it was attacking until well passed the minima of the IDC. Two constraints are used: one distance constraint between the attacking nucleophile and the carbon center of S_N2 inversion, and another maintaining the attack angle between the nucleophile, the carbon center and the leaving group at 180°. These two constraints are adequate for this study because the reaction examined in this paper is symmetric; the nucleophile and the leaving group are identical. If this was not the case, it would be important to examine the departure of the leaving group with additional points and constraints during and after the transition state. Additionally, calculations were made of each individual reagent in order to properly calculate the energy of the reactants and products at 'infinite separation.' With the listed constraints imposed, geometries of several different points of reaction coordinate progress were optimized using the selected functional at the 6-31+g(d,p) level of

theory. Energies were then recalculated at the 6-311++g(3df,2p) level at the previously optimized geometry, each performed using an ultrafine integration grid and tight convergence criteria. After the reaction profile was created, it was verified to be representative of the reaction through cross examination with the intrinsic reaction coordinate (IRC) profile.

Where possible, results in the gas phase are compared with computational benchmark results from the W1' method by Parthiban *et al.* [134] The W1' method is a variation on the Weizmann-1 (W1) procedure to improve performance on second-row compounds, using a combination of high basis set level B3LYP, CCSD and CCSD(T) calculations to extrapolate to infinite basis set limits and accurately account for an-harmonic zero-point energy, relativistic effects and other computationally challenging energetics. Typically, this procedure yields results that are accurate within 0.25 kcal/mol for thermochemical data.[154, 155]

2.2.2 Parameterization of the SRP AM1 Hamiltonian

Once the high-level reference gas phase reaction profile was obtained, it was used to generate an objective function for the parameterization of the new AM1 Hamiltonian. A weighted chi-squared function was used, taking into account energetics and molecular geometries of each of the points in the reaction profile outlined in the previous section. Including many points along the reaction coordinate allows the new model to capture subtle features over the course of the reaction, such as curvature of the adiabatic energy surface. Identical constraints as those in the benchmark quantum calculations were also imposed on the optimization procedure for regeneration of the gas phase reaction profile. Extra weight was given to the important minima and saddle points in the chi squared function, as recapitulating the reaction barrier was thought of as the most important characteristic to obtain. Reagents and other similar molecules were also included into the objective function. Important physical observables, such as the dipole moments of relevant reagents and electron affinities of select particles where also taken into account. The parameterization scheme used involved a multi-step optimization in which the objective function weights were modulated in order to quickly converge the parameters of the new Hamiltonian. Initially, geometric quantities were given higher weights relative to recapitulating energetics because small changes to bond lengths and angles result

in drastic shifts in potential energy which could cause instabilities in the parameter optimization . Once these initial parameters are converged, meaning that the parameters for the geometric quantities are relaxed, the objective function weights are then adjusted such that the energetic terms contribute in a more significant degree to the chi squared function. Then parameters were re-optimized and more finely tuned in order to capture reaction energetics. Full details of the merit function and optimization protocol used to derive the new parameters, including each molecular species and their starting geometries, the molecular refinement procedure including all constraints, all properties and their reference values and chi squared weight can be found as a documented input file which was used to run the calculation in the Appendix B. Optimization departed from the original AM1 parameters for chlorine and was accomplished using a direction set method[156] in conjunction with the MNDO97[157] software package.

2.2.3 Protocol used for Molecular Simulation

After a new Hamiltonian was parametrized, molecular dynamics simulations were carried out using AMBER12.[158] Potential of Mean Force (PMF) simulations were performed using both the TIP3P[159] and TIP4P-Ew[160] water models according to the following protocol: 1.) Generate structure and solvent box using the leap series of programs. Structures were generated using ANTECHAMBER from optimized TS Gaussian output files and force field modification files for the parameters of methyl chloride and the chloride anion were generated. Lennard-Jones parameters for all chlorine atoms were selected from the Joung and Cheatham work[161] for each of their respective solvents because the authors felt that, over the coarse of the reaction, the chlorine atoms maintain a strong chloride-like charge-state and character. All Lennard-Jones parameters used can be seen in Table 2.1. Using Leap, the solute was given an 15 Å solvent buffer radius in a cubic cell for periodic simulations.

	TIP3P		TIP4P-Ew	
Element	$R_{min,ii}/2$	ϵ_{ii}	$R_{min,ii}/2$	ϵ_{ii}
OW	1.7683	0.1520	1.775931	0.16275
HW	0.0000	0.0000	0.0000	0.0000
C	1.9080	0.0860	1.9080	0.0860

H	1.4870	0.0157	1.4870	0.0157
Cl/Cl^-	2.5130	0.0355910	2.760	0.0116615

Table 2.1: **Lennard-Jones parameters used in TIP3P and TIP4P-Ew explicit water QM/MM simulations.** Parameters for the chlorine and chloride atoms are taken from the work of Joung *et al.*[161]. $R_{\min,ij}$ and ϵ_{ij}, values are in Å and kcal/mol respectively.

2.) Equilibration of the solvent box was performed for 250 ps using an NPT ensemble with the solute harmonically restrained with a 500 kcal/mol/Å^2 force constant. The purpose of this step is to relax the solvent box around the initial state and allow the solvent box density to stabilize. 3.) Several different umbrella windows along the reaction coordinate are created. The system is then allowed to equilibrate at each umbrella windows. This step of equilibration was performed for 150 ps in an NPT ensemble. 4.) Finally production was ran for 150 ps in an NVT ensemble. Production trajectories were then analyzed using MBAR[75] in conjunction with a kernel density estimator in order to calculate the solution reaction profile.

2.3 Results and Discussion

The goal of this chapter is to create a SRP AM1 Hamiltonian for chlorine for the attack of chloride on methyl chloride. Parameterization was carried out by only relying on gas phase data for refinement of that Hamiltonian and then the parameterization was tested in solution as to its performance in capturing experimental after augmentation of all parameters was complete. The results of individual states used to obtained optimized SRP parameters and details, such as the weights of the chi-squared function used in parameter optimization can be found in Appendix B. The performance of the SRP Hamiltonian are examined in the following subsections, in both the gas phase and in solution. All reaction barrier and geometric data is compared to selected existing semiempirical Hamiltonians and literature experimental data where applicable.

2.3.1 Performance and Evaluation of Various QM *ab initio*, Density Functional and Semiempirical Methods

A variety of different mid-to-high level quantum mechanical methods were tested in order to determine a standard on which the new semiempirical Hamiltonian could be parametrized. The ideal test functional would achieve chemical accuracy for the reaction of interest while reducing the computational cost of parametrization as much as possible. Results comparing the complexation energy (the energy difference between the IDC and the free reagents) and the central attack barrier (the difference between the TS and the IDC) are shown in Table 2.2. The values shown compare each functional to the experimentally observed value, as well as the high-level W1' reference.

	Transition State		Ion Dipole Comp.		
Method	d_H	Barrier	d_N	d_L	Energy
Expt.	–	13.66[162, 163]	–	–	-10.53[164]
W1'	2.355[a]	(0.01)	1.846[a]	3.191[a]	(-0.05)
HF	2.381	(2.60)	1.818	3.350	(1.49)
MP2	2.287	(2.22)	1.806	3.151	(-0.30)
B3LYP	2.354	(-4.57)	1.844	3.180	(0.71)
BH&HLYP	2.562	(-0.71)	1.816	3.194	(0.59)
ωB97x-D	2.324	(-0.58)	1.817	3.186	(0.04)
M06-2X	2.300	(0.06)	1.817	3.101	(-1.16)
AM1	2.154	(-4.61)	1.785	2.872	(1.97)
MNDO	2.148	(-3.15)	1.830	3.346	(3.25)
MNDO/d	2.173	(5.93)	1.805	3.516	(4.45)
PM3	2.189	(-4.00)	1.806	2.843	(1.65)
MeCl SRP	2.300	(0.04)	1.819	3.088	(-1.12)

Table 2.2: **Gas phase quantum chemical calculations results, including energetics and geometries for the attack of chloride/methylchloride symmetric S_N2 reaction.** Gas phase stationary point calculations were performed at the 6-311+g(3df,2p) level of theory with a variety of quantum metods describing the attack of chloride anion on methyl chloride. Adiabatic energy results are presented as differences from experiment. Included with the methods tested in this study is the W1' functional as a high-level benchmark for readers. All energies and energy differences are reported in kcal/mol. Bond distances are also given for the TS and IDC. All distances are reported in Å. Distances marked by, *a*, had their geometeries calculated at the B3LYP/cc-pVTZ+(X) level of theory.

	Cl^-	Cl_2	MeCl		CCl_4
	EA	BL	C-Cl BL	Dipole	C-Cl BL
Expt.	83.31[165]	1.988[166]	1.776[167]	1.870[168]	1.769[169]
HF	55.35	1.974	1.781	2.119	1.760
MP2	82.89	1.985	1.773	2.080	1.762
B3LYP	84.90	2.010	1.795	1.950	1.780
BH&HLYP	79.32	1.983	1.777	1.991	1.761
ωB97x-D	84.88	1.986	1.781	1.943	1.768
M06-2X	84.15	1.984	1.779	1.931	1.764
AM1	37.65	1.918	1.741	1.513	1.760
PM3	51.20	2.035	1.764	1.377	1.747
MNDO	83.71	1.996	1.795	1.973	1.782
MNDO/d	83.40	1.984	1.779	1.859	1.786
MeCl SRP	83.65	1.988	1.771	1.817	1.795

Table 2.3: **Selected physical properties for a variety of related chlorine-containing compounds used in to evaluate different quantum chemical methods.** Each of the properties shown are used as part of the objective function for in the parameterization of the SRP Hamiltonian. EA is the electron affinity and BL is the bond length. All energies are reported in kcal/mol, dipole moments are reported in Debye and all distances are reported in Å.

Looking at the energetics of the chloride/methyl chloride reaction, the experimentally known central reaction barrier of 13.66 kcal/mol [162, 163] is best reproduced by the high-level benchmark W1' model chemistry, which underestimated the barrier by 0.01 kcal/mol. As previously stated, due to the computational cost of this method, it is not an ideal choice to use in the parametrization scheme despite its high accuracy. The next most accurate tested method was the M06-2X functional, which had similarly accurate performance of over estimating the barrier by 0.06 kcal/mol. The binding energy of the IDC is most accurately reproduced by the ωB97x-D functional and W1' method with deviations from the experimental value of 0.04 and 0.05 kcal/mol, respectively. All of the tested high-level quantum methods have similar calculated geometries for the various reagents and reaction complexes. A majority of the quantum methods estimated d_H to be approximately 2.3 Å, with the largest deviation from the set being BH&HLYP which estimated that bond length to be about 0.2 Å larger than the other functionals. Likewise, all high-level QM methods performed similarly when evaluating d_N and d_L, with the largest outlier being the HF calculation of d_L, indicating that the distance should be about 0.15 Å longer than the other functionals. The HF d_L distance discrepancy is possibly related to it significantly lower electron affinity for chlorine, as also seen in Table 2.3, thus requiring longer distances in the IDC because electronic transfer will happen more readily at closer distances than the other methods.

As for physical properties and geometries for related compounds; M06-2X reproduced the experimental electron affinity for a chlorine atom as well as the dipole moment of methyl chloride the best among the quantum methods that were surveyed. These results would indicate that the M06-2X most accurately describes the electronic nature of chlorine and chlorine containing molecules, as it is the only density functional which takes into account the second derivative of the electronic density, from the set tested. As a whole, geometries for the relevant small molecules were similarly well reproduced by all functionals with the exception of B3LYP which consistently overestimated carbon-to-chlorine bond lengths by approximately 0.05 Å.

When evaluating QM methods to act as the model function for the parameterization of the SRP semiempirical Hamiltonian, it is important to consider how accurate the the chosen method would perform in capturing overall reaction energies as well as how

computationally demanding the method is to perform the needed calculations. The best overall performance in capturing geometries and energies for the entire data set would fall upon ωB97x-D, being able to calculate both the central barrier and IDC energies with sub kcal/mol accuracies and being able to closely recapitulate known geometries for the related molecules. However, the authors felt that the central barrier of the reaction was the most important aspect of the potential energy surface to correctly calculate, especially due to reasons discussed earlier. When in solution, which is the intended medium for which this Hamiltonian is designed, the reagents of a reaction are greatly stabilized by their interactions with the surrounding water while the IDC (if one exists) is destabilized relative to the reagents, all of which is controlled by the molecules' charge distribution and the atoms' Lennard-Jones terms. In this regard, the complexation energy is less important to recapitulate in the gas phase as it becomes mostly insignificant when compared to the reaction barrier in solution. As such, the M06-2X functional was chosen. This functional is able to accurately capture the reaction barrier, the key feature the SRP Hamiltonian is targeted to produce, while being more tractable than the W1' functional.

After selection, constrained scans of the chloride attack on methyl chloride, as defined by the distance and angle between them, were performed at a mixed basis set level in order to map out the reduced adiabatic energy surface of the reaction. These scans can be seen in Figure 2.2.

2.3.2 Performance of the Specific Reaction Parameterization AM1 Hamiltonian

The newly parametrized SRP Hamiltonian is shown to reproduce a higher level QM gas phase profile. Several aspects of the reaction were taken into account including the geometry and energetics of the reaction along several different points, but also several important properties of relevant compounds. For general comparison of gas phase performance of different semiempirical Hamiltonians for this reaction, gas phase optimization of each the AM1, MNDO, MNDO/d, and PM3 Hamiltonians were performed with an identical protocol as used in the SRP optimization, with constraints on the angle of attack and the distance between the attaching chloride ion and the methyl chloride carbon. Results of this test can be seen in Figure 2.2. Similarly to the higher level

quantum data, differences from experimental reaction barriers and physical properties were collected and can be seen in Table 2.3.

To accurately capture the reaction barrier in solution, and in order for the results to be physically meaningful, the gas phase profile must be accurately reproduced. To this end a series of standard semiempirical Hamiltonians were tested in order to determine if a special parametrization was needed for the chloride/methyl chloride reaction and, of the Hamiltonians tested, none of them performed adequately in capturing the reaction barrier in the gas phase. All of the semiempirical Hamiltonians tested severely underestimated the reaction barrier besides the MNDO/d functional which over estimated the barrier by 5.93 kcal/mol. It should also be noted that all of the semiempirical Hamiltonians under estimate the energy for forming the IDC, however it is interesting that the slope of the energy surface for the approach of the chloride ion into the IDC is consistently underestimated for all of the semiempirical Hamiltonians, indicating that while in the gas phase each semiempirical Hamiltonian predicts that the chloride ion will interact with the methyl chloride at a much longer range than the higher-level M06-2X functional would indicate. This trend is most notable the MNDO-type Hamiltonians. When looking at the physical properties of the semiempirical Hamiltonians, the AM1 and PM3 Hamiltonians both underestimate the dipole moment of methyl chloride and correspondingly the carbon to chloride distance in the IDC and the TS are too short as compared to the M06-2X model functional.

Due to the lack of accuracy in existing semiempirical models for capturing the gas phase and solution phase barriers for the chloride/methylchloride, a new parametrization of the AM1 Hamiltonian was made, the final of which can be found in Table 2.4. As seen in Table 2.2 and in Figure 2.2, the specifically parametrized chlorine Hamiltonian can readily reproduce the stationary point values quite accurately (within less than one tenth of a kcal/mol), and general curvature, of the reaction profile. A comparison of the Intrinsic Reaction Coordinate (IRC) profiles between the newly parameterized Hamiltonian and the high-level M06-2X reference can be found in Figure 2.3. for the IRC calculation, the SRP Hamiltonian in general is slightly lower in energy than the M06-2X reference along the reaction pathway but then widens and reaches the same energy extrema, resulting in a marginally narrower IRC profile. The IRC itself is also more narrow than the gas phase reaction profile, in general, but the transition state

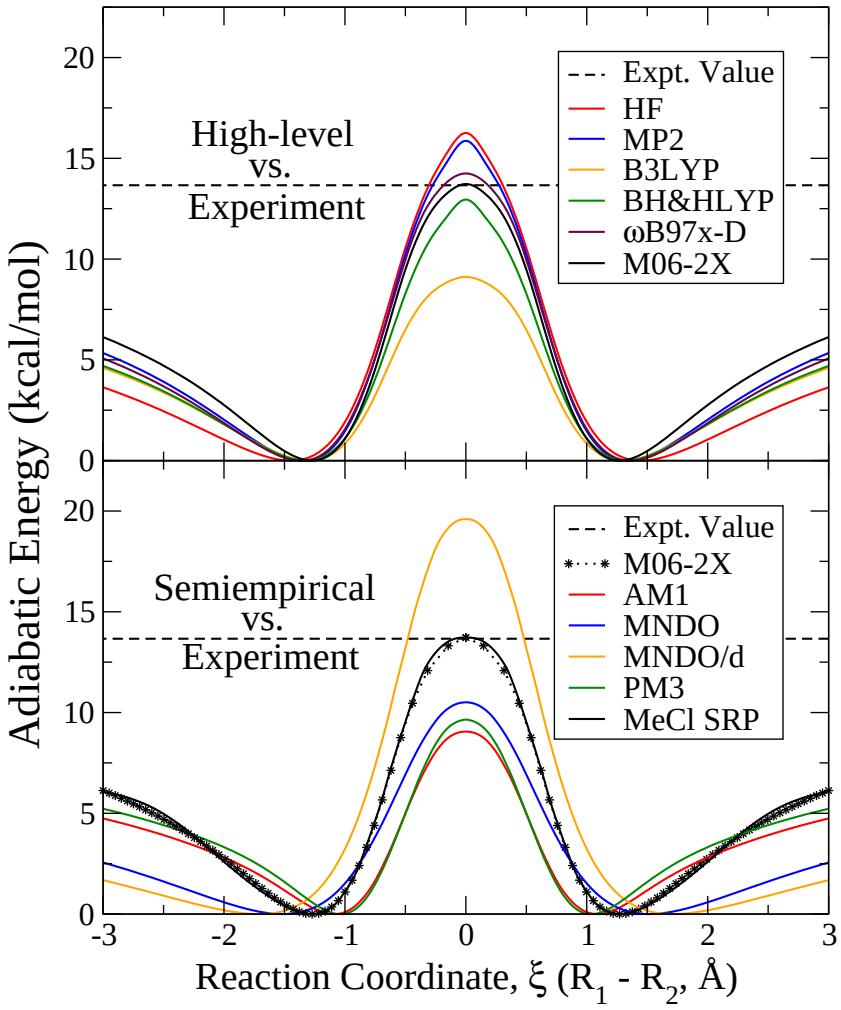


Figure 2.2: **Comparative gas phase reaction profiles for the attack of the chloride anion on methylchloride for both high-level and semiempirical quantum methods, including the new SRP Hamiltonian.** Results for both a range of different high-level quantum mechanical techniques (top) and with a variety of semiempirical Hamiltonians (bottom) as compared to the experimentally known barrier. From the high-level data, M06-2X reproduces the barrier most accurately. The newly parameterized semiempirical Hamiltonian (SRP) closely mimics the higher level data (shown as the dashed line) and captures the experimental barrier.

Parameter	Elmnt	Value
USS	Cl	-123.7250428683
UPP	Cl	-77.4465820238
BETAS	Cl	-22.0636103143
BETAP	Cl	-15.0606629218
HSP	Cl	2.3232148995
GSS	Cl	25.4542629717
GPP	Cl	10.9136964876
GSP	Cl	13.2195152507
GP2	Cl	9.9406477970
ZS	Cl	3.6491091946
ZP	Cl	2.0509924622
ALP	Cl	2.8891075086
GNN	Cl	1.0000000000
FN11	Cl	0.0896900889
FN21	Cl	3.9710619231
FN31	Cl	1.3583210593
FN12	Cl	0.0494738744
FN22	Cl	3.9499284925
FN32	Cl	2.1151236026
FN13	Cl	0.0058588832
FN23	Cl	2.3173905973
FN24	Cl	1.7521185936
FN34	Cl	5.1796862803
FN33	Cl	5.1189314112
FN14	Cl	-0.0121500094

Table 2.4: **Final optimized parameters for chlorine in the new SRP Hamiltonian.** Values of the optimized Hamiltonian for the chloride/methylchloride reaction, for use with an AM1 semiempirical Hamiltonian for use in QM/MM simulations.

and energy basins are well reproduced. Additionally, normal mode analysis of the transition state were calculatedwas performed for both the M06-2X reference as the SRP Hamiltonian for comparision. Results can be found in Table 2.5. Each report one imaginary mode (negative eigenvalue) and the frequencies reported are comparable in value. Finally, as noted in 2.3, several of the higher-level properties such as carbon to chlorine bond lengths and the methyl chloride dipole moment are reproduced.

mode	M06-2X	SRP
1	-454.50	-484.65
2	213.96	202.18
3	214.30	202.21
4	229.96	219.83
5	984.48	964.33
6	985.23	964.38
7	1093.59	1167.04
8	1410.66	1327.53
9	1412.45	1327.55
10	3205.93	3171.69
11	3400.75	3179.53
12	3403.36	3179.64

Table 2.5: **Vibrational frequencies for the transition state of chloride/methylchloride symmetric reaction as obtained via normal mode analysis for both the reference M06-2X and SRP AM1 methods.** Frequency values for each mode from the normal mode analysis of the TS of reaction of methylchloride with a chloride anion for both the M06-2X reference, at the 6-311++G(3df,2p) level, and the newly parameterized SRP Hamiltonian are shown. The results from each are similar, indicating that the energy landscape between the two models are comparable.

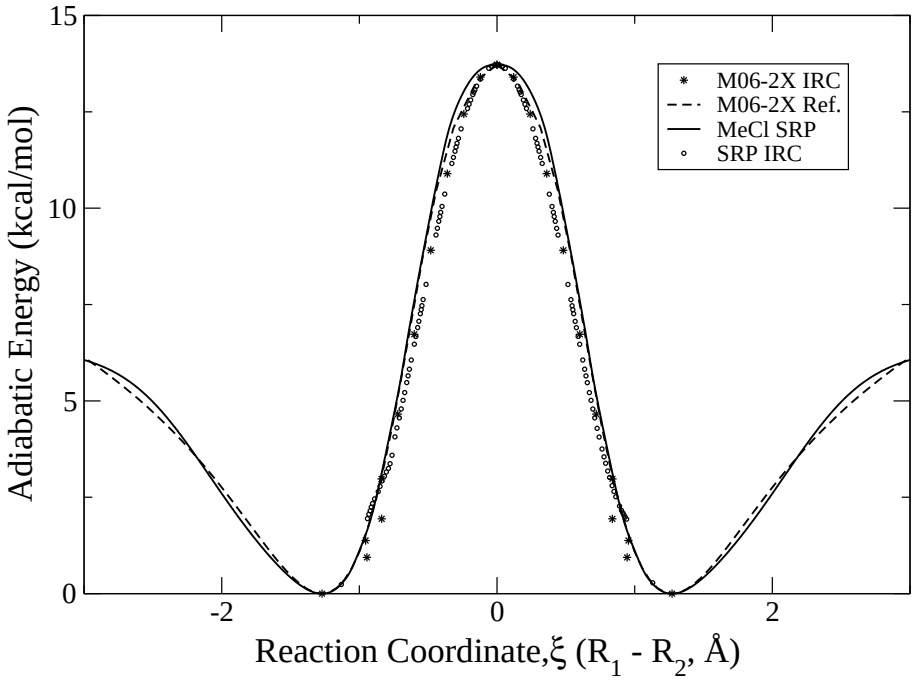


Figure 2.3: **The Intrisic Reaction Coordinate canculation and gas phase profiles of the attack of the chloride anion on methylchloride for the new SRP model and a higher-level M06-2X reference.** Comparison between M06-2X/6-311++(3df,2p) and the semiempirical AM1 SRP Hamiltonian. Differences between the models, for both the reaction profile and the IRC calculation are minimal and reach the same energy basins.

In order to capture the reaction profile in solution, PMF simulations were performed in solution for each semiempirical Hamiltonian tested in the gas phase. Results of these simulations can be seen in Figure 2.4. The profiles are compared to the experimental reaction barrier. The parametrized semiempirical Hamiltonian performed the best out of the selected Hamiltonians for both the TIP3P and TIP4P-Ew water simulations. Errors on reported values are on the order of ±0.1 kcal/mol in the free energy barriers.

	TIP3P		TIP4P-Ew	
Hamiltonian	$<d_H>$	$\Delta\ G(\xi)^{\ddagger}$	$<d_H>$	$\Delta\ G(\xi)^{\ddagger}$
Expt. Value	–	26.5	–	26.5
AM1	2.170	(-2.6)	2.169	(-2.8)
MNDO	2.170	(-0.9)	2.170	(-1.0)
MNDO/d	2.182	(5.9)	2.181	(5.9)
PM3	2.198	(-3.2)	2.199	(-3.6)
MeCl SRP	2.345	(0.0)	2.346	(0.1)

Table 2.6: **Solution phase QM/MM simulation data for the chloride/methylchloride reaction for various semiempirical methods in both TIP3P and TIP4P-Ew water.** The carbon to chloride distance (d_H) and the attack barrier ($\Delta\ G(\xi)^{\ddagger}$) for each are shown. Reaction barriers are reported as differences from experimentally[170] known values. Distances are reported in Å and energy barriers are reported in kcal/mol.

The SRP Hamiltonian best captures the experimentally observed solution phase reaction barrier of 26.5 kcal/mol[170] from the set of semiempirical Hamiltonians tested, and did so within statistical error of the simulation. Numerical data from the PMF simulations can be found in Table 2.6. The standard Hamiltonian results follow much in the same trend as was seen in the gas phase adiabatic profile calculations, however while in the gas phase PM3 more accurately reproduces the experimental barrier than AM1 while in solution phase the opposite was the case. Given that the 'size' of the particles involved in the reaction are set using static Lennard-Jones parameters, the differences between the solvation effect of each Hamiltonian can mostly be attributed to their

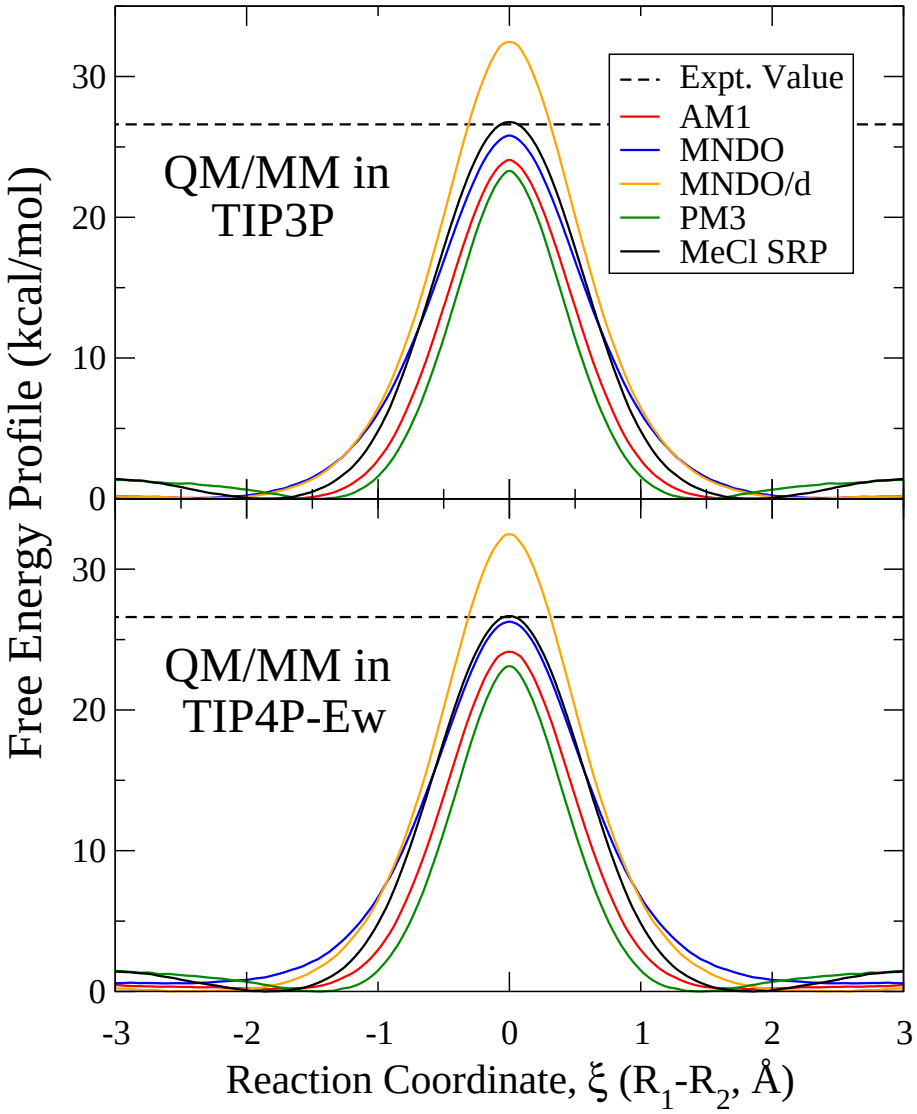


Figure 2.4: **QM/MM free energy profiles for the symmetric reaction of the attack of chloride anion on methylchloride of several semiempirical Hamiltonians in both TIP3P and TIP4P-Ew.** TIP3P (top) and TIP4P-Ew (bottom) water are presented. Simulations were performed in AMBER and the free energy profiles were calculated using umbrella sampling in conjunction with MBAR utilizing a kernel density estimator.

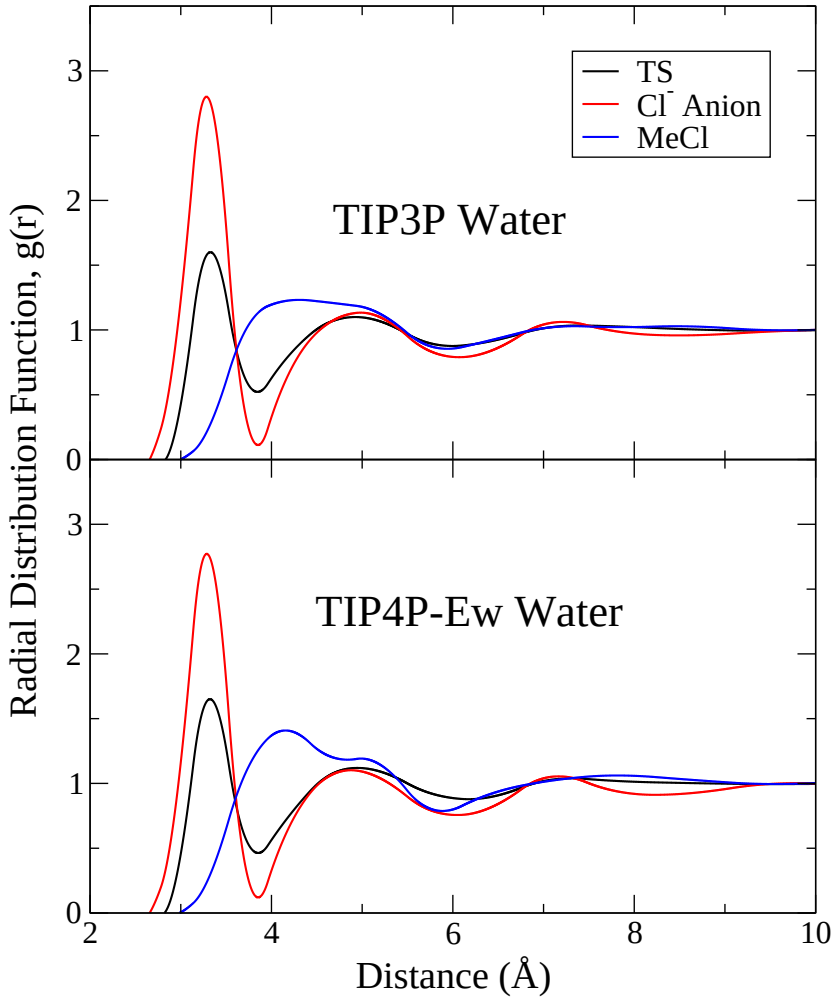


Figure 2.5: **The radial distribution functions of chlorine atoms in the reactant and transition states of the chloride/methylchloride reaction QM/MM simulations in both TIP3P and TIP4P-Ew water.** Shown are the RDF of the chlorine/chloride atoms involved in the S_N2 reaction between chloride and methyl chloride and the oxygen of the surrounding waters. RDF plots are generated from retained simulations in both TIP3P (top) and TIP4P-Ew (bottom) water, comparing the solvent structure at the transition state to that of the separate reagents. Functions are generated by integrating the direct count of the species of interest and then smoothed using binned splines. Numerical results from these simulations can be found in Table 2.7.

electronic nature. Therefore, it is believed that because the PM3 Hamiltonian drastically underestimated the dipole moment of methyl chloride, it would in turn destabilize the methyl chloride's solvation which would lead to more energetically unfavorable reactants and a shorter central barrier height. Similarly, the MNDO semiempirical Hamiltonian performed admirably in solution despite drastically underestimating the barrier in the gas phase, this change could be due to the over predicted dipole moment of methyl chloride. As the reagents of the reactant would be more stabilized due to stronger interactions with the surrounding water, the reaction would appear to have an enhanced central barrier.

Bond lengths in solution for the TS, d_H, mimic the same trend as those seen in the gas phase, with the SRP Hamiltonian having a much larger bond length than the other standard Hamiltonians. When moved into solution, all of the semiempirical Hamiltonians experienced a minor elongation of d_H. This behavior is expected because of increased interactions with the surrounding water would favor geometric relaxation of the d_H bond, insomuch as the competing solvent destabilization energy would allow the complex to expand. None of the semiempirical Hamiltonians showed any strong solution-type dependence in either d_H or the barrier height. Working in an NDDO basis frame work, with non-polarizable chlorine atoms surrounded by non-polarizable waters, relegates the direct interaction between the Hamiltonian and the solvent to be largely electrostatic without any type of electronic response. Given that the Lennard-Jones parameters used were specifically designed to account for changes in the solvent it is not surprising that there is no strong dependence for changes in the solvent.

	TIP3P		TIP4P-Ew	
	TS	Cl^-	TS	Cl^-
r_{max}	3.28±0.03	3.18±0.02	3.26±0.02	3.20±0.02
r_{min}	3.73±0.02	3.84±0.03	3.75±0.03	3.73±0.01
N_{crd}	3.95±0.08	6.93±0.01	4.05±0.18	6.52±0.01
r^*_{max}	–	3.05-3.25±0.2	–	3.05-3.25±0.2
N^*_{crd}	–	6.0-6.5±0.5	–	6.0-6.5±0.5

Table 2.7: **Data of the first solvation shell around the chloride anion and the TS chlorine atom in the chloride/methylchloride reaction as modeled by the SRP Hamiltonian in TIP3P and TIP4P-Ew.** r_{max} and r_{min} are the positions of the maximum and the minimum of the RDF between chloride and the oxygen of water, respectively. N_{crd} is the calculated coordinated number of waters in the first solvation shell (between the beginning of the RDF and and r_{min}). r^*_{max} and N^*_{crd} are the experimentally known values, taken from the work of Zhang *et al.*[171]

The radial distribution function (RDF), using the SRP Hamiltonian are, for the chlorine to oxygen of water (OW) are also included for both the TIP3P and TIP4P-Ew simulations. Differences between the two different water models are nominal, however the differences in the number of coordinated water between the chloride anion, around the each chlorine in the TS and methyl chloride are drastic (as seen in Table 2.7). The calculated coordination number for the chloride anion is 6.93±0.01 and 6.52±0.01 for TIP3P and TIP4P-Ew respectively, which compared favorably to the experimental number range of 6.0 to 6.5±0.5,[171] and appears to have fairly well organized second and possibly third solvation shells. As the system moves through the transition state the coordination number for the chlorine drops to 3.95±0.08 and 4.05±0.18, as the local charge of the chlorine changes from negative 1.0 to approximately negative 0.7 charge units and the chloride attacks the methyl chloride. The results for methyl chloride are not too meaningful, despite the fact that one would expect the interactions of the chlorine atom to be lessened as the local charge of that atom decreases, because the methyl group attached to the chlorine atom disrupts the RDF in non-trivial ways. As such, the first and second solvation shells around the chlorine atom appear to overlap.

2.3.3 Critical assessment of the Specific Reaction Parameterization AM1 Hamiltonian

The new AM1 SRP Hamiltonian performs quite admirably in capturing experimentally known observables in both the gas and solution phases. Nonetheless, one must be critical in evaluating the limitations of the procedure described herein, and acknowledge that the very close agreement with experiment might arise due to a somewhat fortuitous

cancellation of errors. Notably, the solvent models used in this study are not explicitly polarizable, and the van der Waals radii used for the chloride ion were derived to accurately reproduce the solvation free energy of a non-polarizable MM chloride ion for use with each specific water model used.[161] However, due to the minimal basis set used in the QM model employed here, the QM chloride ion when in isolation is also not polarizable due to the lack of available virtual orbitals.

While the model performed well within the current framework, one should expect that as the reaction proceeds through the reaction coordinate that the electronic density around the chlorine atoms to shift in accordance to their bonding environment. These changes should be reflected in the repulsive exchange and correlation dispersion potentials, requiring the van der Waals radii to adjust. [172, 173] The static Lennard-Jones parameters used in this study were developed to obtain solvation free energy of a chloride ion in non-polarizable water, as is the case for the QM/MM system, and do not consider the solvation free energy of methyl chloride. It is therefore likely to be in error. However in the case of this study, the explicit coupling of polarization and exchange effects might not be necessary in order to obtain an accurate reaction barrier in solution. Thus, recognizing these limitations of the model used in the present work, one must concede that the very close agreement with the experimental barrier in solution is likely a result of the fortuitous cancellation of errors.

2.4 Conclusions

Accurately predicting reaction barriers in the condensed phase is of pinnacle importance in the fields of computational chemistry and biology. In order to be able to achieve this feat, precise models must be made in a reliable and robust fashion. In this work, such an approach has been applied for the development of SRP quantum model for chloride ion attack to methyl chloride based on gas-phase reaction profiles that are transferable to QM/MM simulations in solution. The new model reproduces many important observables, while being orders of magnitudes faster than conventional high-level QM methods. In the gas phase, reference experimental geometries and dipole moments are recapitulated, as well as the experimental reaction barrier. QM/MM simulations in solution demonstrate that the new model provides excellent agreement

with the experimentally observed free energy barrier estimated from kinetic experiments. The overall free profiles show little dependence on the form of the water model (TIP3P versus TIP4P-Ew), although there is some differences in the predicted solvent structure along the reaction coordinate. While these results are quite encouraging, one must also be forthcoming with respect to the limitations of the model, and in particular, the use of van der Waals radii that are fixed along the reaction coordinate. Overall, the results of this work provide an accurate SRP model for an important prototype S_N2 reaction in chemistry that can be used in the gas phase to build potential energy surfaces or in QM/MM simulations in solution to build free energy profiles. Analysis of the simulation results provide a detailed characterization of the solvent structure along the reaction coordinate, and insight into the nature of solvation effects on the reaction.

2.5 Acknowledgments

The authors are grateful for financial support provided by the National Institutes of Health (GM62248 to DY). Computational resources from The Minnesota Supercomputing Institute for Advanced Computational Research (MSI) and from Rutgers, the State University of New Jersey were utilized in this work. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575.

Chapter 3

Charge-dependent Many-body Exchange and Dispersion Interactions for use in Hybrid QM/MM Simulations

With the specific reaction parameterization of the AM1 semiempirical Hamiltonian outlined in Chapter 2, which provides an accurate quantum model for both the gas phase and solution reaction of the attack of a chloride anion on methylchloride, the parameterization of QXD, a charge-dependent exchange-repulsion and dispersion model for QM/MM simulations, could be completed. This chapter will detail the derivation, parameterization and integration of the QXD model into an existing molecular dynamics software package, AMBER[158, 174]. In order to parameterize this model a specialized technique called thermodynamic integration (TI) will be used extensively. For a general background on this technique, please refer to Section 1.3.2.

[2] This chapter compiles material with permission sumbitted for publication in The Journal of chemical physics from the manuscript Kuechler, Erich R., Timothy J. Giese and Darrin M. York. "Charge-dependent many-body exhchange and dispersion interactions in combined QM/MM simulations" The Journal of chemical physics **submitted, revised**. Copyright 2015, AIP Publishing LLC.

3.1 General Background

Combined quantum mechanical/molecular mechanical (QM/MM) simulations are powerful tools in the study of chemical reactions taking place while embedded within large biological scaffolds and/or in condensed phase environments. These hybrid methods are typically orders of magnitudes more efficient than fully quantum mechanical approaches, although recent advances in the development of reactive quantum mechanical force fields have greatly narrowed this gap. [175, 176, 177, 178] QM/MM methods have been successfully applied in the study of enzymes,[179, 180, 181] catalytic RNAs,[182, 183, 184] ligand binding,[185, 186, 187] acid dissociation constants[188, 189, 49] and small molecule reactions occurring in solution.[190, 191, 192, 193]

Combined QM/MM approaches leverage the strengths associated with of both QM methods and MM simulations by allowing a typically small, localized reactive region requiring explicit treatment of electronic degrees of freedom to be modeled within a typically very large, complex condensed phase environment. The overall reliability of QM/MM simulations depends critically on both the accuracy of the chosen QM and MM models as well as the treatment of the QM/MM interaction. Considerations such as the size of the reactive region and the specific chemistry occurring within it must be considered when choosing the QM method. The accuracy of MM force field model depends simultaneously on its functional form and the empirical parameters used to describe the underlying potential. The MM model is typically chosen as to best represent key features of the environment surrounding the QM region. The most advanced force field treatments include accurate descriptions of electrostatic interactions and have explicit consideration of many-body polarization. While such potentials are very promising, these force fields are typically less computationally efficient than their traditional counterparts and, in many cases, have parameters which either have a somewhat limited chemical scope or have not been as fully matured through critical assessment to the level of those employed by the more simple force fields. Because of these reasons, the vast majority of QM/MM simulations currently utilize conventional, point-charge force fields in the study of condensed phase reactions.

The QM and MM regions interact with each other through a QM/MM boundary. Most generally, both bonded and non-bonded interactions are treated over this division.

Bonded interactions are required when the QM/MM boundary falls between two or more covalently bound atoms. A variety of specialized methods; including link atoms,[44, 194] pseudoatoms,[43, 195] and generalized orbital methods;[196, 197] have been developed to model this interaction. Non-bonded interactions can be further subdivided into electrostatic and non-electrostatic terms.

The QM/MM non-bonded electrostatic interactions can be treated using different embedding techniques that can be broadly separated into three categories: mechanical, electrostatic or polarized embedding.[26] Electrostatic embedding enters the electrostatic potential due to fixed, static MM point charges directly into the QM Hamiltonian as an external potential to polarize the QM density that results from a self-consistent field (SCF) procedure. This scheme generally provides greater accuracy than mechanical embedding,[198] where static point charges are assigned to QM atoms and electrostatics are treated classically as Coulomb interactions between point charges. If the MM force field is itself an explicitly polarizable model then a polarized embedding[199, 200] scheme can be adopted, where MM polarizable charges/multipoles are allowed to respond to the QM charge density in some way.

The QM/MM non-electrostatic non-bonded interactions include short-ranged exchange repulsion and mid/long-ranged dispersion interactions. Conventional QM/MM methods handle this term using the Lennard-Jones (LJ) model, [201, 202, 203] which is also the most commonly employed non-electrostatic non-bonded model used by MM force fields. However, unlike the electrostatic QM/MM interactions in electrostatic and polarized embedding methods, the LJ model for non-electrostatic non-bonded QM/MM interactions is not explicitly coupled to the quantum mechanical wave function, meaning these interactions cannot adjust to changes in the chemical environment. In order to account for this issue, atoms in different chemical environments are assigned different LJ interactions, and much like many MM force fields adopt the concept of an *atom type* - a prescription in which parameters for an atom are inherited based on specific local electronic and chemical bonding environment. Within this framework, each atom type is given a unique set of fixed LJ parameters.[106] Given that traditional MM force fields do not allow for the formation or destruction of chemical bonds this strategy provides a practical mechanism for modeling these interactions, provided that the scope of LJ parameters in the force field is wide enough to model a diverse range of intermolecular

interactions. However, in the case of QM/MM simulations, the bonding environment can and in fact often does change. As such, adhering to the notion of assigning fixed atom types to reactive QM atoms is impractical and unsound. In careful QM/MM studies, LJ parameters for QM atoms are developed to produce an overall reasonable, balanced reaction profile that strives to eliminate artifacts due to having atoms whose assigned atom types might misrepresent changes in the chemical identity of those atoms over the course of the modeled chemical reaction. However, in doing so the accuracy of the solvation properties of individual states along the reaction pathway is sacrificed. Frequently, it is not possible to select a single LJ parameterization which can simultaneously reproduce quantitative reaction barrier data while obtaining accurate solvation free energies for the reactant and product states.

One partial solution to this issue is to use so-called adaptive QM/MM methods,[204, 205, 206] where the QM and MM representations of molecules turn on/off as they enter/leave a larger "reactive zone". In principle QM models account for changes in exchange repulsion and dispersion interactions as a function of the electronic structure. By using these methods and by pushing the QM/MM boundary further away from the reaction zone, the problems associated with static LJ potentials could be greatly reduced. While this approach is conceptually appealing, and has been demonstrated to be successful if carefully applied, it is also more computationally demanding than traditional methods and will only have clear benefits if the QM model used is superior to the MM force field at representing solvation properties. For instance, the application of approximate semiempirical, Hartree-Fock or small basis set density-functional QM models have been shown to often be inferior to a carefully parameterized MM water model for describing the physical properties of liquid water.

Alternatively, one could replace the conventional, static LJ QM/MM interaction with one which is explicitly coupled to the QM electronic structure. Recent studies have pioneered this type of approach by directly linking the extent of charge delocalization to the electronic structure of the system, effectively allowing the charge of more negative species to appear more diffuse.[207, 208] This work introduces a new model that explicitly couples atomic charge with many-body exchange and dispersion interactions. The QXD (Charge-dependent eXchange and Dispersion) model allows the

seamless modulation of non-electrostatic, non-bonded interactions in a chemically intuitive and meaningful way with changes in local atomic charge. As atoms become more negatively charged, they become larger and softer with regard to their interactions with surrounding MM atoms. In this way, the model is able to better represent the QM/MM interaction non-locally along the reaction coordinate.

In the present work, we examine a well-studied benchmark reaction of the chloride ion attack on methylchloride using a specific reaction parameterization semiempirical QM model[209] in a solution of non-polarizable MM water while treating electrostatic QM/MM interactions via electrostatic embedding. This combination of QM and MM models was made based which methods where most widely used in combined QM/MM studies of biological reactions; a target application area for which the hope is that the methods introduced here will have future impact. The remainder of the chapter will be broken into four main parts: 1) a background theory section in which the QXD model is developed, 2) a computational methods section which outlines details as to how the QXD model was parameterized and integrated into a QM/MM framework, and 3) a results and discussion section which presents, compares and discusses simulation results for solvation free energy calculations and reaction free energy profiles involving chloride ion attack to methylchloride, and 4) a section which summarizes the main conclusions of the work and places its importance into a broader context.

3.2 Computational Methods

This section describes computational and theoretical details involved in the implementation, parameterization, and validation of the QXD model for interaction energies, molecular dynamics simulations, and free energy simulations. First, a brief outline of the framework used in traditional QM/MM simulations is provided, and how the QXD model will be interaction will be used to replace the LJ potential in the QM/MM interaction energy term. Then a series of analytic expressions will be provided which allow the QXD potential to interact with traditional force fields, providing a mechanism for the integration of the QXD model into nearly any existing QM/MM program. Third, protocols and calculation details for comparing LJ and QXD QM/MM interactions to

high-level quantum data using gas-phase scans are provided, followed by the molecular dynamics (MD) simulation protocols to calculate the solvation free energies used in the parameterization of the QXD model are discussed. Finally, the MD protocol for free energy profile simulations used to validate the QXD QM/MM interaction model is described.

The QXD model replaces the traditional LJ treatment of non-classical interactions in the QM/MM Hamiltonian for QM atoms. It allows QM atoms to effectively 'change size' in accordance to fluctuations in local charge, as determined by the single-particle density matrix. QXD does this task through a series of equations which closely follow previous work of Giese *et al.*[172] However, unlike the initial work where the energy was added as a post-SCF correction term, present here is a variational formulation that is integrated with the SCF procedure which provides simple expressions for the gradients needed to obtain forces when used in molecular dynamics calculations.

Solvation free energies of chlorine containing compounds and a prototype reaction of a chloride anion attacking methylchloride will be examined to validate the QXD QM/MM interaction model. This reaction has been extensively studied[115, 116, 117, 210, 211] and is generally well understood, lending credence as a system for benchmarking new computational models. For all QM/MM calculations reported here, the Specific Reaction Parameterization (SRP) AM1 Hamiltonian for Cl^- attack on CH_3Cl discussed in Chapter 2 was employed in both the gas phase and in solution.[209] Additionally, the LJ parameters used for all non-QXD atoms in QM/MM simulations and calculations can be found in Table 3.1.

3.2.1 The QXD QM/MM Interaction Potential

During traditional QM/MM simulations the total potential energy, E, can be written as

$$E = E_{\mathrm{QM}}(\mathbf{R}, \mathbf{P}) + E_{\mathrm{MM}}(\mathbf{R}, \mathbf{q}) + E_{\mathrm{QM/MM}}(\mathbf{R}, \mathbf{P}, \mathbf{q}) \tag{3.1}$$

where $\mathbf{R}$ is an array of atomic positions, $\mathbf{q}$ is a set of MM charges, and $\mathbf{P}$ is an atomic orbital (AO) representation of the QM total density matrix. This term is defined as

$$\mathbf{P} = \mathbf{P}^{\alpha} + \mathbf{P}^{\beta} \tag{3.2}$$

	Element	$R_{\min,c}$	ε_c
Solvent	O	1.775931	0.16275
	H	0.0000	0.0000
Solute	C	1.9080	0.0860
	H	1.4870	0.0157
	LJ_{Cl^-}: Cl	2.7600	0.0116615
	LJ_{CH_3Cl}: Cl	1.4489	0.4578662

Table 3.1: **Lennard-Jones parameters used for non-QXD atoms in molecular dynamics simulations** Parameters used to determine the non-electrostatic, non-bonded interactions in conventional QM/MM and other molecular dynamics simulations. The LJ_{Cl^-} chlorine parameters are taken from the work of Joung *et al.*[161] and are specufically tailored to be used with a TIP4P-Ew solvent, while the LJ_{CH_3Cl} were independently parameterized to capture the solvation free energy of methylchloride. Other parameters are from the ff99 AMBER force field.[212] $R_{\min,c}$ and ε_c, values are in Å and kcal/mol respectively.

where $\mathbf{P}^\sigma$ is the spin-resolved density matrix

$$P^\sigma_{\mu\nu} = \sum_{\sigma,i} n^\sigma_i C^\sigma_{\mu i} C^\sigma_{\nu i} \tag{3.3}$$

and $\mathbf{n}^\sigma$ and $\mathbf{C}^\sigma$ are the spin-resolved orbital occupation numbers and molecular orbital (MO) coefficients, respectively, and where μ and ν index atomic orbital basis functions. $E_{\mathrm{QM}}(\mathbf{R}, \mathbf{P})$ is the QM region's *ab initio* self-energy, $E_{\mathrm{MM}}(\mathbf{R}, \mathbf{q})$ is the MM energy, and $E_{\mathrm{QM/MM}}(\mathbf{R}, \mathbf{P}, \mathbf{q})$ is the interaction energy of the QM region with the MM region.

For a given set of atomic positions and MM charges, the total energy is minimized with respect to changes in the QM region's MO coefficients under the constraint that the MOs remain orthonormal. When the energy is extremized under these constraints, the following stationary condition (the Roothaan-Hall equation) is satisfied:

$$\mathbf{F}^\sigma \cdot \mathbf{C}^\sigma = \mathbf{S} \cdot \mathbf{C}^\sigma \cdot \mathbf{E}^\sigma \tag{3.4}$$

where $S_{\mu\nu} = \int \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) d^3r$ is the AO overlap matrix, and $F^\sigma_{\mu\nu}$ is the spin-resolved

Fock (or Kohn-Sham) matrix

$$\begin{aligned} F^{\sigma}_{\mu\nu} =& \left.\frac{\partial E_{\mathrm{QM}}}{\partial P^{\sigma}_{\mu\nu}}\right|_{\mathbf{R},\mathbf{q}} + \left.\frac{\partial E_{\mathrm{QM/MM}}}{\partial P^{\sigma}_{\mu\nu}}\right|_{\mathbf{R},\mathbf{q}} \\ =& F^{\sigma,0}_{\mu\nu} + \left.\frac{\partial E_{\mathrm{QM/MM}}}{\partial P^{\sigma}_{\mu\nu}}\right|_{\mathbf{R},\mathbf{q}} \end{aligned} \tag{3.5}$$

$F^{\sigma,0}_{\mu\nu}$ is QM method's Fock matrix in the absence of the MM atoms, and $\partial E_{\mathrm{QM/MM}}/\partial P^{\sigma}_{\mu\nu}|_{\mathbf{R},\mathbf{q}}$ is an effective external electronic chemical potential that causes the QM density to respond to the MM environment. The self-consistent field procedure is as follows:

1. Given a set of trial MOs, compute the density matrix.
2. From the density matrix, compute the energy and Fock matrix.
3. If the change in the energy is small, and Equation 3.4 is satisfied, then exit.
4. Construct a new set of trial MOs and go to step 1. Typically, a new guess is made by solving Equation 3.4 for $\mathbf{C}^{\sigma}$.

In conventional QM/MM implementations, $E_{\mathrm{QM/MM}}$ is expressed in terms of non-bonded electrostatic interactions and LJ potentials:

$$\begin{aligned} E_{\mathrm{QM/MM}}(\mathbf{R},\mathbf{P},\mathbf{q}) =& E^{\mathrm{el}}_{\mathrm{QM/MM}}(\mathbf{R},\mathbf{P},\mathbf{q}) \\ &+ E^{\mathrm{bond}}_{\mathrm{QM/MM}}(\mathbf{R},\mathbf{P}) \\ &+ E^{\mathrm{nb}}_{\mathrm{QM/MM}}(\mathbf{R}) \end{aligned} \tag{3.6}$$

where $E^{\mathrm{el}}_{\mathrm{QM/MM}}$ is the electrostatic interaction, $E^{\mathrm{bond}}_{\mathrm{QM/MM}}$ is the bonded interaction and $E^{\mathrm{nb}}_{\mathrm{QM/MM}}$ is the non-bonded interaction term. As discussed in the Background Information section, $E^{\mathrm{bond}}_{\mathrm{QM/MM}}$ can be entirely avoided if the QM/MM boundary does not cross any covalent bonds. The present work will be examining systems which are entirely subsumed by the QM region and no bonds intersect the boundary, so this term will be ignored to avoid unnecessary complication. Furthermore, if the LJ interaction is

assumed to represent the non-classical QM/MM interactions, the total QM/MM interaction can be expressed as

$$\begin{aligned} E_{\mathrm{QM/MM}}(\mathbf{R},\mathbf{P},\mathbf{q}) = &- \sum_{\substack{\mu\nu \\ c\in\mathrm{MM}}} P_{\mu\nu} q_c (\mu\nu|c) \\ &+ \sum_{\substack{a\in\mathrm{QM} \\ c\in\mathrm{MM}}} E_{\mathrm{LJ},ac}(R_{ac}) \end{aligned} \tag{3.7}$$

where

$$(\mu\nu|c) \equiv \int \frac{\chi_\mu(\mathbf{r})\chi_\nu(\mathbf{r})}{|\mathbf{r}-\mathbf{R}_c|} d^3r \tag{3.8}$$

and $E_{\mathrm{LJ},ac}$ is the LJ potential[201, 213, 214]

$$E_{\mathrm{LJ},ac}(R_{ac}) = \varepsilon_{ac}\left[\left(\frac{R_{\mathrm{min},ac}}{R_{ac}}\right)^{12} - 2\left(\frac{R_{\mathrm{min},ac}}{R_{ac}}\right)^{6}\right] \tag{3.9}$$

The LJ expression can be separated into two distinct, but coupled parts. The first representing the exchange-repulsion and the latter models the attractive, dispersion interactions. The dispersion term has quantitatively accurate behavior as the R_{ac}^{-6} form is the leading term of the multipole expansion that arises from the correlation interaction of electrons on two different centers. However the R_{ac}^{-12} repulsive term is simply a matter of computational ease; being efficient to calculate after the dispersion interaction has been determined. A more correct representation of the repulsive behavior of atoms would be modeled through a potential with exponential character,[215, 216] which has been adopted by more sophisticated models.[217]

Furthermore, it is well known that polarization[218, 219], exchange-repulsion and correlation-dispersion interactions[220, 216] are fundamentally quantum mechanical in nature. Modeling these interactions through a static LJ potential completely negates any meaningful response the exchange-repulsion and dispersion could have to these interactions. In fact, if one were to try to capture these subtle changes within a LJ-like formalism, fundamental physical trends would be violated. For instance, as electrons are added onto an atomic center, the magnitude of the correlation-dispersion should increase and the atom should become softer. Atomic hardness is well known to be related to the steepness of the steepness of the exchange-repulsion between atoms,[221],

and as atoms become softer this repulsive potential should become less steep. However, given the formulation of the LJ potential, it is impossible to capture both of these phenomena simultaneously. As correlation-dispersion increases, as seen by an increase in the ε_{ac} LJ parameter, the repulsive wall of the LJ interaction becomes more steep – indicating that the atoms being modeled are in fact becoming harder. Examples of this can be seen in Figure 3.1.

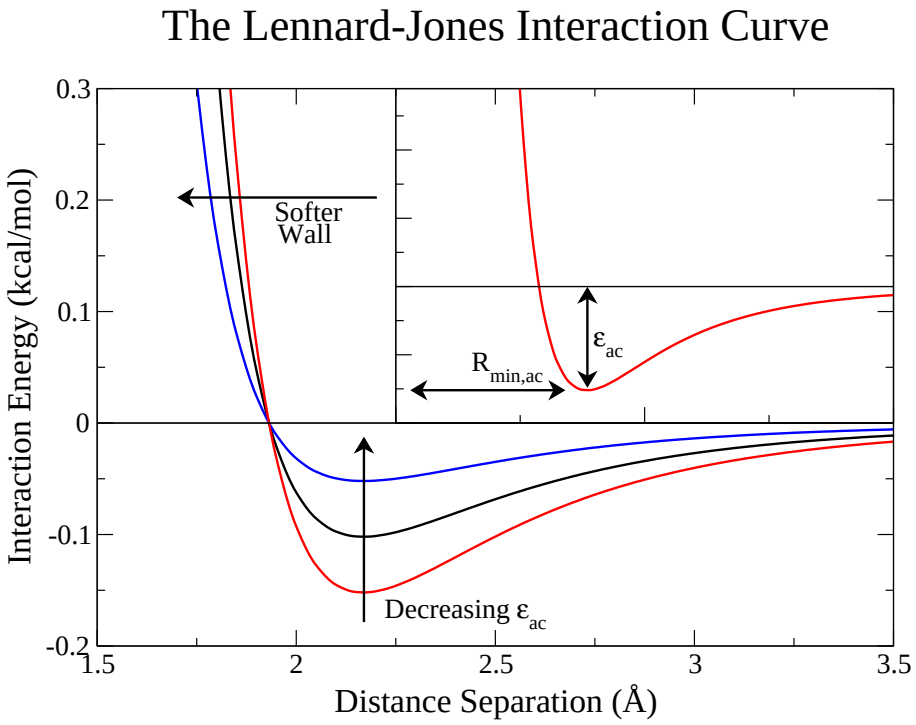


Figure 3.1: **The effect of changing ε_{ac} on the Lennard Jones interaction potential.** The LJ potential is used in modern chemical simulations to approximate exchange-repulsion and dispersion effects in MM force fields and in QM/MM simulations. It is controlled through two different pair-wise parameters $R_{\mathrm{min},ac}$ and ε_{ac} (seen in the upper right) which in themselves are controlled through atom types. The lower three curves show the effect of decreasing the magnetude of the ε_{ac} parameter on the LJ potential at a constant value of $R_{\mathrm{min},ac}$.

To address these concerns, this chapter introduces the QXD QM/MM interaction model. This model replaces QM/MM LJ interactions with a charge-dependent new functional form. The following will detail the equations and implementation of this model.

The systems considered in this chapter do not contain any instances where the QM/MM boundary crosses a covalent bond, so any specialized treatment of this type of interaction will be ignored. Any formalisms presented here can be safely extended to contain these interactions, however they were avoided in this case to mitigate any unnecessary complications. Continuing with these considerations in mind, we may write the QXD total energy as:

$$E = E_{\mathrm{QM}}(\mathbf{R}, \mathbf{P}) + E_{\mathrm{MM}}(\mathbf{R}, \mathbf{q}) + E_{\mathrm{QXD}}(\mathbf{R}, \mathbf{P}, \mathbf{q}) \tag{3.10}$$

where

$$\begin{aligned} E_{\mathrm{QXD}}(\mathbf{R}, \mathbf{P}, \mathbf{q}) = &- \sum_{\substack{\mu\nu \\ c\in\mathrm{MM}}} P_{\mu\nu} q_c (\mu\nu|c) \\ &+ \sum_{\substack{a\in\mathrm{QM} \\ c\in\mathrm{MM}}} E_{\mathrm{XD},ac}(R_{ac}; Q_a) \end{aligned} \tag{3.11}$$

where $Q_a \equiv Q_a(\mathbf{P})$ is the Mulliken charge of QM atom a

$$Q_a(\mathbf{P}) = Z_a - \sum_{\mu\nu} P_{\mu\nu} S_{\mu\nu} W_{\mu\nu}^{(a)} \tag{3.12}$$

$$W_{\mu\nu}^{(a)} = \begin{cases} 1 & \text{if } \mu\nu \in a \\ \frac{1}{2} & \text{if } \mu \in a, \nu \notin a \text{ or } \nu \in a, \mu \notin a \\ 0 & \text{otherwise} \end{cases} \tag{3.13}$$

and Z_a is the nuclear charge of atom a.

The charge-dependent QXD van der Waals energy consists of exchange-repulsion and dispersion components

$$\begin{aligned} E_{\mathrm{XD},ac}(R_{ac}; Q_a) = &E_{\mathrm{X},ac}(R_{ac}; Q_a) \\ &+ E_{\mathrm{D},ac}(R_{ac}; Q_a) \end{aligned} \tag{3.14}$$

$E_{\mathrm{X},ac}$ treats the exchange-repulsion as being proportional to the overlap of atomic densities, which Equation 3.18 models as the overlap of Slater functions $S(r;\zeta) = (\zeta^3/8\pi)e^{-\zeta r}$. The approximate relationship between exchange-repulsion and atomic

overlap was recognized 40 years ago[215] and then further developed into an anisotropic repulsion model by Wheatley.[216] Recently, similar formalisms have gained popularity in the GEM force field.[222] The model used in the present manuscript is based on previous work by Giese[172], which replaces the rigorous treatment of the electron density with an empirical model that can be tuned for high-accuracy while providing a means for incorporating charge-dependence. Specifically, the charge dependence is introduced via the Slater exponents:

$$\zeta_a \equiv \zeta_a(Q_a) = \zeta_{0,a} e^{-\zeta_{q,a} Q_a} \tag{3.15}$$

The charge dependence in this expression is handled by the $\zeta_{q,a}$ whose derivative yields a negative function. There are a few physical constraints on the form of $\zeta_a(Q_a)$; the value of this function should never be zero as the overlap between this could cause the overlap between atoms would be independent to changes in distance, the functional form of this function should be monotonic as there are no local extrema in atomic size in relation to atomic size, the function must be a positive and have a negative derivative because anions are larger and more diffuse than cations.

$E_{\mathrm{D},ac}$ is a damped[223] R^{-6} dispersion model, whose C_6 dispersion coefficient is proportional to the atomic dipole polarizability, α, and an "effective number of electrons", N_{eff}. Charge dependence is introduced by making α and N_{eff} a function of atomic charge:[224, 225]

$$N_{\mathrm{eff},a} \equiv N_{\mathrm{eff},a}(Q_a) = N_{\mathrm{eff},0,a} - Q_a \tag{3.16}$$

$$\alpha_a \equiv \alpha_a(Q_a) = \alpha_{0,a} e^{-\alpha_{q,a} Q_a} \tag{3.17}$$

The QXD model depends on the empirical parameters: s_a, $\zeta_{0,a}$, $\zeta_{q,a}$, $N_{\mathrm{eff},0,a}$, $\alpha_{0,a}$, $\alpha_{q,a}$. The values of $N_{\mathrm{eff},0,a}$, which depend on atomic number, were developed by Pellenq and Nicholson[224] and are used directly without modification; that is, $N_{\mathrm{eff},0,a}$ is not treated as a parameter to be optimized. The expressions for $E_{\mathrm{X},ac}$ and $E_{\mathrm{D},ac}$ are briefly summarized as follows: The repulsive and attractive terms can be expanded and written as

$$E_{\mathrm{X},ac}(R_{ac}; Q_a) = s_a s_c \zeta_{ac} (\Delta_{ac} - \Delta_{ca}) \tag{3.18}$$

and

$$E_{\mathrm{D},ac}(R_{ac};Q_a) = -S_{6,ac}\frac{C_{6,ac}}{R_{ac}^6} \tag{3.19}$$

where

$$\Delta_{ac} = \frac{\zeta_c e^{-\zeta_a R_{ac}}}{R_{ac}}[4\zeta_a + R_{ac}(\zeta_a^2 - \zeta_c^2)] \tag{3.20}$$

and

$$\zeta_{ac} = \frac{1}{8\pi}\frac{\zeta_a^3\zeta_c^3}{(\zeta_a^2 - \zeta_c^2)^3}. \tag{3.21}$$

In the limit $\zeta_c \to \zeta_a$, Equation 3.18 reduces to

$$E_{\mathrm{X},ac}(R_{ac};Q_a) = s_a s_c \frac{e^{-\zeta_a R_{ac}}\zeta_a^3}{192\pi}(3 + 3R_{ac}\zeta_a + R_{ac}^2\zeta_a^2). \tag{3.22}$$

The charge-dependent dispersion term is

$$C_{6,ac} = \frac{3}{2}\frac{\eta_a\eta_c}{\eta_a + \eta_c}\alpha_a\alpha_c \tag{3.23}$$

where

$$\eta_a = \sqrt{\frac{N_{\mathrm{eff},a}}{\alpha_a}}. \tag{3.24}$$

To get the total dispersion interaction energy in the QXD model the previous term is multiplied by a charge-dependent switching function

$$S_{6,ac} = 1 - e^{-b_{ac}R_{ac}}\sum_{k=0}^{6}\frac{(b_{ac}R_{ac})^k}{k!} \tag{3.25}$$

which is directly related to the repulsive potential, Equation 3.18 through

$$b_{ac} = (\Delta_{ac} - \Delta_{ca})^{-1}\frac{d}{dR_{ac}}(\Delta_{ca} - \Delta_{ac}). \tag{3.26}$$

It should be noted that all of these equations can be reduced down into expressions which consist of old single-bodied parameters, meaning that these values only depend on the atom for which they are assigned and are not influenced by the local chemical or bonding environment in ways similar to those in the LJ potential.

These equations then influence the QM calculation by entering into the Fock matrix as a correction, which for the QXD QM/MM model (Equation 3.11) is

$$\begin{aligned}\left.\frac{\partial E_{QXD}}{\partial P^{\sigma}_{\mu\nu}}\right|_{\mathbf{R},\mathbf{q}} =& \sum_{a} \left.\frac{\partial E_{QXD}}{\partial Q_a}\right|_{\mathbf{R},\mathbf{q}} \left.\frac{\partial Q_a}{\partial P^{\sigma}_{\mu\nu}}\right|_{\mathbf{R},\mathbf{q}} \\ & - \sum_{c\in \mathrm{MM}} q_c(\mu\nu|c)\end{aligned} \tag{3.27}$$

where

$$\left.\frac{\partial Q_a}{\partial P^{\sigma}_{\mu\nu}}\right|_{\mathbf{R},\mathbf{q}} = -S_{\mu\nu}W^{(a)}_{\mu\nu} \tag{3.28}$$

and $\partial E_{\mathrm{QXD}}/\partial Q_a|_{\mathbf{R},\mathbf{q}}$ is obtained from elementary chain-rule derivatives of Equations 3.14-3.26, which are detailed in Appendix C.

More modern charge-fitting procedures could be used to more accurately represent local atomic charge, such as ESP fitting. However, early tests indicated that differences in between these models and Mulliken charges largely did not affect the end results as absolute deviations were largely canceled out and subsumed into the charge-dependent parameters; rather the general trend of the local charges was vastly more important, which is followed by a Mulliken charge scheme.[226] These findings combined with the ease of including the QXD model while using Mulliken charges into the Fock matrix, the wide-spread use of Mulliken charge partitioning in QM/MM simulations (and incorporation into QM/MM software packages) and comparably lower computational burden for calculating local charge as a simulation progresses lead to the adoption of this simple charge model was over other more complex methodologies.

Upon reaching SCF convergence, the atomic gradients of atom a are

$$\begin{aligned}\left.\frac{\partial E}{\partial X_a}\right|_{\mathbf{P}} =& \frac{dE_{\mathrm{MM}}}{dX_a} + \left.\frac{\partial E_{\mathrm{QM}}}{\partial X_a}\right|_{\mathbf{P}} \\ & + \left.\frac{\partial E_{\mathrm{QXD}}}{\partial X_a}\right|_{\mathbf{P},\mathbf{q}} - \sum_{\mu\nu} Q_{\mu\nu}\frac{dS_{\mu\nu}}{dX_a}\end{aligned} \tag{3.29}$$

where

$$Q_{\mu\nu} = \sum_{\sigma}\sum_{k} n^{\sigma}_{k} E^{\sigma}_{kk} C^{\sigma}_{\mu k} C^{\sigma}_{\nu k} \tag{3.30}$$

3.2.2 Interfacing the QXD Interaction with Traditional Force Fields

The first step in implementing the QXD correction is to interface it with existing QM/MM frameworks. The QXD model would be inconvenient to use if special parameters would have to be redeveloped for all MM atom types. Fortunately, it was discovered that the LJ parameters used in current MM force fields can be suitably "mapped" into QXD parameters through a series of approximate expressions.

The prescription for mapping the MM atom LJ parameters to QXD parameters begins by removing their charge-dependence within the QXD model $\zeta_{q,c} = 0$ and $\alpha_{q,c} = 0$. The approximate relations between the remaining adjustable parameters in the QXD model to the $\varepsilon_c \equiv \sqrt{\varepsilon_{cc}}$ and $R_{\min,c} \equiv (1/2)R_{\min,cc}$ LJ parameters are:

$$s_c = A\varepsilon_c^B R_{\min,c}^C \tag{3.31}$$

$$\zeta_{0,c} = D\varepsilon_c^E R_{\min,c}^F \tag{3.32}$$

$$\alpha_{0,c} = G_c \left(\varepsilon_c R_{\min,c}^6\right)^{2/3}. \tag{3.33}$$

Parameters A-F do not depend on atom-type, nor atomic number, and can be used universally (the numerical values are listed in atomic units): $A = 9.4423$, $B = 0.4111$, $C = 2.8208$, $D = 3.7893$, $E = -0.0192$, $F = -0.7249$. The G_c parameter depends on atomic number, as to keep the mapping expressions consistent with the Pellenq and Nicholson values of $N_{\text{eff},0,c}$ used within the dispersion potential.[224] Values for G_c have been parameterized for elements H, C, N, O, and Cl by fitting the QXD van der Waals interaction energies to reproduce the LJ energy for a large number of atom-type pairs. The numerical values of these parameters are (listed in atomic units): $G_{\text{H}} = 32.7324$, $G_{\text{C}} = 19.0422$, $G_{\text{N}} = 18.2066$, $G_{\text{O}} = 17.5986$, and $G_{\text{Cl}} = 15.1066$. A detailed comparison of the mapped QXD interactions with respect to the LJ potential is presented in Sec. 3.3.1.

Initially, when trying to create a method of relating QXD potentials to LJ potentials, charge-static QXD parameters were optimized to try to best reproduce the LJ via minimizing a chi-squared function utilizing a direction set optimization scheme with 16 equidistant points along the particle-particle separation distance with a chi-squared weight proportional to their Boltzmann weight probability. Additionally, extra points

were included at the location where the LJ potential was equal to $1.0k_B$T, the contact distance (σ_{ac}), the energy at the minimum of the potential (ε_{ac}) and the distance of the minimum of the potential ($R_{min,ac}$) to increase their importance in the fitting procedure. To better fit these selected points and to maintain consistency with the charge-dependent dispersion expressions of Pellenq and Nicholson,[224] an additional parameter was introduced into the optimization fitting procedure. Values of s_c and $\zeta_{0,c}$ were obtained for a grid of LJ parameter pairs, with $R_{min,ac}$ and ε_{ac} being systematically changed within the ranges of 0.5Å to 3.75Å and 0.05 kcal/mol to 0.75 kcal/mol, respectively. It was then noted that the resulting values of the charge-static QXD parameters could be decomposed back into their LJ origins by fitting exponential equations. Shown in Figure 3.2, the LJ $R_{min,ac}$ is plotted against the value of the optimized $\zeta_{0,c}$ and at different values of ε_{ac} (the different lines). An exponential was fit to these curves, then the LJ ε_{ac} was plotted against the coefficients of the previously obtained expressions. Similar protocols for s_c were followed and yielded comparable results.

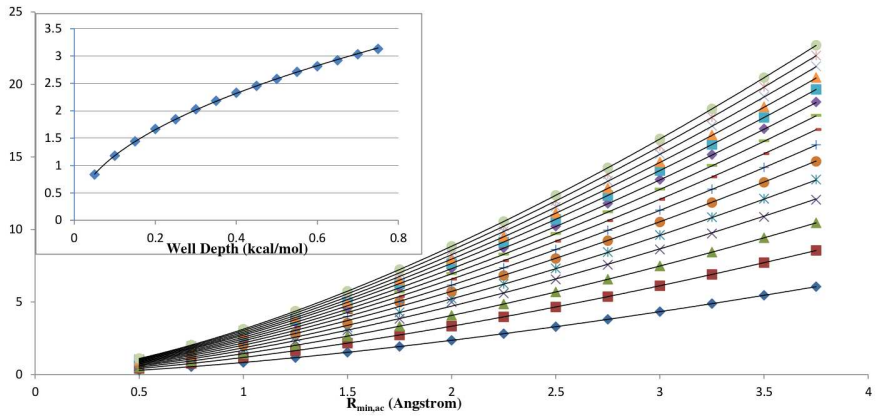


Figure 3.2: **Example plot from early tests for generating the LJ to charge-static QXD parameter expressions.**

Once these relationships were established, an direction set optimization protocol for the value of the exponential factors and coefficients of these expressions. A wider range of of $R_{min,ac}$ and ε_{ac} was examined for each element, in conjunction with the G_c factor. The total optimization protocol was evaluated to best reproduce the LJ interaction energy of different elements with ranges of LJ parameters as to have a generalized, but accurate method of converting LJ parameters into charge-static QXD parameters

on-the-fly during a simulation.

Therefore, the QXD parameters of the MM atoms are obtained directly from their LJ parameters using the prescription described above, whereas the QXD parameters of the QM atoms are parameterized for the specific QM Hamiltonian used. The description of the QM atom QXD parameters are shown in Table 3.3 and discussed in Sec. 3.3.3.

3.2.3 High-level Quantum Chemical Calculation Protocol

Gas-phase adiabatic scans of a water probe of various molecules of interest were carried out via constrained optimization to demonstrate the viability and flexibility of the QXD interaction model, while also achieving a base line departure point for solution phase parameters. In these scans the quantum-level water is constrained to the TIP4P-Ew[160] geometry. The electronic degrees of freedom and the geometry of the molecule of interest were allowed to relax. The water probe is oriented with the hydrogen atoms facing the particle of interest the and the axis of the scan bifurcating the hydrogen-oxygen-hydrogen angle. High-level calculations were performed with the Gaussian09 software suite[152] in conjunction with the GaussView tool.[153] using the M06-2X functional[148] at the mixed 6-31+g(d,p)//6-311++g(3df,2p) level using an ultrafine integration grid for both the geometry and energy evaluations, with tight SCF and optimization criteria. Gas phase QXD scans were performed in the AMBER12 software package[158] with the water probe represented by an MM TIP4P-Ew model and an infinite non-bonded cutoff was employed. QXD parameters were optimized via steepest descent optimization in conjunction with a direction set minimization routine, utilizing a Boltzman weighted chi-squared protocol with additional weighting for reproducing the zero and minimum of the interaction energy.

3.2.4 Solution Phase Simulation Methods

Solution phase Thermodynamic Integration (TI) simulations were performed with the AMBER12 software package[158] in TIP4P-Ew[160] to parameterize the QXD model for condensed phase simulations. All simulations use periodic boundary conditions (PBCs) with a particle mesh Ewald treatment for long-range electrostatics, the QM/MM long-range interaction switch fix, tight density convergence, and an SCF convergence criteria

set to 1.0x10^{-10}. A tighter density and SCF convergence criteria were desirable due to the importance of accurately modeling the electronic density as it now imapcted the QXD interaction and, in turn, affect the surrounding MM environment through the exhange corrilation and disperion interactions.

Parameters are refined through the use of dual-topology, single coordinate TI[83, 227, 84, 228] simulations to accurately calculate solvation free energies using the thermodynamic cycle shown in Figure 3.3. TI mixes Hamiltonians of two different potential states (V_0 and V_1) to calculate the free energy between those states by moving the system along an artificial, alchemical pathway.

$$\Delta G = G(\lambda = 1) - G(\lambda = 0) = \int_0^1 \langle \partial V / \partial \lambda \rangle_\lambda d\lambda \tag{3.34}$$

where

$$V(\lambda) = (1 - \lambda)^k V_0 + [1 - (1 - \lambda)^k] V_1 \tag{3.35}$$

Simulation protocol was tested for accuracy, examining box size and long-range electrostatic effects. The protocol was adjusted accordingly to eliminate any systematic errors. Following the thermodynamic cycle in Fig 3.3, contributions from each leg can be analyzed separately for their contributions to the total solvation energy. Of particular note; the gas phase legs of the cycle involving adjusting non-bonded terms can be ignored as, by design, the contribution to the overall solvation free energy is zero. The gas phase MM to QM leg, designated as ΔG_6, is required to correct for differences in arbitrary zeros of the QM and MM representations. Parameterization of QXD is hastened through intelligent division the cycle and by minimizing the free energy differences experienced in each individual step. To this end, trial LJ parameters were created for each species to approximate their solvation free energy accurately, allowing for the ΔG_4 step to be repeatedly run to test small adjustments of QXD parameters without needing additional simulations of other areas in the thermodynamic cycle. Each leg of the thermodynamic cycle is divided into eleven evenly spaced windows form $\lambda = 0.0$ to $\lambda = 1.0$ with a linear coupling scheme ($k = 1$) with the exclusion of the calculation of ΔG_1. In this transformation a softcore potential scheme[229] is used to avoid the "end-point catastrophe" when disappearing a LJ sphere in solution. The endpoint windows in these legs are moved to $\lambda = 0.05$ and $\lambda = 0.95$. Data was processed using a smoothing Akima spline

(and in softcore cases, also extrapolated to the endpoint values of 0.0 and 1.0) to the free energy $\partial V/\partial\lambda$ plots and then those splines were integrated with respect to λ to calculate the free energy contribution from each change. NPT equilibration of the the system was run for 250ps to allow for solvent and box size relaxation. Afterwards, each λ window was NPT equibrated for an additional 250ps. Finally, NVE production was ran for 250ps, analyzed and used in parameterization of the QXD model.

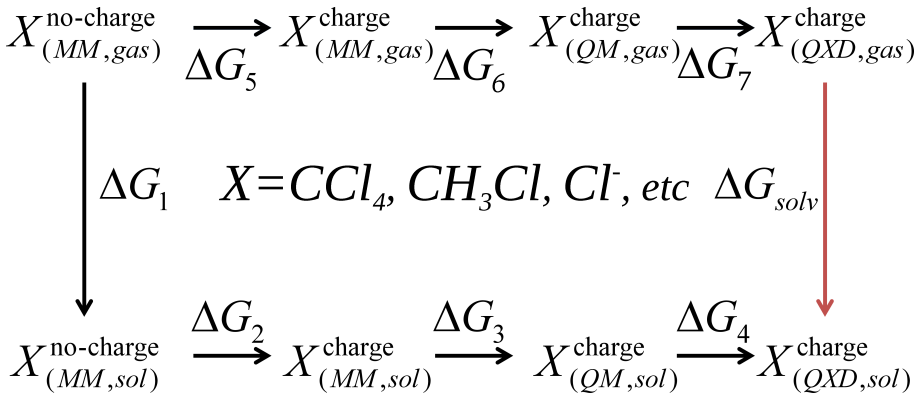


Figure 3.3: **Detailed thermodynamic cycle use in the calculation of the solvation free energies of small chlorine-containing compounds for the parameterization of the QXD model.** Shown is the cycle used for calculating the solvation free energies for the parameterization of the charge-dependent QM/MM interaction model, QXD. Cycle legs shown in black are simulated using alchemical perturbation (where needed) and the leg in red is the target free energy. For cycles not using QXD, those legs are ignored.

Parameters for the QXD model, due to the length of simulation and detailed analysis required, were adjusted by hand in order to obtain the best results. Parameters for the chloride anion, methylchloride, and carbon tetrachloride were initially primed based off of high-level, gas phase QM calculations whose interactions were then fit with LJ parameters which then provided charge-static QXD parameters via the relationship between QXD and LJ parameters developed herein. To accompany these rudimentary initial guesses, a wide swath of parameter space was probed in a systematic and incrementally fashion. Then, upon reaching a region in which a minima of parameter space was believed to reside, several hundred similar parameter sets were tested within the local region in order to determine an optimized set. This procedure was repeated for each different region for which the initial scan indicated the possibility of meaningful

results. This procedure was hastened after obtaining a parameterization of QXD which produced the solvation free energy for a given particle, allowing the interaction curve generated by those parameters with an MM water probe to act as a screening agent for other possible parameterization candidates. Trial parameter sets could be refined using an automated, direction set minimization to the previously obtained interaction curve before simulation. While automation of the entire procedure would be desirable, it is not be computationally tractable. Each step in the optimization procedure requires at least one TI simulation and extensive data analysis. Additionally, the expressions in the QXD model are non-linear, non-independent equations whose parameters do not necessarily have well-behaved gradients with respect to the desired thermodynamic observables, which could lead many automation procedures astray and does not lend itself to some of the more modern optimization techniques.

Then, after initial parameterization, the attack of a chloride ion on methylchloride will be examined as an example reaction where there is a large amount of charge transfer along the reaction coordinate (RxC).[230] General simulation protocol is identical to the QM legs of the thermodynamic cycle which were outline in the previous section. Initial NPT simulations for solvent and box size equilibration were ran for 250ps followed by another 150ps of NPT equilibration at each different RxC window. Finally, 150ps of NVE production was ran and used for analysis. The RxC is defined as a linear combination of orthogonal degrees of freedom, R_1-R_2, where R_1 and R_2 are defined as the distance between the attacking/leaving chloride and the carbon atoms. Umbrella windows were evenly spaced every 0.2Å between RxC values of -3.5Å and 3.5Å with force constants ranging from 60 kcal/mol/Å and 24 kcal/mol/Å depending on location in the RxC. Additional windows between -0.8Å and 0.8Å with force constants ranging between 200kcal/mol/Å and 184 kcal/mol/Å were used to ensure adequate sampling about the transition state of the reaction. Simulation results were analyzed using vFEP[79, 78] to obtain the solution phase free energy profile.

3.3 Results and Discussion

In this section the main results this chapter are discussed. First, the performance of the interface between QXD and traditional QM/MM models will be reviewed and discussed. Next, comparisons will be drawn between the QXD model and the LJ model in their ability to reproduce high-level gas phase interaction curves. Third, the parameterization of QXD for reproducing experimental solvation free energies of various chlorine containing species will be assessed. Fourth, free energy profile results of the previous parameterization for the the attack of chloride on methylchloride will be examined. Lastly, an overall perspective of the QXD model and its future directions are discussed.

3.3.1 Performance of the Lennard-Jones to QXD Interface Relations

Integration of QXD into conventional QM/MM frameworks is key to facilitate widespread use of the model. Toward this end, a empirical model outlined through Eqs. 3.31-3.33 was developed that, as seen in Figure 3.4, is capable of transforming traditional LJ parameters into charge-static terms which operate within a QXD framework. This way, one can avoid having to generate MM QXD parameters for any number of solvents/counter ions/etc which already posess finely tuned LJ parameters. The mapping equations accurately reproduce potentials for any LJ pair with atom $R_{\min,c}$ and ε_c between 0.5-3.5 Å and 0.001-0.9 kcal/mol, respectively, with a charge-static QXD potential. All of the parameters in these expressions are independent of the local electronic environment atoms, avoiding the shortcomings that limit current atom typing models. Differences between the mapped QXD and the LJ interaction curves were tested with thermodynamic integration simulations and found to be negligible with respect to statistical uncertainties typically on the order of 0.1 kcal/mol or less. With this empirical mapping, quantum mechanical atoms that have fully charge-dependent QXD parameters can interact with MM atoms that have static (fixed-charge) QXD parameters derived from their LJ parameters.

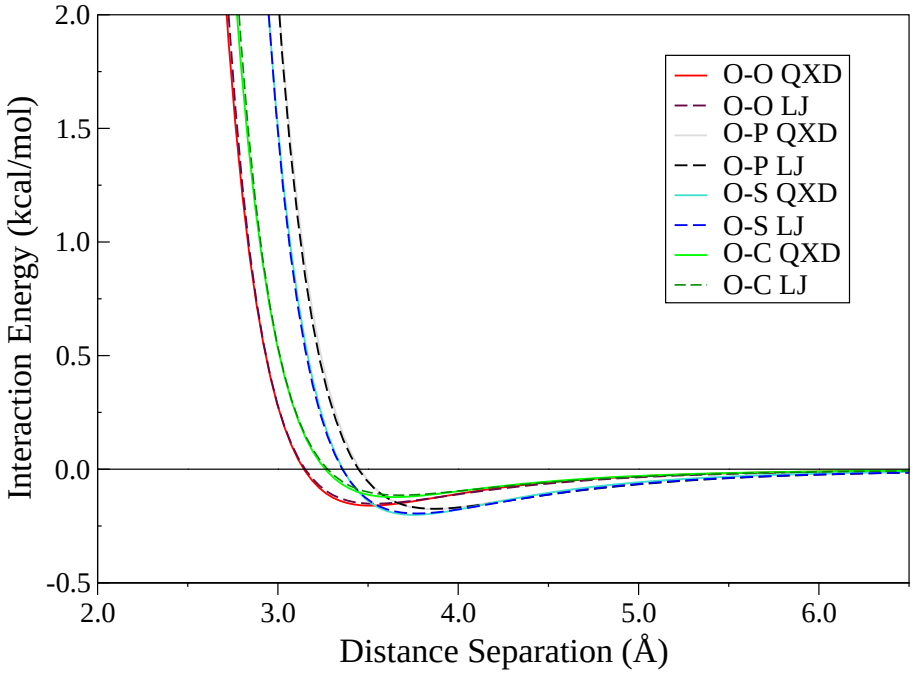


Figure 3.4: **Lennard-Jones interactions compared with corresponding mapped parameter QXD interactions.** Shown are various atom-atom (X-O; X=O,P,S,C) interactions for both traditional Lennard-Jones and QXD models, with QXD charge-independent parameters generated from the interfacing equations which allow for on-the-fly conversion of LJ interactions into a form which can be used with the new QXD charge-dependent interaction model.

3.3.2 Robustness of the QXD Model as Compared to High-Level Gas Phase Interactions

High-level adiabatic QM scans were created to demonstrate the flexibility of the QXD model in reproducing interactions of systems through various charge-states. Initially, two different sets of LJ parameters were optimized to reproduce the high-level scans of the chloride anion and of the methylchloride molecule, respectively. Then, a signle set of charge-dependent QXD parameters was optimized to best reproduce all three of the reference potentials shown. Potentials from these optimizations and the resulting scans are shown in Figure 3.5.

The charge-dependent QXD model stands as a marked improvement when compared to the conventional LJ model in its ability to capture the high-level gas-phase data.

The two different LJ 12-6 potentials each reasonably model the interaction for which they were parameterized to reproduce, however each fails to simultaneously capture the other interactions within a single parameterization. When trying to recapitulate the Cl^- interaction, the LJ potential drastically under-stabilizes the neutral chlorinated species. Likewise, when the LJ potential is parameterized to model the methylchloride interaction the neutral particles (CH_3Cl and CCl_4) interaction curves reasonably well, however the parameterization significantly over-stabilizes the chloride anion. Without the ability to respond to changes in local electronic environment, the LJ potential lacks the flexibility to represent atoms over a wide range of charge states. The QXD model, on the other hand, is able to accurately capture all of the high-level quantum data with the small exception that there is a slight deviation in the Cl^-/water interaction curve.

The discrepancy between QXD and the high-level reference data at the minimum of the interaction curve is likely due to the limited electronic degrees of freedom in the semiempirical and MM models. Despite the high-level water having its geometry fixed to that of the TIP4P-Ew water model, differences between the reference and the QXD scans arise because the electronic density about the atoms of the high-level reference are still allowed to fully polarize. The QM/MM simulations do not allow the electronic density about the chloride anion (and, clearly, around the MM water) to relax as freely as the high-level reference and thus the quantum scans gain additional stabilization beyond what the QXD model can compensate for. Nonetheless, the QXD model demonstrates greater flexibility and robustness in capturing interactions at varied local charge state than the traditional LJ model and represents a significant improvement over the LJ potential in recapitulating high-level quantum data.

3.3.3 Thermodynamic Integration and Solvation Free Energy Data

QM/MM solvation free energies were calculated through the use of TI and were used in the parameterization of chlorine QXD atoms for solution phase simulations. Parameters for carbon and hydrogen atoms were provided from the mapping equations with additional charge dependence assigned to these atoms to prevent artificial charge transfer between atoms. Chlorine parameters were adjusted to best reproduce the experimental solvation free energy data; the resulting parameters can be found in Table 3.3.

	ΔG, Solvation Free Energy			
Compoud	Experimental[231, 232, 233, 234, 235, 236]	LJ_{Cl^-}	LJ_{CH_3Cl}	QXD
Cl^- Anion	-89.1	-89.21	-125.67	-89.64
CH_3Cl	-1.02 to 0.31	5.64	0.34	0.53
CH_2Cl_2	-1.90 to -0.65	8.11	0.14	-1.37
$CHCl_3$	-1.78 to -0.38	11.43	0.62	0.33
CCl_4	-0.64 to 1.10	14.66	2.37	1.20
CCl_3CHCl_2	-2.32 to -0.24	17.63	1.51	0.28
CCl_3CCl_3	-1.57 to -0.93	33.84	4.53	2.94

Table 3.2: **Comparitive solvation data for two parameterizations of the traditional Lennard-Jones potentials and a single parameterization of the QXD model.** Calculated QM/MM solvation free energies for a series different chlorine-containing compounds, using two parameterizations of the traditional Lennard-Jones model, one made to produce the chloride solvation free energy and the other to capture the CH_3Cl solvation free energy, and the a single parameterization of the QXD density-dependent interaction model. Free energies provided are in kcal/mol.

The QXD model for non-bonded QM/MM interactions shows promise in accurately capturing experimentally known solvation free energies over a series of chlorine-containing compounds and represents a marked improvement over the results obtained when employing a traditional LJ model. While individual sets of parameters in either model could be generated to reproduce the solvation free energy for a single given species, the QXD model is able to accurately replicate experimental solvation free energies of several different species in a variety of local electronic states, while the LJ potential is incapable of this task. For comparison, data from QXD and two sets of LJ parameters can be found in Table 3.2. The LJ parameter sets used where chosen to represent two different chlorine 'atom types': one set that accurately reproduces the solvation free energy of a chloride anion, LJ_{Cl^-}, and another set that captures the methylchloride solvation free energy, LJ_{CH_3Cl}.

The LJ_{Cl^-} set of parameters accurately reproduces the chloride anion solvation free energy, at -89.21 kcal/mol, however for all other species it consistently undersolvates the charge neutral species. This is especially notable on molecules containing several chlorine atoms, such as hexachloroethane which is undersolvated by nearly 35 kcal/mol.

This behavior is expected for this parameterization as a strong repulsive potential is required to overcome the extremely favorable charge/partial-charge interactions that the chloride ion would experience in solution. As the traditional LJ model cannot respond to changes in the local electronic density, the 'size' of the chlorine atoms represented in the non-ionic species are the same as those for the chloride anion. From a physical standpoint, given that the local atomic charge around a chlorine atom in a neutral species is considerably less than that of a chloride anion with full -1 charge, the interactions should reflect the smaller, harder character and have a steeper repulsive wall and closer contact distance. Thus, the LJ_{Cl^-} parameter set cause the chlorine atoms in the molecular compounds to interact as if they were too large and soft, reducing the ability of water to effectively solvate.

The LJ_{CH_3Cl} set of parameters captures the solvation free energy of methylchloride, reproducing the upper experimental range at 0.34 kcal/mol. Also, because the methylchloride is electronically more similar to several other molecules in the test set than the chloride anion, the LJ_{CH_3Cl} parameterization generally out performs the LJ_{Cl^-} set as a whole. However, it still fails dramatically is several cases. Most predominately, the chloride anion is oversolvated by approximately 35 kcal/mol. Furthermore, while the methylchloride has the appropriate amount of solvation, the hexachloroethane molecule remains undersolvated by approximately 5.5 kcal/mol. Similar arguments could be made as in the previous paragraph as to why these results are not unexpected. The static nature of the LJ potential means that if the atom being modeled is in a different local electronic environment, then new parameters must be generated to accurately model the non-bonded interactions.

The QXD model, however, is able to capture the solvation free energies for nearly all species with errors of approximately 1.0 kcal/mol or less. The QXD model provides a much more accurate representation of the non-bonded interactions, reproducing solvation free energies for both the chloride anion, at -89.64 kcal/mol, and the methylchloride molecule, at 0.53 kcal/mol. Error can be seen in the hexachloroethane molecule, undersolvating the species by about 4 kcal/mol, however this number is still in better agreement with experiment than either of the individual LJ models. QXD overcomes the limitations of the LJ model by allowing the non-bonded interaction to adjust to

changes in the local electronic density and to behave in a more physically realistic manner. As seen in Table 3.2, more accurate solvation results can be achieved when using the QXD model within a single parameterization than with the tradition LJ potential. The accuracy gained through QXD model does have a cost; requiring the evaluation of additional, comparatively complex expressions for the exchange-correlation and dispersion interactions during the QM SCF procedure has been noted to increase the computational burden of this evaluation by approximately a factor of two in simulations similar to those presented in this work as compared to simulations utilize traditional LJ interactions for the current, non-optimized implementation of the model. The ability of the QXD model to adjust to different electronic environments may ultimately allow the need for atom types to be eliminated in QM/MM calculations, which will have particular impact in studies of chemical reactions where local charge changes along the reaction coordinate. It is believed that the errors in the QXD solvation data likely arise from the nature of the standard QM/MM model, specifically from the fact that the surrounding TIP4P-Ew water does not polarize appropriately to the solute and, so a lesser but still significant extent, the lack of electronic degrees of freedom within the QM Hamiltonian itself. Further discussion on this topic can be found in Section 3.3.5.

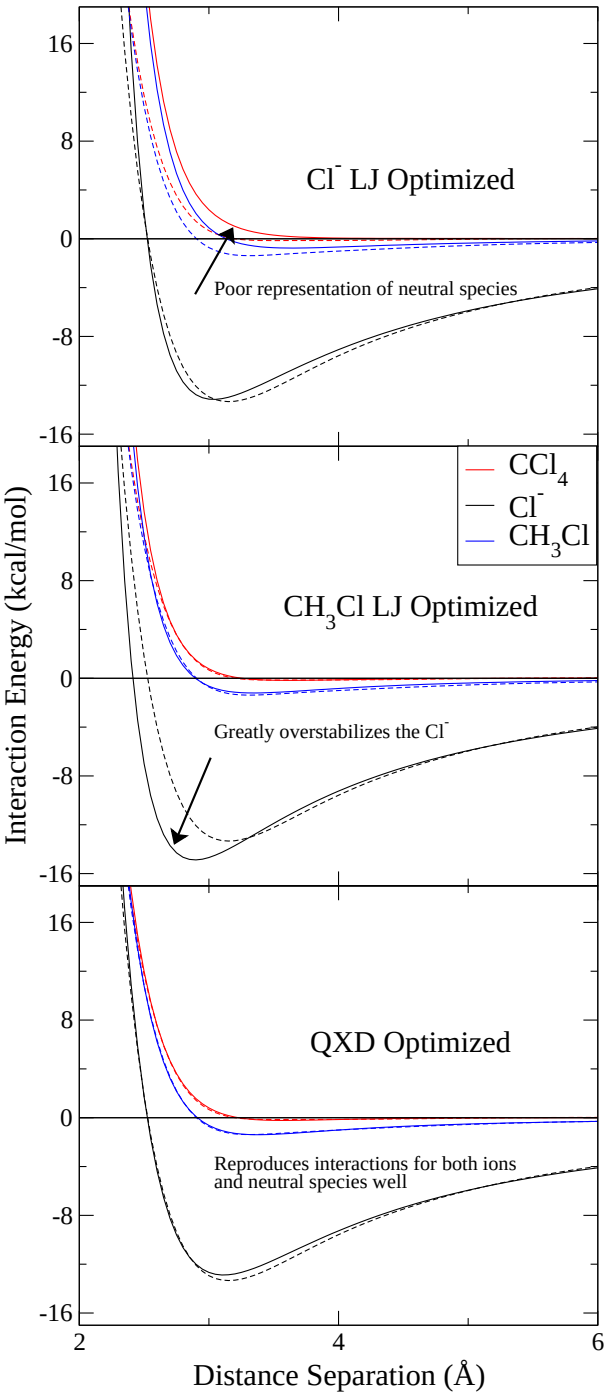


Figure 3.5: **Comparative X-H_2O gas phase LJ and QXD interactions.** Gas phase QM/MM adiabatic energy curves (solid lines) of chlorine containing compounds (X=Cl^-,CH_3Cl and CCl_4) interacting with a water probe are compared to a high-level quantum benchmark (dashed lines). Water geometry was fixed to emulate a TIP4P-Ew water. Shown are the results from interactions with LJ parameters optimized to reproduce the Cl^- reference energy (top), LJ parameters optimized to reproduce the CCl_4 reference energy (middle), and QXD with full charge dependent interactions (bottom).

3.3.4 Free Energy Profile Simulation Results

Free energy profile simulations were performed in order to test the effects of the QXD model on the attack barrier of a chloride anion reacting with methylchloride. In previous studies,[209] the traditional LJ model was shown to accurately capture this reaction barrier while using the LJ_{Cl^-} parameters set. The QXD parameter set used for these simulations is the set used to recapitulate experimental solvation free energies outlined in the previous section. Results of these simulations, and comparisons to the standard LJ model with both the LJ_{Cl^-} and LJ_{CH_3Cl} parameter sets, can be seen in Figure 3.6.

Type	s_i	$\zeta_i(0)$	$\zeta_{q,i}$	$\alpha_i(0)$	$\alpha_{q,i}$	$N_{eff,i}$
Cl/Cl^-	11.300	1.970	-0.225	26.31	0.30045	5.551
C	21.308	2.487	-0.215	8.553	0.00000	2.657
H	5.2423	3.078	-0.205	1.746	0.00000	0.824

Table 3.3: **Optimized QXD QM/MM interaction parameters for chlorine containing compounds.** Shown are the parameters used to accurately predict the solvation free energy for the series of chlorine containing compounds with the QXD charge-dependent QM/MM interaction model. Parameters for chlorine were optimized by hand and Parameters for hydrogen and oxygen were obtained by the mapping equations and slight modulation to the charge dependent parameters to avoid non-physical charge relocation during reaction simulations.

The QXD model predicts the attack barrier for the reaction of a chloride anion on methylchloride to be 25.4 kcal/mol, roughly 1 kcal/mol less than the experimentally estimated value of 26.5 kcal/mol. The LJ_{Cl^-} and LJ_{CH_3Cl} parameter sets bracket the QXD reaction barrier: the LJ_{Cl^-} set, which as discussed above produces the worst solvation free energies for neutral chlorine-containing compounds, matches very closely the experimental Cl^- attack barrier, whereas the LJ_{CH_3Cl} set underestimates the barrier by about 3 kcal/mol. The reason for the somewhat fortuitous agreement of the LJ_{Cl^-} parameter set with the experimentally estimated barrier is discussed in more detail in Section 3.3.5.

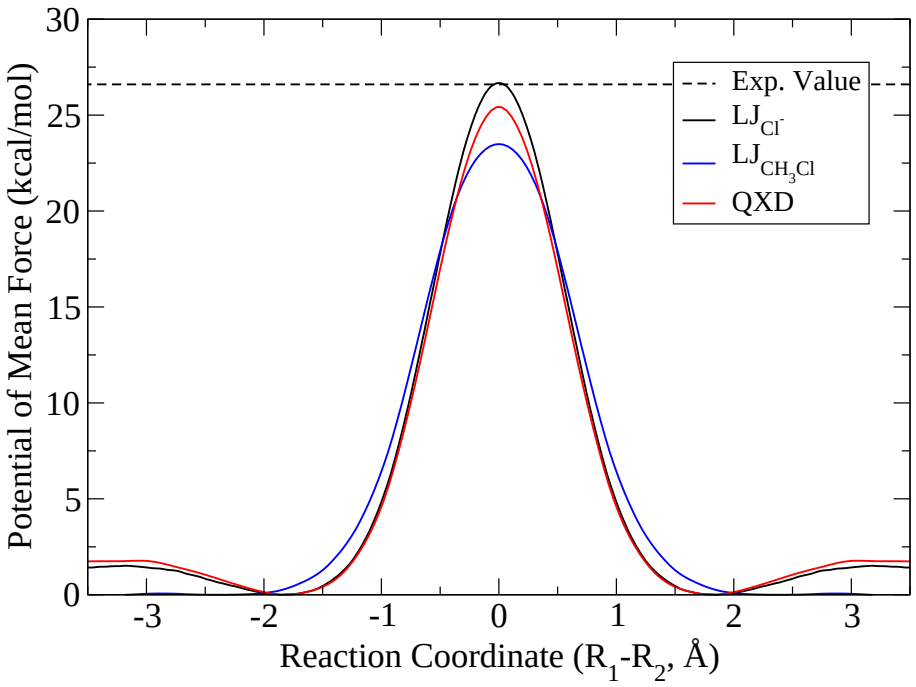


Figure 3.6: **Comparative LJ and QXD free energy profiles for the attack of chloride on methylchloride.** Free energy profiles were run to examine the effects of the different parameter sets for the traditional LJ potential and the QXD charge-dependent potential had on the attack barrier of the symmetric reaction of Cl^- with methylchloride. The simulated free energy barriers for the LJ_{Cl^-}, LJ_{CH_3Cl}, and QXD (26.5, 23.5 and 25.4 kcal/mol) parameterizations are compared to the experimental value (26.5 kcal/mol estimated from the rate constant).

3.3.5 Critical Assessment of the QXD Model and Discussion of Errors

While the results of the current study are encouraging, it is nonetheless important to point out limitations that remain within the scope of the current work. The QXD model, as represented in this manuscript, is hindered by the application of a standard QM/MM protocol. Errors arise from 1.) the lack of explicit many-body response of the MM solvent and 2) the lack of flexibility inherent in the minimal valence basis set used in the semiempirical quantum model. Both of these shortcomings factor directly into the parameterization of the models presented here, and thus warrant discussion in order to properly interpret results.

Without the appropriate many-body response of the MM solvent, differently charged species will experience electrostatic interactions with the same set of MM charges on the water molecules. Similarly, the minimal valence basis set used in the semiempirical quantum model manifests limitations mainly as a systematic underestimation of the electronic polarization response that becomes worse as anionic charge increases. [172] In the extreme cases, such as that of a Cl^- anion in isolation, the orbitals are fully occupied, leaving no virtual orbitals available for polarization. Each of these effects require compensating adjustment of the non-electrostatic non-bonded parameters (i.e., LJ or QXD) to overcome their shortcomings in an effort to reproduce experimental solvation free energies.

There is a large body of literature [237, 238, 239, 240, 241, 242] that specifically examines the importance of polarization and quantum many-body effects for modeling a solvated chloride anion. While the degree to which polarization plays a roll in correctly predicting the solvation free energy and long-range solvent structure of the chloride anion is still debated, it is clear that some amount of polarization (in both the solute and the solvent) is required if both of these observables are to be accurately be represented. The mismatch of the polarization effect on Cl^- and CH_3Cl at different stages along the reaction coordinate is more difficult to correct within the scope of the current work. For instance, in the previous section, Section 3.3.4, the QXD model predicts the attack barrier for the reaction of a chloride anion on methylchloride to be 25.4 kcal/mol, roughly a kcal/mol under the the experimentally observed value of 26.5 kcal/mol. The LJ_{Cl^-} parameter set, on the other hand, agrees with experiment to within the statistical error of the calculation which, in retrospect, may not be all that surprising because the LJ

terms are consistent with the solvation free energy of a non-polarizable MM Cl^- anion, and due to the minimal valence basis set, the QM Cl^- anion also lacks polarization. The lower barrier of the LJ_{CH_3Cl} parameter set, which models the Cl atoms as being smaller, reflects the preferential stabilization of the transition state where both Cl atoms carry a significant charge (approximately 0.89 e in the simulation, see Appendix D).

However, even with this knowledge, the authors chose to pursue a route in which a non-polarizable force field and a minimal valence semiempirical Hamiltonian were used in the study for a number of reasons. First, the authors wanted to demonstrate, in as direct a manner as possible, the importance that the non-bonded non-electrostatic interactions have on QM/MM simulations without convoluting the overall picture with the inclusion of more complex, albeit more physically meaningful, MM models. Additionally, the authors wanted to show the impact of the inclusion of such a correction on conventional QM/MM simulations and that QXD could yield improved results even without the inclusion of other, more rigorous models. Clearly however, the parameterization of this interaction is intrinsically linked to not only the QM method and the manner in which atomic charge is partitioned among atoms but also the treatment of the MM surroundings. If aspects of the QM/MM framework are changed, small alterations in the parameter set may be required in order to obtain the best results.

A physically meaningful QXD parameterization may not be achievable at this time, due to limitations within the current QM/MM framework. If QXD were to be applied to a system where the QM region were to have all of the flexibility required to accurately account for atomic/molecular polarization response (studies suggest something on the order of B3LYP 6/311++g(3df,3pd))[242] in conjunction with a polarizable MM region then such a real, physically meaningful parameters could be derived. However, with current computational limitations such simulations are intractable for use in parameterization and much less so in long-time simulation of nearly any system. If such a parameterization were to be attempted, or even completed, the resulting model would exist primarily for academic interest as the application of the model would be extremely narrow; limited in scope due to the highly demanding computational rigor demanded by the calculation and the particular, rather uncommon, QM/MM protocol used.

Future work has promise to overcome both of the limitations described herein. Recently, methods have been recently introduced [172, 243] that use “chemical potential

equalization" [244] to overcome the problem of poor modeling of electronic response properties with approximate QM models without increasing the size of the minimal valence basis set (and thus remaining highly efficient). Further, the ongoing development of quantum mechanical force fields (QMFFs) poses a possible solution to seamlessly allow mutual polarization and other many-body quantum effects to be modeled for very large systems with practical efficiency. [245, 177, 178]

3.4 Conclusion

Herein we develop a charge-dependent QXD model for exchange and dispersion interactions in QM/MM simulations that is demonstrated to accurately capture the reaction barrier of the attack of a chloride anion on methylchloride (within approximately 1 kcal/mol) while simultaneously predicting experimental solvation free energies for a series of chlorine-containing compounds. This represents a feat which conventional LJ models fail to accomplish. Additionally, QXD has demonstrated superior flexibility over LJ models by being able to accurately recapitulate high-level gas-phase intermolecular interactions.

As it currently stands, the QXD QM/MM interaction model offers an attractive alternative to traditional LJ interactions for chemical reactions or other processes where changes in local atomic charge occur. An advantage of the QXD model is that is surmounts the problem of pre-assigning non-bonded parameters based on 'atom types' corresponding to a particular chemical environment. The QXD model may therefore be useful in simulations of pK_a and pK_a shifts, allowing for the correct response in the non-classical terms of key residues in large biopolymers as they pass catalytic steps in a reaction pathway. Furthermore, the QXD methodology could be applied to next-generation quantum mechanical force fields, which still do not explicitly couple the non-bond, non-electrostatic interactions directly to the underlying electronic structure.

3.5 Acknowledgements

The authors are grateful for financial support provided by the National Institutes of Health (GM107485 to DY). Computational resources were provided by the Minnesota

Supercomputing Institute for Advanced Computational Research (MSI) and Rutgers, the State university of New Jersey.

Chapter 4

VRSCOSMO: Charge-dependent Implicit Solvent Model Implemented with DFTB3/3OB

4.1 General Background

Computational study of biological reactions is of great interest to the scientific community seeking to gain molecular-level insight into reaction mechanisms. These studies encompass a wide range of archetypes including, but not limited to: reactions happening within protein environments,[188, 246, 247] model phosphoryl transfer reactions,[248, 249, 250] and studies of RNA catalysis;[182, 183, 184, 251] all of which are readily applicable to the deign of new drugs and therapeutic treatments.[77, 252, 253] Biological reactions almost exclusively take place in solution, having profound ramifications on the energetic pathways of the reaction mechanism.[191, 192, 209] This phenomena is especially apparent when studying highly-charged systems, such as RNA-like model systems[190, 193] as shown in Figure 4.1.

[3] This chapter compiles material with permission sumbitted for publication in The Journal of chemical physics from the manuscript Kuechler, Erich R., Timothy J. Giese and Darrin M. York. "VRSCOSMO: a smooth charge-dependent implicit solvent model implemented with DFTB3/3OB" The Journal of chemical physics **submitted**. Copyright 2015, AIP Publishing LLC.

An attractive method for reducing the computational burden of modeling the condensed phase environment is to employ an implicit solvent model.[254, 255, 256, 257] Several different implicit models; Poisson-Boltzmann (PB),[58] Minnesota solvation (SMx),[59, 60, 61] conductor-like screening (COSMO),[62, 63, 64, 258, 259] general Borne (GB),[65, 66] and polarizable continuum (PCM)[67, 68, 69] models; levy this advantage by removing explicit solvent from the system and then represent the condensed phase environment in some average way to alleviate the computational bottle-neck of sampling explicit solvent degrees of freedom. After which, if explicit solvent is desired, hybrid quantum mechanical/molecular mechanical (QM/MM) simulations[106, 26, 260] can be run using informed biasing potentials and reaction coordinates with the insight gained from the implicit solvent calculation.

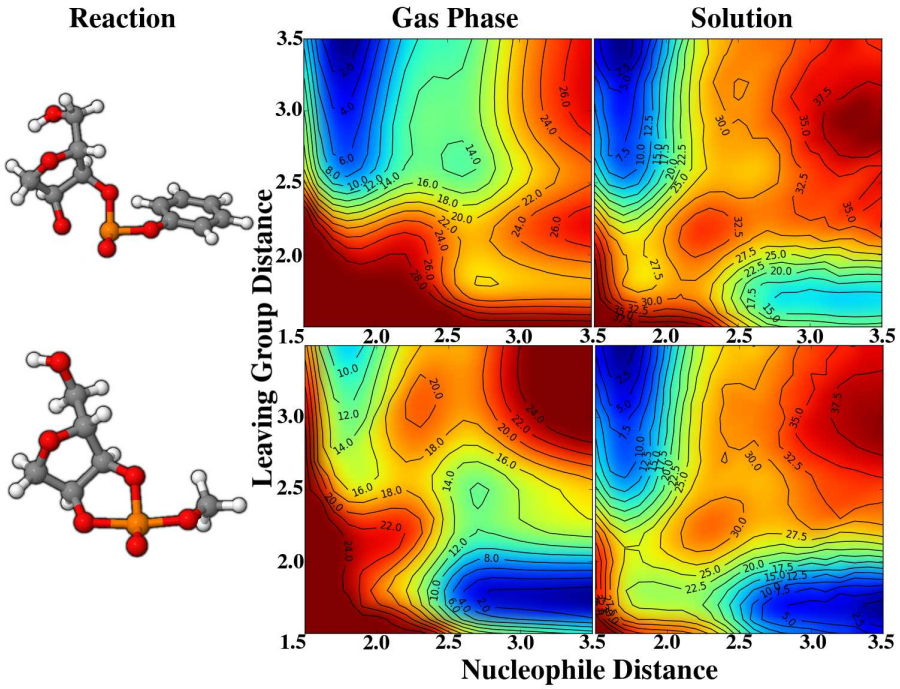


Figure 4.1: **Comparative gas and solution phase reaction energy surfaces for two model RNA-like reactions.** Solvation has a profound effect on the topology of potential energy surfaces, pathways and barriers for chemical reactions of charged systems. Special note should be made that the color scale between the gas and solution surfaces are different. All distances and energies are reported in Å and kcal/mol respectively.

Even when using an implicit solvation model the cost of high-level quantum mechanical (QM) techniques can still be computationally prohibitive, especially when the system size is large. Therefore, it can be worthwhile to consider approximate QM methods for use with implicit solvation models. Recent advancements have greatly increased the accuracy and transferability of these methods, such as those in the density functional tight-binding (DFTB) models.[261, 262] DFTB models have shown themselves to be comparable to other semiempirical methods, such as AM1[114, 263] and PM3,[264, 265] in terms of accuracy and computational efficiency[266, 267] and special interest has been garnered with the development of a a DFTB model parameterization that has been specially tuned for chemistry involving sulfur and phosphorus atoms;[268] opening avenues for the study of RNA and RNA model compounds.

RNA-like reactions are highly-charged in nature and employing an implicit solvation model whose solvation cavity is defined by atom-typed radii, which remain static during optimization and change discontinuously as a function of local charge state, will result in a poorly modeled reaction energy surface. As such, the present chapter develops a charge-dependent model. Following the pioneering work of Hou *et al.* which introduced charge-dependent radii into a nonlinear PB model (SCC-DFTBPR)[269] as well as others,[270, 271, 272] an implementation of the smooth conductor-like screening model (SCOSMO) implicit solvation model has been developed for use with the biologically focused parameterization of DFTB3/3OB-OPhyd.[273, 274, 268] This variable-radii SCOSMO model (VRSCOSMO) allows the solvation cavity of atomic species to change in accordance to the local atomic charge while leveraging the analytically smooth gradients afforded by the SCOSMO framework. Parameterizations of both SCOSMO and VRSCOSMO are developed independently to reproduce a set of absolute solvation energies and are compared Hou's charge-dependent PB model (Ref. 269). The VRSCOSMO model is then used to construct two-dimensional, relaxed potential energy surfaces of biologically motivated reactions and the results of these calculations compared to available experimental data. Given the specific interest of quickly identifying potentially important energetic pathways for RNA catalysis this new model, which has been designed for this specific derivation of DFTB, has impact on the greater biochemical community.

4.2 Computational Methods

This chapter uses the semiempirical DFTB3/3OB Hamiltonian [274] supplemented with a specific reaction parameterization, OPhyd,[268] which modulates the O and P parameters for improving performance on phosphate hydrolysis reactions. The DFTB3/3OB-OPhyd parameters nor the functional form for the solute-solute self-interaction energy have been altered in the current work. Therefore, for further discussion of technical details of the DFTB3 model the reader is referred to Ref. 274, 268, 273. It can be stated that standard gas-phase DFTB3 calculations depend on the nuclear positions and the atomic orbital density matrix

$$E_{\text{gas}} \equiv E_{\text{DFTB3}}(\mathbf{R}, \mathbf{P}) \tag{4.1}$$

where $\mathbf{R}$ is a vector of atomic positions,

$$\mathbf{P} = \mathbf{P}^{\alpha} + \mathbf{P}^{\beta}, \tag{4.2}$$

$$P^{\sigma}_{\mu\nu} = \sum_{i} n^{\sigma}_{i} C^{\sigma}_{\mu i} C^{\sigma}_{\nu i}, \tag{4.3}$$

$\mathbf{n}^{\sigma}$ and $\mathbf{C}^{\sigma}$ are spin-resolved orbital occupation numbers and molecular orbital coefficients, respectively, in the atomic orbital basis of μ and ν.

4.2.1 The SCOSMO Implicit Solvation Model

The SCOSMO model[275, 276] treats the solvent environment as if it were a block of conducting metal whose dielectric constant is chosen to mimic solvent on a macroscopic scale. Then, a cavity with a dielectric constant of unity, $\epsilon_0 = 1$, is carved out of the metal and the solute is placed within it. Given the discontinuous nature between the cavity and the bulk solvent dielectric constants at the cavity's boundary, the electrostatic response of the metal occurs solely at the cavity surface. The cavity is constructed from a union of solute spheres whose radii, $R_{\text{rad},a}$, are parameters of the model. The electrostatic response on the cavity's surface is modeled through discretized Gaussian functions,

$$g_t(\mathbf{r} - \mathbf{R}_t; \zeta_t) = (\zeta_t/\pi)^{3/2} e^{-\zeta_t |\mathbf{r} - \mathbf{R}_t|^2} \tag{4.4}$$

such that the solvent's charge density response at the surface is

$$\sigma(\mathbf{r}) \approx \sum_t s_t g_t(\mathbf{r} - \mathbf{R}_t; \zeta_t), \tag{4.5}$$

where $\mathbf{s}$ is the vector of solvent response coefficients, s_t, that remain to be determined. In this notation, t indexes a discretized point on the surface and

$$\mathbf{R}_t = \mathbf{R}_a + R_{\mathrm{rad},a}\hat{\mathbf{R}}_{\mathrm{q},t} \tag{4.6}$$

is the location of that point. In other words, each point t is tethered to one atom $t \in a$ separated by $R_{\mathrm{rad},a}$ in the direction $\hat{\mathbf{R}}_{\mathrm{q},t}$, where $\hat{\mathbf{R}}_{\mathrm{q},t}$ is a Lebedev quadrature point associated with the Lebedev quadrature weight $w_{\mathrm{q},t}$ and the number of discretized points on a sphere is $N_{\mathrm{q},a}$.[277, 278]

The solvation of free energy of a compound can be modeled as an adiabatic energy of that compound in the presence of SCOSMO implicit

$$E_{\mathrm{aq}} \equiv E_{\mathrm{gas}}(\mathbf{R}, \mathbf{P}) + E_s(\mathbf{s}, \mathbf{q}), \tag{4.7}$$

where

$$\begin{aligned} E_s(\mathbf{s}, \mathbf{q}) =& \frac{1}{2f(\epsilon)}\mathbf{s}^T \cdot \mathbf{A} \cdot \mathbf{s} + \\ & \sum_{a,t\in a} \gamma w_{\mathrm{q},t} p_t R^2_{\mathrm{rad},a} + \mathbf{s}^T \cdot \mathbf{B} \cdot \mathbf{q} \end{aligned} \tag{4.8}$$

and

$$f(\epsilon) = \frac{\epsilon - 1}{\epsilon} \tag{4.9}$$

is the scaled deviation of the dielectric medium of the implicit solvent from an ideal conductor. Throughout this work, the dielectric $\epsilon = 78.4$ (unitless) is used for water. The scaled self-interaction of the polarized surface charges of residing upon the cavitation

barrier, **A**, takes the form

$$A_{tt'} = \int\int \frac{g_t(\mathbf{r}-\mathbf{R}_t;\zeta_t)g_{t'}(\mathbf{r}'-\mathbf{R}_{t'};\zeta_{t'})}{|\mathbf{r}-\mathbf{r}'|} d^3r d^3r' \times \begin{cases} p_t^{-1}, & \text{if } t = t' \\ 1, & \text{otherwise} \end{cases} \tag{4.10}$$

The switching function p_t smoothly transitions between 0 and 1, preventing surface elements from interacting with other surface elements which have been partially or fully removed by the intersection of the cavitation sphere from another particle,

$$p_{t\in a} = \prod_{b\neq a} S_{\text{on}}(R_{bt}^2, R_{\text{in},b}^2, R_{\text{out},b}^2). \tag{4.11}$$

When $p_t = 0$ dimensionality of the arrays can be decreased safely without causing discontinuous changes in the energy or gradients. In this case, $R_{\text{in},a}$ and $R_{\text{out},a}$ are the limits of the switching region, whose explicit definitions can be found in Appendix F and the switching function, S_{on}, takes the form

$$S_{\text{on}}(x, x_{\text{lo}}, x_{\text{hi}}) = \begin{cases} 0 & \text{if } x < x_{\text{lo}} \\ 1 & \text{if } x > x_{\text{hi}} \\ 1 - 10u^3 - 15u^4 + 6u^5 & \text{otherwise} \\ u \equiv (x_{\text{hi}} - x)/(x_{\text{hi}} - x_{\text{lo}}) & \end{cases} \tag{4.12}$$

Continuing, γ is the 'surface tension' of the cavity which accounts for the free energy cost associated with the formation of the cavity. This term is treated as a parameter for optimization in the current protocols. **B** is the interaction between the previously defined surface charges and the solute charge distribution, **q**. Traditionally, when interacting with a point-charge distribution, this term would be represented as

$$B_{ta} = \frac{\text{erf}(\sqrt{\zeta_t} R_{at})}{R_{at}} \tag{4.13}$$

where

$$\zeta_{t\in a} = \frac{1}{w_{\mathrm{q},t}} \left(\frac{\zeta(N_{\mathrm{q},a})}{R_{\mathrm{rad},a}} \right)^2 . \tag{4.14}$$

Values of $\zeta(N_{\mathrm{q},a})$ constants are tabulated in Ref. 279 and are completely determined upon defining the discretization level of each atom, $N_{\mathrm{q},a}$. Special note should be taken that in the present work the interaction of the electrostatic potential with the surface response elements has been modified. The standard monopole representation is replaced with an auxiliary set of atomic multipole moments, constructed from the underlying DFTB3 density matrix. Further details will be discussed in the next section, Sec 4.2.2.

The response coefficients, s_t, are determined by minimizing

$$\delta \left\{ E_s(\mathbf{s}, \mathbf{q}) - \lambda \left(\sum_t s_t + f(\epsilon) Q \right) \right\} = 0 \tag{4.15}$$

where

$$Q = \sum_a q_a \tag{4.16}$$

is the total solute charge, λ is the Lagrange multiplier used to enforce the constraint

$$\int \sigma(\mathbf{r}) d^3 r = -f(\epsilon) Q \tag{4.17}$$

and q_a is the atomic charge of atom a. Following this constraint, one finds

$$\mathbf{s} = f(\epsilon) \left(\lambda \mathbf{A}^{-1} \cdot \mathbf{v} - \mathbf{A}^{-1} \cdot \mathbf{B} \cdot \mathbf{q} \right) \tag{4.18}$$

and

$$\lambda = \frac{\mathbf{v}^T \cdot \mathbf{A}^{-1} \cdot \mathbf{B} \cdot \mathbf{q} - Q}{\mathbf{v}^T \cdot \mathbf{A}^{-1} \cdot \mathbf{v}} \tag{4.19}$$

where $v_t = 1 \, \forall t$.

The SCOSMO model variationally enters the solution for the molecular orbital coefficients through the Fock matrix

$$\begin{aligned} F^{\sigma}_{\mu\nu} &= \left. \frac{\partial E_{\mathrm{aq}}}{\partial P^{\sigma}_{\mu\nu}} \right|_{\mathbf{R}} \\ &= \left. \frac{\partial E_{\mathrm{gas}}}{\partial P^{\sigma}_{\mu\nu}} \right|_{\mathbf{R}} + \sum_a \left. \frac{\partial E_s}{\partial q_a} \right|_{\mathbf{R},\mathbf{s}} \left. \frac{\partial q_a}{\partial P_{\mu\nu}} \right|_{\mathbf{R}} \end{aligned} \tag{4.20}$$

Upon reaching self-consistent convergence, the molecular orbitals satisfy the generalized eigenvalue equation

$$\mathbf{F}^{\sigma} \cdot \mathbf{C}^{\sigma} = \mathbf{S} \cdot \mathbf{C}^{\sigma} \cdot \mathbf{E}^{\sigma} \tag{4.21}$$

where $\mathbf{S}$ is the atomic orbital overlap matrix

$$S_{\mu\nu} = \int \chi_{\mu}(\mathbf{r}) \chi_{\nu}(\mathbf{r}) d^3 r \tag{4.22}$$

and the Cartesian gradient X_a of atom a becomes

$$\begin{aligned} \frac{dE_{\mathrm{aq}}}{dX_a} = \left.\frac{\partial E_{\mathrm{gas}}}{\partial X_a}\right|_{\mathbf{P}} - \sum_{\mu\nu} Q_{\mu\nu} \frac{dS_{\mu\nu}}{dX_a} \\ + \left.\frac{\partial E_s}{\partial X_a}\right|_{\mathbf{s},\mathbf{q}} + \sum_b \left.\frac{\partial E_s}{\partial q_b}\right|_{\mathbf{R},\mathbf{s}} \left.\frac{\partial q_b}{\partial X_a}\right|_{\mathbf{P}} \end{aligned} \tag{4.23}$$

where

$$Q_{\mu\nu} = \sum_{\sigma \in (\alpha,\beta)} \sum_i n_i^{\sigma} E_{ii}^{\sigma} C_{\mu i}^{\sigma} C_{\nu i}^{\sigma} \tag{4.24}$$

and E_{ii}^{σ} is a spin-resolved molecular orbital eigenvalue.

4.2.2 Correction to the Solute Charge Density Interaction

The functional form of the DFTB3/3OB energy happens to use Mulliken charges to interact the solute atoms with the other solute, atoms[274] however, previous work has found that the interaction of DFTB3 molecules with an "external environment" can be improved by choosing a second charge representation.[245] This approach has been adopted in the present work. In essence, a second set of atomic charges and higher-order atomic multipoles are chosen to interact with the surrounding implicit solute environment while leaving the DFTB3 solute-solute interactions unchanged from their original form. The solute charge density can be approximated by a sum of atomic point

multipoles

$$\begin{aligned} q(\mathbf{r}) &= \sum_a Z_a \delta(\mathbf{r} - \mathbf{R}_a) - \sum_{\mu\nu} P_{\mu\nu} \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) \\ &\approx \sum_{a, lm \in a} q_{a,lm} \frac{C_{lm}(\nabla_a)}{(2l-1)!!} \delta(\mathbf{r} - \mathbf{R}_a) \end{aligned} \tag{4.25}$$

where Z_a is the nuclear core charge of atom a, $q_{a,lm}$ is a multipole moment, $\mathbf{R}_a$ is vector of atomic positions, ∇_a is the Cartesian gradient operator acting on the coordinates of a, and $C_{lm}(\nabla)$ is a spherical tensor gradient operator which is constructed by replacing the Cartesian coordinate arguments of the real-valued regular solid harmonic with their corresponding derivative operators. In this notation,

$$\chi_\mu(\mathbf{r}) \equiv \chi_\mu(r) Y_{l_\mu m_\mu}(\Omega), \tag{4.26}$$

where $Y_{lm}(\Omega)$ is a real-valued spherical harmonic. Following Refs. 245 and 178, the atomic charges are a biased Mulliken partitioning of the density matrix and the higher-order multipole moments are constructed from the one-center blocks of the density matrix:

$$q_{a,lm} = \begin{cases} Z_a - \frac{b_{aa}}{2} - \sum_{b \neq a} f_{ab}(b_{ab}) b_{ab} & \text{if } l = 0 \\ \sum_{\mu\nu \in a} P_{\mu\nu} M^{(l)}_{\mu\nu} \sqrt{\frac{4\pi}{2l+1}} \times & \text{if } l > 0 \\ \int Y_{lm}(\Omega) Y_{l_\mu m_\mu}(\Omega) Y_{l_\nu m_\nu}(\Omega) d\Omega & \end{cases} \tag{4.27}$$

where b_{ab} is a Mulliken bond order,

$$b_{ab} = 2 \sum_{\substack{\mu \in a \\ \nu \in b}} P_{\mu\nu} S_{\mu\nu}, \tag{4.28}$$

$S_{\mu\nu}$ is the atomic orbital overlap matrix [Eq. (4.22)] and f_{ab} is a "charge bias." This bias is constructed such that at $f_{ab} = 1/2$ a Mulliken charge decomposition is produced, but it is generalized to switch from f^{s}_{ab} to f^{d}_{ab} as the Mulliken bond order changes from single bond character, b^{s}_{ab}, to double bond character, b^{d}_{ab},

$$f_{ab}(b_{ab}) = f^{\mathrm{s}}_{ab} + S_{\mathrm{on}}(b_{ab}, b^{\mathrm{s}}_{ab}, b^{\mathrm{d}}_{ab})(f^{\mathrm{d}}_{ab} - f^{\mathrm{s}}_{ab}). \tag{4.29}$$

$S_{\text{on}}(x, x_{\text{lo}}, x_{\text{hi}})$ is described in Eq. (E.2) and $M_{\mu\nu}^{(l)}$ are the one-center radial integrals

$$M_{\mu\nu}^{(l)} = \int_0^\infty \chi_\mu(r)\chi_\nu(r)r^{2+l}dr \tag{4.30}$$

which are treated as parameters to empirically improve molecular electrostatic potentials. The parameters of this model; f_{ab}^{s},f_{ab}^{d},b_{ab}^{s},b_{ab}^{d}, and $M_{\mu\nu}^{(l)}$; are listed in the supporting information of Ref. 178, and are not modified in this work. The effects of the inclusion of this correction can be seen in Figure 4.2 for butanoic acid, showing how the use of multipoles can improve the description of sp^3 oxygen lone pairs. Previous work has noted improvements to the description of sp^2 carbon bonds, and sp^3 sulfur and sp^2 nitrogen lone pairs.

The inclusion of this correction into the SCOSMO framework only modifies one term in the formalism, the interaction of the surface response elements with the solute charge distribution, Eq. (4.13). It would now take the form

$$\begin{aligned} B_{ta} =& \frac{C_{l_a m_a}(\nabla_a)}{(2l_a-1)!!} \\ & \times \int\int \frac{g_t(\mathbf{r}-\mathbf{R}_t;\zeta_t)\delta(\mathbf{r}'-\mathbf{R}_a)}{|\mathbf{r}-\mathbf{r}'|} d^3r d^3r' \\ =& \frac{C_{l_a m_a}(\mathbf{R}_{at})}{(2l_a-1)!!}\left(\frac{d}{dR_{at}^2}\right)^{l_a} \frac{\text{erf}(\sqrt{\zeta_t}R_{at})}{R_{at}}. \end{aligned} \tag{4.31}$$

Additionally, care must be taken when evaluating the Fock matrix derivatives, Eq. (4.20)-(4.23); that is, to make them with respect to $q_{a,lm}$ instead of q_a.

4.2.3 The VRSCOSMO Implicit Solvation Model

The VRSCOSMO model modifies the SCOSMO framework by incorporating charge-dependent behavior into the definition of the solvation radii

$$\begin{aligned} R_{\text{rad},a}(q_{a,00}) =& R_{\text{rad},a} + \alpha(q_{a,00} - q_a^{(0)}) \\ & + \frac{1}{2}\beta(q_{a,00} - q_a^{(0)})^2 \end{aligned} \tag{4.32}$$

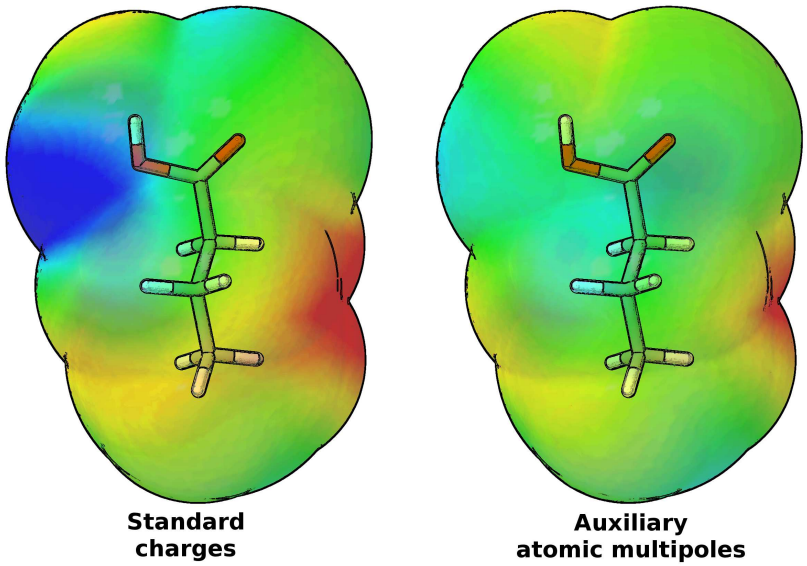


Figure 4.2: **Comparison between electrostatic surface differences generated using DFTB3/3OB with and without the multipole correction with respect to and B3LYP reference.** Standard DFTB3/3OB monopole representation (left) and DFTB3/3OB using an auxiliary set of atomic multipoles (right) evaluated on the solvent accessible surface of butanoic acid. Blue and red indicate that the semiempirical model requires more negative charge and more positive charge relative to B3LYP/6-311++G** reference, respectively. Colors are bounded with a range ±0.003 au. The multipole correction causes the electrostatics to appear more B3LYP-like, without significantly impacting computational efficiency.

where $R_{\mathrm{rad},a}$, α, β, and $q_a^{(0)}$ are parameters. With this charge, the surface response Gaussian positions $\mathbf{R}_t$ and exponents ζ_t [Eq. (4.14)] also become charge-dependent. Upon replacing $R_{\mathrm{rad},a}$ with $R_{\mathrm{rad},a}(q_{a,00})$, the derivatives appearing in Eq. (4.20)-(4.23) require additional chain-rules; that is,

$$\begin{aligned}\left.\frac{\partial E_s}{\partial q_{a,lm}}\right|_{\mathbf{R},\mathbf{s}} &= \left.\frac{\partial E_s}{\partial q_{a,lm}}\right|_{\mathbf{R},\mathbf{s},\zeta,\mathbf{R_t}} \\ &+ \delta_{l0}\delta_{m0}\sum_{t\in a}\left.\frac{\partial E_s}{\partial \zeta_t}\right|_{\mathbf{R},\mathbf{s},\mathbf{R_t}}\frac{d\zeta_t}{dR_{\mathrm{rad},a}}\frac{dR_{\mathrm{rad},a}}{\partial q_{a,00}} \\ &+ \delta_{l0}\delta_{m0}\sum_{t\in a}\left.\frac{\partial E_s}{\partial X_t}\right|_{\mathbf{R},\mathbf{s},\zeta}\frac{dX_t}{dR_{\mathrm{rad},a}}\frac{dR_{\mathrm{rad},a}}{\partial q_{a,00}}.\end{aligned} \tag{4.33}$$

4.2.4 Parameterization Protocol for the VRSCOSMO Solvation Model

Parameters for the SCOSMO and VRSCOSMO models, both utilizing the multipole correction, were optimized for elements H, C, N, O, and P to reproduce the absolute solvation free energy of a subset of the SM6 database[59] consisting of 169 molecules (139 neutral, 16 anionic, 14 cationic). A table listing these molecules can be found in Appendix F. The absolute solvation free energy is approximated by the difference in the adiabatic electronic energy of the molecule in the implicit solvent and gas phase environments upon geometry optimization in its respective state.

$$\Delta G_{\text{solv}} \equiv E_{\text{aq}}(\mathbf{R}_{\text{aq}}, \mathbf{P}_{\text{aq}}) - E_{\text{gas}}(\mathbf{R}_{\text{gas}}, \mathbf{P}_{\text{gas}}) \tag{4.34}$$

The parameterizations of the SCOSMO and VRSCOSMO models were treated independently. Parameters obtained for both models are not constrained to have the same values. Parameter optimization was performed by direction set chi-squared minimization and geometry optimization was performed using the DL-Find software package.[280] Initial parameterization steps where performed with the optimized gas phase geometry, as simultaneous optimization of solvation model parameters and molecular geometry can be computationally prohibitive. However, after initial parameterization, geometries and parameters were allowed to relax interchangeably until convergence was reached. In the final step of optimization, both geometries and parameters were relaxed simultaneously in the neighborhood of the previously obtained values. The optimized parameters are listed in Table 4.1. The training set mean signed error (MSE), mean unsigned error (MUE) and root mean squared error (RMSE) for the SCOSMO and VRSCOSMO models are listed in Table 4.2 and model performance, as compared to experiment, is shown in Figure 4.3.

After parameter optimization, the SCOSMO and VRSCOSMO models were evaluated on a separate test set of molecules, consisting of a total of 65 molecules (34 neutral, 17 anionic, 14 cationic) taken from Ref. 269. A table listing these molecules can be found in Appendix G. These molecules were used in the parameterization of the SCC-DFTBPR implicit solvation model developed by Hou *et al.*, which utilized a charge-dependent non-linear PB model implemented with an earlier variant of the DFTB3 semiempirical method.[281, 282] Comparisons of SCOSMO, VRSCOSMO and

this reference model's performance is tabulated in Table 4.3. Similarly, performance of these models as compared to experimental values can be found in Figure 4.4.

4.2.5 Phosphoryl Transfer Reaction Protocol used with VRSCOSMO

Using the parameters obtained from absolute solvation free energy optimization, the VRSCOSMO model was used to examine various biologically motivated reactions. Specifically, a series of reactions that serve as informative models for reactive RNA systems. The cleavage of the phosphodiester bond in these hydrolysis/transfer reactions can be seen in many catalytic motifs: in the protein RNase A[283, 284] and in the hairpin,[285, 251] hammerhead,[286, 287, 288] *glm*S,[289, 290] hepatitis delta virus[291, 292] ribozymes. An understanding of the fundamental chemical pathway in these reactions can glean further insight into these complex biochemical processes.

Two-dimensional reaction energy surfaces were generated, with reaction coordinates defined as the distance between the attacking and leaving group oxygen to phosphorus center, respectively. Distances were constrained to the value of each reaction coordinate and the system was allowed to relax using the DL-Find geometry optimization software suite.[280] After which, energy landscapes were constructed using two-dimensional spline interpolation. Stationary point calculations were performed with no restraints to obtain reaction barriers and product state calculations were completed at infinite separation.

Two different types of phosphoryl reactions were modeled: the phosphate hydrolysis reaction of trimethylphosphate (TMP) and a series of phosphoryl transesterification transfer reactions corresponding to the self-attack of the O2' oxygen on the phosphate center in methyloxyribose (MOR) with seven different leaving groups [HO^-, CH_3O^-, $CH_3CH_2O^-$, $CH_3CH_2CH_2O^-$, $(CH_3)_2CHO^-$, CH_3COO^- and PhO^-]. The TMP PES is shown in Figure 4.5 and the MOR PESs are shown in Figure 4.7. Table 4.4 compares the modeled free energy barriers of the MOR reactions to previously calculated and experimental values and Figure 4.6 displays geometric information for the MOR transition states.

4.3 Results and Discussion

This section is divided into two distinct parts; the first discusses the parameterization and performance of the VRSCOSMO model in reference frame of accurate calculation of the solvation free energy data and the second discusses the performance of DFTB3/3OB in conjunction with the VRSCOSMO solvation model using the multipole long range electrostatic correction in capturing the reaction barriers of methyloxyribose transesterification and hydrolysis reactions by analyzing calculated reaction surfaces and comparing them to previously calculated and experimental data.

4.3.1 VRSCOSMO Parameterization and Modeled Solvation Free Energy Performance

Initial tests performed during the preparation of this manuscript found that the standard monopole representation of the DFTB3/3OB electrostatic potential encountered difficulties in distinguishing different functional groups. Large, systematic errors between in the absolute solvation free energy for molecules containing oxygen atoms with different numbers of lone pairs, such as when comparing alcohol and acid functional groups. Additionally, conjugated rings were frequently found to produce an electrostatic potential where at the ring center the potential too positive when compared to a B3LYP reference. The implicit solvent model response is caused by the electrostatic potential of the solute, therefore, an auxiliary set atomic multipole moments from the DFTB3 density matrix was constructed using the prescription developed in Ref. 245. The inclusion of these multipole moments lead to the alleviation or complete elimination of these systematic errors, thus informing the decision to abandon parameterization efforts using the conventional DFTB3 monopole approximation.

Overall, the VRSCOSMO model performs well for a vast majority of the molecules in the training set, shown in Table 4.2 and Figure 4.3. The SCOSMO non-charge-dependent and VRSCOSMO charge-dependent parameterizations show similar performance for neutral molecules, having MUEs of approximately 1 kcal/mol. However, the VRSCOSMO model presents MUE 2-3 kcal/mol lower for the charged species. For both models, the largest errors occur in cationic molecules containing oxygen atoms whose sum of net bond orders is 3. That is, molecules containing an “additionally protonated”

Element	R_0	q_0	α	β
C	3.75	-0.0339	-0.2493	-0.000
H	2.20	0.0954	-0.0502	-0.1794
O	3.19	-0.7056	0.0179	-0.4350
N	3.17	-0.2450	0.0128	-0.000
P	5.30	-0.2000	–	–

Table 4.1: **Optimized parameters for the VRSCOSMO implicit solvation model for use with DFTB3/3OB with the OPhyd specific reaction parameterization and multipole long-range electrostatic correction.** Optimized parameters used in the calculation of modeled solvation free energy and the phosphoryl transfer reaction data reproted here. Radii are reported in Bohr and charge are in $-e$.

alcohol compounds or molecules containing a protonated oxygen which also contains a double bond. These compounds are systematically undersolvated by approximately 10 kcal/mol, which is likely due to the over-simplification of proton chemical bonding in the absence of explicit solvent. Better estimations of solvation free energies could likely be obtained if explicit water molecules were included as part of the solute. Due to this logic, these types of compounds were assigned a smaller weight during chi-squared minimization as to not bias the parameterization into over-correcting these errors. For comparison purposes, additional solvation free energy data for the more stable cations, bereft of these compounds, has been added to Tables 4.2 and 4.3 labeled with an asterisk.

After parameter optimization the SCOSMO and VRSCOSMO models were evaluated against a separate test set of molecules. Absolute solvation data from these calculations can be seen in Table 4.3 and Figure 4.4. These results are also compared to those generated by Hou *et al.*[269], whom developed a similar charge-dependent nonlinear PB model, SCC-DFTBPR. The molecules within this test set were used as part of the SCC-DFTBPR training set and do not appear within this manuscript's training set. Again, the largest errors occur for the cationic compounds containing unusual oxygen protonation states. As expected, the SCC-DFTBPR model performs better for these unusual cases; a significant number of these compounds appeared in the SCC-DFTBPR training set thus the parameterization is better conditioned for these cases. However if these compounds where removed, the MSE and MUE of the test set for VRSCOSMO would be 0.3 kcal/mol and 1.6 kcal/mol, respectively while the SCC-DFTBPR model would report MSEs and MUEs of 0.1 kcal/mol and 2.1 kcal/mol, respectively. Both VRSCOSMO and SCC-DFTBPR perform significantly better than the SCOSMO parameterization.

	VRSCOSMO			SCOSMO		
Type	MSE	MUE	RMSE	MSE	MUE	RMSE
neutral	-0.0	1.4	1.9	0.4	1.2	1.7
anion	0.8	2.5	3.0	3.9	5.6	6.3
cation	1.4	2.8	4.3	0.7	4.6	5.7
cation*	-0.1	1.5	1.9	-1.0	3.5	3.9
total	0.2	1.6	2.4	0.8	2.0	3.0

Table 4.2: **Modeled solvation free energies with VRSCOSMO and SCOSMO using DFTB3/3OB for the training set.** A breakdown of the compounds is shown here between neutral, anionic and cationic species. The additionally protonated oxygen compounds (those resulting in oxygens with three bonds) are systemtically erronious by approximately 10 kcal/mol for these models and are excluded from the data in the cation* row for comparision purposes and is not included into the the 'total' error analysis. Modeled solvation free energy errors are reported in kcal/mol.

	VRSCOSMO		SCOSMO		SCC-DFTBPR[269]	
Type	MSE	MUE	MSE	MUE	MSE	MUE
neutral	0.4	0.9	0.6	1.0	0.2	1.6
anion	0.2	3.2	5.6	6.6	0.7	3.1
cation	3.5	4.6	3.2	6.1	1.1	3.4
cation*	0.0	1.6	-0.2	3.9	-1.4	1.8
total	1.1	2.3	2.4	3.6	0.5	2.4

Table 4.3: **Modeled solvation free energies with VRSCOSMO and SCOSMO using DFTB3/3OB for the test set.** A breakdown of the compounds is shown here between neutral, anionic and cationic species. Similar for the training set data , the cation* is shown for comparion and is not included into the the 'total' error analysis. However, if these compounds where removed the MSE and MUE for VRSCOSMO for this set would be 0.3 kcal/mol and 1.6 kcal/mol, respectively. Modeled solvation free energy errors are reported in kcal/mol.

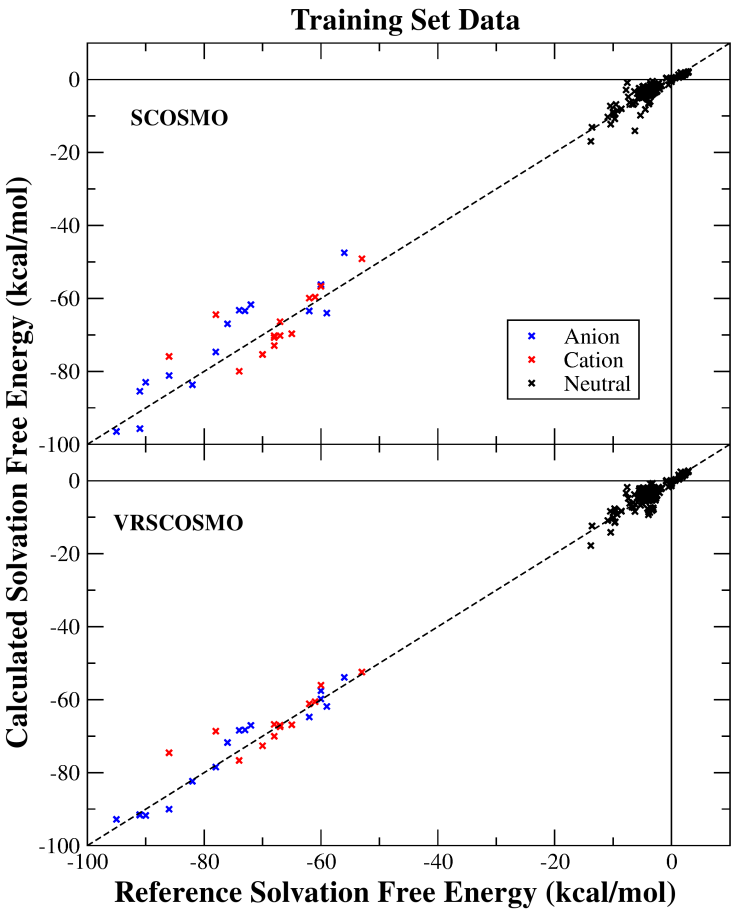


Figure 4.3: **Comparison between modeled and reference solvation free energies for the training set for the SCOSMO and VRSCOSMO implicit solvation models with DFTB3/3OB for the training set.** Both the SCOSMO and VRSCOSMO modeled solvation free energies are ploted against the reference solvation free energies for comparison, with differentiation for charged and neutral species. Calculations were performed using the DFTB3/3OB Hamiltonian[268] with a multipole long-range electrostatic correction [177]. All values are reported in kcal/mol.

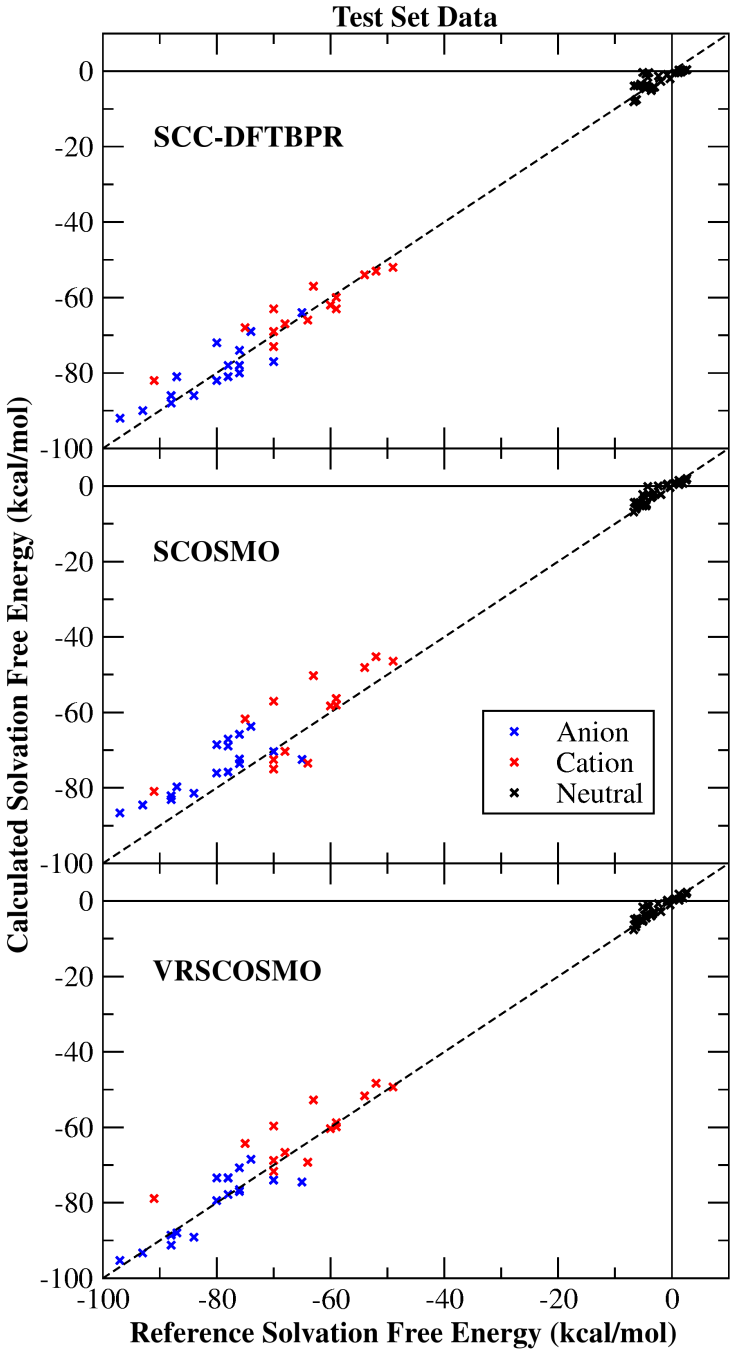


Figure 4.4: **Comparison between modeled and reference solvation free energies for the training set for the SCOSMO model , VRSCOSMO model and a literacture reference for the test set.** All three (Hou, SCOSMO and VRSCOSMO) modeled solvation free energies are ploted against the reference solvation free energies for comparison, with differentiation for charged and neutral species. It should be noted that Hou's model[269] were performed with the SCC-DFTB Hamiltonian[293] while the SCOSMO and VRSCOSMO calculations are performed with the DFTB3/3OB[268] Hamiltonian. All values are reported in kcal/mol.

4.3.2 VRSCOSMO Performance Capturing Phosphoryl Transfer Reaction Pathways

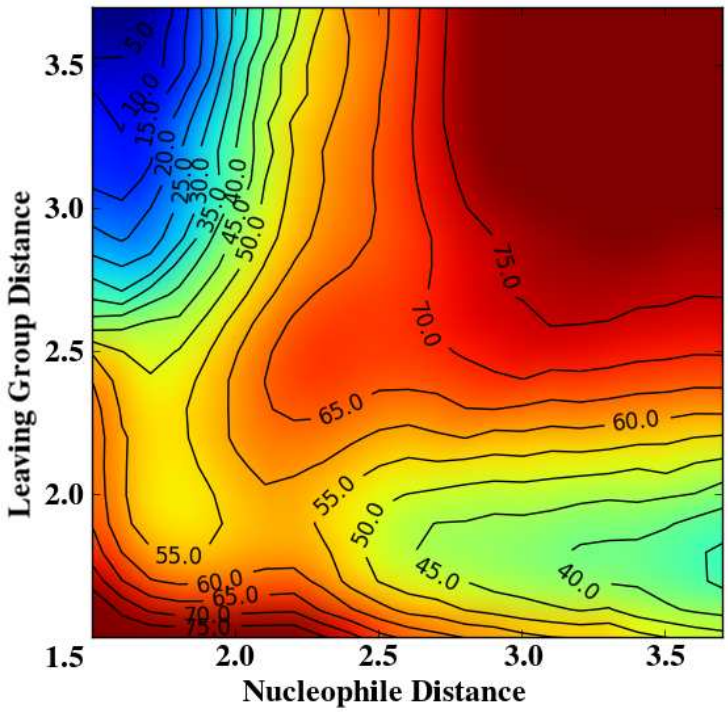


Figure 4.5: **Reaction surface for the hydrolysis reactions of trimethylphosphate using VRSCOSMO with DFTB3/3OB.** The reaction surface for TMP hydrolysis with the DFTB3/3OB Hamiltonian while under the influence of the VRSCOSMO charge-dependent solvation model with long-range electrostatic corrections. Reaction shows an early transition state, as noted by the relatively short leaving group bond distance (the y-axis) as the reaction proceeds through the reaction coordinate at the transition state. The reaction barrier obtained from the surface, and stationary point calculations, is consistent with previous studies and experiment at 23.19 kcal/mol. Distances and energies are reported in Å and kcal/mol, respectively.

Two different types of phosphoryl reactions were used to test the phosphate hydrolysis applicability of the DFTB3+VRSCOSMO framework to RNA-like systems, a phosphate hydrolysis reactions and a series phosphoryl transesterification transfer reactions with different leaving groups. These reactions have a large amount of local charge transfer, totaling a net $2e$ charge, which makes them ideal candidates for use with a charge-dependent model with an aim of accurately modeling biological reactions. These

	Calculated		Experiment	
Model/Ref.	VRSCOSMO	PCM[294]	Expt. I[295]	Expt. II[296]
HO^-	24.90	–	–	–
CH_3O^-	28.10	24.43	27.09	–
$CH_3CH_2O^-$	26.60	25.84	28.55	23.37
$CH_3CH_2CH_2O^-$	27.38	24.26	–	–
$(CH_3)_2CHO^-$	28.87	25.96	30.76	26.02
CH_3COO^-	8.19	11.91	–	–
PhO^-	22.72	14.76	22.62	–

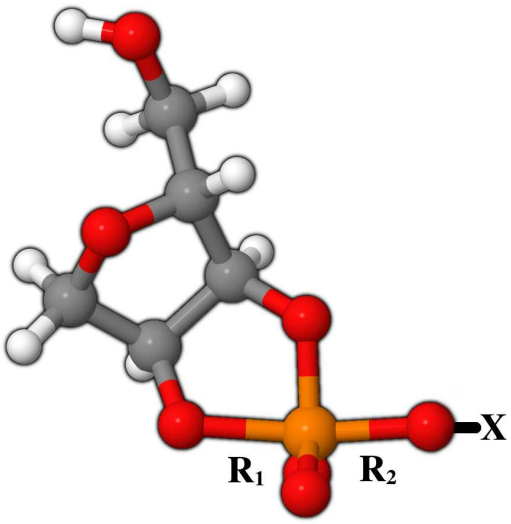
Table 4.4: **Modeled and experimental reaction barriers for the self-attack of MOR with different leaving groups.** Barriers are reported in kcal/mol. These methyloxyribose transesterification reactions serve as model systems for RNA cleavage reactions and have a large amount of localized charge which is transfered from atom to atom over the coarse of the reaction. PCM calculations are performed with the UFF radii at the M06-2X/6-311++G(3df,2p)//6-311++G(d,p) level of theory[294]. Expt. I is work completed by Brown on 2-hyroxypropyl phosphate[295] and Expt. II was completed by Kosonen by examining various esters of uridine 3-phosphate

reactions in particular serve as analogs for a key step in the self-cleavage of the RNA backbone seen in many ribozymes.

A two-dimensional reaction energy surface for hydrolysis of TMP can be seen in Figure 4.5. The energy barrier is found to be 23.19 kcal/mol which is in good agreement with the experimentally known value of 24.6 kcal/mol.[297] The model predicts that the transition state for this reaction appears to be quite "early" in the reaction coordinate, having formed be the P–O attacking bond well before the cleavage of the leaving group bond. This observation is consistent with previous computational studies of this system.[269] Given that the Hamiltonian used in this calculation is specifically parameterized to accurately model gas-phase phosphate hydrolysis reactions, this result indicates that the VRSCOSMO parameterization is not grossly misrepresenting the solvent environment.

Similarly, DFTB3+VRSCOSMO reaction surfaces for the phosphoryl transesterification reactions of MOR with different leaving groups can be found in Figure 4.7. A summary of the reaction barriers as compared to previously calculated *ab initio* barriers using the PCM with UFF radii and experimentally known barriers can be found

in Table 4.4. Additionally, Figure 4.6 details various transition state geometries for the modeled reactions as well as listing non-rate-limiting reaction barriers for those reactions where two separate transition states were observed. VRSCOSMO attack barriers compare favorably to those calculated with the M06-2X/6-311++G(3df,2p)//M06-2X/6-311++G(d,p) PCM model[294] as well as to the known experiments. It should be noted, however, that the experiments presented here examine reactions with sightly different chemical structures than those which have been computed in this manuscript. Specifically, Expt. I[295] performed experiments with 2-hyroxypropyl phosphate while Expt. II[296] examined various esters of uridine 3'-phosphate. However, despite these structural differences, it is thought that the reaction barriers should not be significantly different and general trends of the data should remain fairly consistent.



	TS 1			TS 2			Exp.
Leaving Group	R_1	R_2	$\Delta G^{\ddagger}$	R_1	R_2	$\Delta G^{\ddagger}$	pK_a
HO^-	2.05	1.72	23.6	1.75	2.00	24.9	15.7
CH_3O^-	2.02	1.73	26.0	1.75	2.07	28.1	15.5
$CH_3CH_2O^-$	2.02	1.72	23.8	1.74	2.08	26.6	16.0
$CH_3CH_2CH_2O^-$	2.00	1.73	24.9	1.75	2.06	27.4	16.1
$(CH_3)_2CHO^-$	2.01	1.72	25.3	1.75	2.08	28.9	17.1
CH_3COO^-	2.60	2.16	8.2	–	–	–	4.46
PhO^-	2.12	1.77	22.7	–	–	–	9.95

Figure 4.6: **Example of the model phosphoryl transesterification system listed with the different leaving groups studied and transition state data.** Shown is the methyloxyribose ring, which undergoes an internal nucleophilic attack to model the backbone cleavage in RNA. Also shown is the attacking and leaving oxygen bond lengths, R_1 and R_2 respectively, as modeled by DFTB3/3OB with the multipole correction while under the influence of VRSCOSMO for each different leaving group, X-O^-. TS 1 and TS 2 refer to the early and late transition states which can be observed for those reactions, respectively. Experimental pK_a values obtained from IUPAC chemical data series No. 23 are provided for the leaving groups as to indicate which reactions may favor early/late transitions.

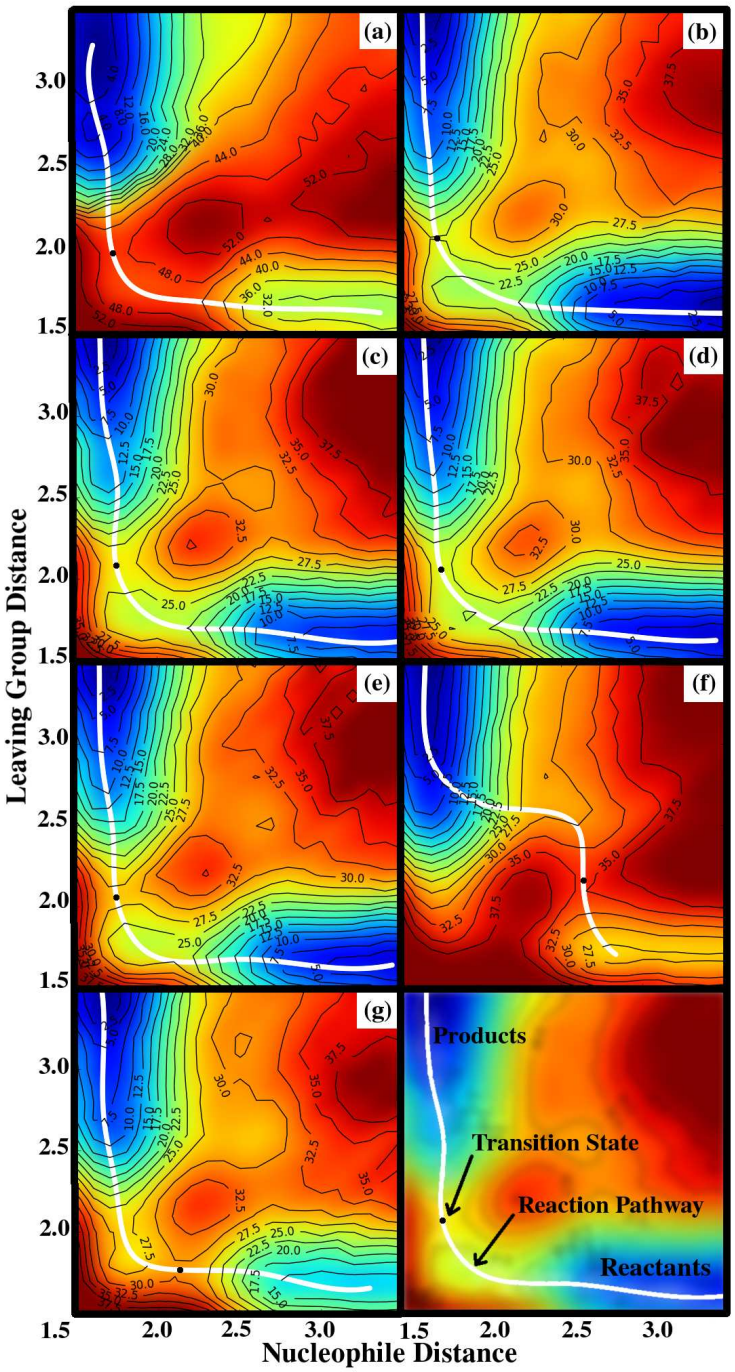


Figure 4.7: **Reaction surfaces for the transesterification phosphoryl transfer reactions studied with VRSCOSMO and DFTB3/3OB with various leaving groups.** All surfaces are calculated while under the influence of the VRSCOSMO charge-dependent solvation model using DFTB3 with corrected long-range electrostatics. he order presented, from left to right and top to bottom is: HO^-, CH_3O^-, $CH_3CH_2O^-$, $CH_3CH_2CH_2O^-$, $(CH_3)_2CHO^-$, CH_3COO^- and PhO^-. Energies are reported in kcal/mol and distances are in Å. The white line on each plot is indicative of the reaction pathway as determined by nudged elastic band minimization. Special note should be taken of the difference in scale for the HO^- plot.

4.4 Conclusion

A charge-dependent variant of the SCOSMO implicit solvation model has been presented. The model, VRSCOSMO, is parameterized using the DFTB3/3OB-OPhyd semiempirical Hamiltonian in conjunction with an auxiliary set of atomic multipoles to interact the DFTB3 solute with the implicit solvent model. SCOSMO and VRSCOSMO modeled solvation energies are provided for both a test and a training set. The two methods produce similar results for neutral molecules, however, the VRSCOSMO model is shown to provide an 2-3 kcal/mol improvement in MUE for charged species. Further comparisons are made to the SCC-DFTBPR charge-dependent nonlinear PB implicit solvation model[269] for test set molecules, with VRSCOSMO showing similar accuracy. DFTB3+VRSCOSMO applications are performed on phosphoryl hydrolysis and phosphoryl transesterification reactions. Reaction barriers are shown to agree well with available experimental data and previously performed *ab initio* calculations. The VRSCOSMO implicit solvation model serves as a efficient technique for probing biochemical reaction landscapes to gain understanding in the underlying mechanistic pathways. Advantages levied by the model include accurately recapitulating neutral and ionic absolute solvation free energies, producing smooth potential energy surfaces with cavitation radii that dynamically adjust themselves to changes in the local charge state of atoms over the coarse of a reaction and possessing the analytically smooth framework afforded by the SCOSMO formalism.

4.5 Acknowledgments

The authors are grateful for financial support provided by the National Institutes of Health (GM107485 to DY). Computational resources were provided by the Minnesota Supercomputing Institute for Advanced Computational Research (MSI) and Rutgers, the State university of New Jersey.

References

[1] Jamal M. Buzayan, Wayne L. Gerlach, and George Bruening. Satellite tobacco ringspot virus RNA: A subset of the RNA sequence is sufficient for autolytic processing. *Proc. Natl. Acad. Sci. USA*, 83:8859–8862, 1986.

[2] Gerry A. Prody, John T. Bakos, Jamal M. Buzayan, Irving R. Schneider, and George Bruening. Autolytic processing of dimeric plant virus satellite RNA. *Science*, 231:1577–1580, 1986.

[3] Jamal M. Buzayan, Arnold Hampel, and George Bruening. Nucleotide sequence and newly formed phosphodiester bond of spontaneously ligated satellite tobacco ringspot virus RNA. *Nucleic Acids Res.*, 14:9729–9743, 1986.

[4] Thomas R. Cech, Daniel Herschlag, Joseph A. Piccirilli, and Anna Marie Pylet. RNA Catalysis by a Group I Ribozyme. *J. Biol. Chem.*, 267(25):17479–17482, 1992.

[5] Thomas R. Cech. Self-Splicing of Group I Introns. *Annu. Rev. Biochem.*, 59:543–568, 1990.

[6] T. R. Cech and O. C. Uhlenbeck. Hammerhead nailed down. *Nature*, 372:39–40, 1994.

[7] Jaime H. Cate, Anne R. Gooding, Elaine Podell, Kaihong Zhou, Barbara L. Golden, Alexander A. Szewczak, Craig E. Kundrot, T. R. Cech, and J. A. Dounda. RNA Tertiary Structure Mediation by Adenosine Platforms. *Science*, 273:1696–1699, 1996.

[8] Thomas R. Cech. RNA finds a simpler way. *Nature*, 428:263–264, 2004.

[9] Thomas R. Cech, Daniel Herschlag, Joseph A. Piccirilli, and Anna Marie Pylet. RNA Catalysis by a Group I Ribozyme. *J. Biol. Chem.*, 262(25):17479–17482, 1992.

[10] Peter B. Moore. The Three-Dimentional Structure of the Ribosome and its Components. *Annu. Rev. Biophys. Biomol. Struct.*, 27:355–379, 1998.

[11] Jamie H. Cate, Marat M. Yusupov, Gulnara Zh. Yusupova, Thomas N. Earnest, and Harry F. Noller. X-ray Crystal Structures of 70S Ribosome Functional Complexes. *Science*, 285:2095–2104, 1999.

[12] Brian T. Wimberly, Ditlev E. Brodersen, William M., Jr. Clemons, Robert J. Morgan-Warren, Andrew P. Carter, Clemens Vonrhein, Thomas Hartschk, and V. Ramakrishnan. Structure of the 30S ribosomal subunit. *Nature*, 407:327–339, 2000.

[13] Nenad Ban, Poul Nissen, Jeffrey Hansen, Peter B. Moore, and Thomas A. Steitz. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Resolution. *Science*, 289:905–920, 2000.

[14] Poul Nissen, Jeffrey Hansen, Nenad Ban, Peter B. Moore, and Thomas A. Steitz. The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. *Science*, 289:920–930, 2000.

[15] Dipali G. Sashital and Samuel E. Butcher. *Ribozymes and RNA Catalysis*, chapter Is the spliceosome a ribozyme?, pages 253–269. RSC Biomolecular Series. RSC Publishing, Cambridge, 2008.

[16] Saba Valadkhan. The spliceosome: a ribozyme at heart? *Biol. Chem.*, 388:693–697, 2007.

[17] Catherine A. Collins and Christine Guthrie. The question remains: Is the spliceosome a ribozyme? *Nature Struct. Biol.*, 7:850–854, 2000.

[18] Walter Gilbert. The RNA World. *Nature*, 319:618, 1986.

[19] F. H. Crick. The origin of the genetic code. *J. Mol. Biol.*, 38(3):367–379, Dec 1968.

[20] Alessio Peracchi, Leonid Beigelman, Nassim Usman, and Daniel Herschlag. Rescue of abasic hammerhead ribozymes by exogenous addition of specific bases. *Proc. Natl. Acad. Sci. USA*, 93:11522–11527, 1996.

[21] Giulia Talini, Sergio Branciamore, and Enzo Gallori. Ribozymes: Flexible molecular devices at work. *Biochimie*, 93:1998–2005, 2011.

[22] Peter B. Rupert, Archna P. Massey, Snorri Th. Sigurdsson, and Adrian R. Ferré-D'Amaré. Transition State Stabilization by a Catalytic RNA. *Science*, 298:1421–1424, 2002.

[23] Philip C. Bevilacqua. Mechanistic considerations for general acid-base catalysis by RNA: Revisiting the mechanism of the hairpin ribozyme. *Biochemistry*, 42:2259–2265, 2003.

[24] Kwangho Nam, Jiali Gao, and Darrin M. York. Quantum mechanical/molecular mechanical simulation study of the mechanism of hairpin ribozyme catalysis. *J. Am. Chem. Soc.*, 130(14):4680–4691, 2008.

[25] Donald H. Andrews. The Relations Between the Raman Spectra and the Structure of Organic Molecules. *Phys. Rev.*, 36:544–554, 1930.

[26] Hans Matrin Senn and Walter Thiel. QM/MM Methods for Biological Systems. *Top. Curr. Chem.*, 268:173–290, 2007.

[27] Christopher J. Cramer. *Essentials of Computational Chemistry: Theories and Models*. John Wiley & Sons, Chichester, England, 2nd edition, 2004.

[28] Pavel Hobza, Martin Kabeláč, Jiří Šponer, Petr Mejzkík, and Jiří Vondrášek. Performance of emperical potentials (AMBER, CFF95, CVFF, CHARMM, OPLS, POLTEV), semiempirical quantum chemical methods (AM1, MNDO/M, PM3), and ab initio Hartree-Fock method for interaction of DNA bases: Comparison with nonemperical beyond Hartree-Fock results. *J. Comput. Chem.*, 18(9):1136–1150, 1997.

[29] Nadine Homeyer, Anselm H. C. Horn, Harald Lanig, and Heinrich Sticht. AMBER force-field parameters for phosphorylated amino acids in different protonation

states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. *J. Mol. Model.*, 12:281–289, 2006.

[30] Raviprasad Aduri, Brian T. Psciuk, Pirro Saro, Hariprakash Taniga, H. Bernhard Schlegel, and Jr. John SantaLucia. AMBER Force Field Parameters for the Naturally Occurring Modified Nucleosides in RNA. *J. Chem. Theory Comput.*, 3:1464–1475, 2007.

[31] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general amber force field. *J. Comput. Chem.*, 25:1157–1174, 2004.

[32] Alexander D. MacKerell, Jr. Empirical Force Fields for Biological Macromolecules: Overview and Issues. *J. Comput. Chem.*, 25:1584–1604, 2004.

[33] Da-Wei Li and Rafael Bru?schweiler. Iterative Optimization of Molecular Mechanics Force Fields from NMR Data of Full-Length Proteins. *J. Chem. Theory Comput.*, 7(6):1773–1782, 2011.

[34] Claude Cohen-Tannoudji, Bernard Diu, and Franck Laloë. *Quantum Mechanics*. John Wiley and Sons, New York, 1977.

[35] Attila Szabo and Neils S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Dover Publications, Inc., New York, 1st edition, 1996.

[36] P. Hohenberg and W. Kohn. Inhomogeneous Electron Gas. *Phys. Rev.*, 136:B864–B871, 1964.

[37] Ciro A. Guido, Stefan Knecht, Jacob Kongsted, and Benedetta Mennucci. Benchmarking Time-Dependent Density Functional Theory for Excited State Geometries of Organic Molecules in Gas-Phase and in Solution. *J. Chem. Theory Comput.*, 9:2209–2220, 2013.

[38] Muneaki Kamiya, Takao Tsuneda, and Kimihiko Hirao. A density functional study of van der waals interactions. *J. Chem. Phys.*, 117:6010–6015, 2002.

[39] Georg Schreckenbach and T Ziegler. Calculation of NMR Shielding Tensors Using Gauge-Including Atomic Orbitals and Modern Density Functional Theory. *J. Phys. Chem.*, 99(2):606–611, January 1995.

[40] L. A. Burns, Á. Vázquez-Mayagoitia, B. G. Sumpter, and C. D. Sherrill. Density-Functional Approaches to Noncovalent Interactions: A Comparison of Dispersion Corrections (DFT-D), Exchange-Hole Dipole Moment (XDM) Theory, and Specialized Functionals. *J. Chem. Phys.*, 134:084107–084132, 2011.

[41] Yan Zhao and Donald G. Truhlar. Hybrid Meta Density Functional Theory Methods for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions: The MPW1B95 and MPWB1K Models and Comparative Assessments for Hydrogen Bonding and van der Waals Interactions . *J. Phys. Chem. A*, 108:6908–6918, 2004.

[42] Yan Zhao and Donald G. Truhlar. Density functional for spectroscopy: no long-range self-interaction error, good performance for Rydberg and charge-transfer states, and better performance on average than B3LYP for ground states. *J. Phys. Chem. A*, 110(49):13126–13130, 2006.

[43] Yingkai Zhang, Tai-Sung Lee, and Weitao Yang. A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *J. Chem. Phys.*, 110:46–54, 1999.

[44] Iris Antes and Walter Thiel. Adjusted Connection Atoms for Combined Quantum Mechanical and Molecular Mechanical Methods. *J. Phys. Chem. A*, 103:9290–9295, 1999.

[45] U. C. Singh and P. A. Kollman. A combined ¡I¿ab initio¡/I¿ quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: applications to the CH_3Cl+Cl^- exchange reaction and gas phase protonation of polyethers. *J. Comput. Chem.*, 7:718–730, 1986.

[46] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, 103:227–249, 1976.

[47] Mireia Garcia-Viloca and Jiali Gao. Generalized hybrid orbital for the treatment of boundary atoms in combined quantum mechanical and molecular mechanical calculations using the semiempirical parameterized model 3 method. *Theor. Chem. Acc.*, 111:280–286, 2004.

[48] Jiali Gao, Patricia Amara, Cristobal Alhambra, and Martin J. Field. A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *J. Phys. Chem. A*, 102:4714–4721, 1998.

[49] Jan H Jensen, Hui Li, Andrew D. Robertson, and Pablo A. Molina. Prediction and Rationalization of Protein pKa Values Using QM and QM/MM Methods. *J. Phys. Chem. A*, 109:6634–6643, 2005.

[50] Timothy J. Giese, Maria T. Panteva, Haoyuan Chen, and Darrin M. York. Multipolar Ewald methods, 1: Theory, accuracy, and performance. *J. Chem. Theory Comput.*, 11:436–450, 2015.

[51] Kwangho Nam, Jiali Gao, and Darrin M. York. An efficient linear-scaling Ewald method for long-range electrostatic interactions in combined QM/MM calculations. *J. Chem. Theory Comput.*, 1(1):2–13, 2005.

[52] Darrin M. York, Alexander Wlodawer, Lee G. Pedersen, and Tom Darden. Atomic level accuracy in simulations of protein crystals. *Proc. Natl. Acad. Sci. USA*, 91:8715–8718, 1994.

[53] Timothy R. Forester and William Smith. SHAKE, rattle, and roll: Efficient constraint algorithms for linked rigid bodies. *J. Comput. Chem.*, 19(1):102–111, 1998.

[54] Vincent Kräutler, Wilfred F. Van Gunsteren, and Philippe H. Hünenberger. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J. Comput. Chem.*, 22:501–508, 2001.

[55] Sang-Ho Lee, Kim Palmo, and Samuel Krimm. WIGGLE: A new constrained molecular dynamics algorithm in Cartesian coordinates. *J. Comput. Phys.*, 210:171–182, 2005.

[56] Shuichi Miyamoto and Peter A. Kollman. SETTLE: An analytic version of the SHAKE and RATTLE algorithms for rigid water models. *J. Comput. Chem.*, 13:952–962, 1992.

[57] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.*, 23:327–341, 1977.

[58] K. Sharp and B. Honig. Calculating Total Electrostatic Energies with the Nonlinear Poisson-Boltzmann Equation. *J. Phys. Chem.*, 94:7684–7692, 1990.

[59] Casey P. Kelly, Christopher J. Cramer, and Donald G. Truhlar. SM6: A Density Functional Theory Continuum Solvation Model for Calculating Aqueous Solvation Free Energies of Neutrals, Ions, and Solute-Water Clusters. *J. Chem. Theory Comput.*, 1:1177–1152, 2005.

[60] David J. Geisen, Gregory D. Hawkins, Daniel A. Liotard, Christopher J. Cramer, and Donald G. Truhlar. A universal model for the quantum mechanical calculation of free energies of solvation in non-aqueous solvents. *Thoer. Chem. Acc.*, 98:85–109, 1997.

[61] Christorpher J. Cramer and Donald G. Truhlar. An SCF Solvation Model for the Hydrphobic Effect and Absolute Free Energies of Aqueous Solvation. *Science*, 256:213–217, 1992.

[62] A. Klamt and G. Schüürmann. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perkin Trans. 2*, 2:799–805, 1993.

[63] Andreas Klamt. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.*, 99:2224–2229, 1995.

[64] Andreas Klamt, Frank Eckert, Michael Diedenhofen, and Michael E. Beck. First Principles Calculations of Aqueous pK_a Values for Organic and Inorganic Acids Using COSMO-RS Reveal an Inconsistency in the Slope of the pK_a Scale. *J. Phys. Chem. A*, 107:9380–9386, 2003.

[65] Daniel A. Liotard, Gregory D. Hawkins, Gillian C. Lynch, Christopher J. Cramer, and Donald G. Truhlar. Improved methods for semiempirical solvation models. *J. Comput. Chem.*, 16:422–440, 1995.

[66] Di Qiu, Peter S. Shenkin, Frank P. Hollinger, and W. Clark Still. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A*, 101(16):3005–3014, 1997.

[67] Robert A. Pierotti. A Scaled Particle Theory of Aqueous and Nonaqueous Solutions. *Chem. Rev.*, 76(6):717–726, 1976.

[68] S. Miertuš, E. Scrocco, and J. Tomasi. Electrostatic interaction of a solute with a continuum. a direct utilization of *ab initio* molecular potentials for the prevision of solvent effects. *Chem. Phys.*, 55:117–129, 1981.

[69] Adrian W. Lange and John M. Herbert. A smooth, nonsingular, and faithful discretization scheme for polarizable continuum models: The switching/Gaussian approach. *J. Chem. Phys.*, 133:244111–244128, 2010.

[70] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, and Augusta H. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1092, 1953.

[71] G. M. Torrie and J. P. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.*, 23:187–199, 1977.

[72] Shankar Kumar, Djamal Bouzida, Robert H. Swendsen, Peter A. Kollman, and John M. Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.*, 13:1011–1021, 1992.

[73] Justin R. Gullingsrud, Rosemary Braun, and Klaus Schulten. Reconstructing Potentials of Mean Force through Time Series Analysis of Steered Molecular Dynamics Simulations. *J. Comput. Phys.*, 151:190–211, 1999.

[74] Erik M. Boczko and Charles L. Brooks III. Constant-Temperature Free Energy Surfaces for Physical and Chemical Processes. *J. Phys. Chem.*, 97:4509–4513, 1993.

[75] Michael R. Shirts and John D. Chodera. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.*, 129:124105, 2008.

[76] Himanshu Paliwal and Michael R. Shirts. A benchmark test set for alchemical free energy transformations and its use to quantify error in common free energy methods. *J. Chem. Theory Comput.*, 7:4115–4134, 2011.

[77] Pavel V. Klimovich, Michael R. Shirts, and David L. Mobley. Guidelines for the analysis of free energy calculations. *J. Comput.-Aided Mol. Des.*, 0:0, March 2015.

[78] Tai-Sung Lee, Brian K. Radak, Ming Huang, Kin-Yiu Wong, and Darrin M. York. Roadmaps through free energy landscapes calculated using the multidimensional vFEP approach. *J. Chem. Theory Comput.*, 10:24–34, 2014.

[79] Tai-Sung Lee, Brian K. Radak, Anna Pabis, and Darrin M. York. A new maximum likelihood approach for free energy profile construction from molecular simulations. *J. Chem. Theory Comput.*, 9:153–164, 2013.

[80] Martin Leitgeb, Christian Schröder, and Stefan Boresch. Alchemical free energy calculations and multiple conformational substates. *J. Chem. Phys.*, 122:084109, 2005.

[81] Arnaud Blondel. Ensemble Variance in Free Energy Calculations by Thermodynamic Integration: Theory, Optimal "Alchemical" Path, and Practical Solutions. *J. Comput. Chem.*, 25:985–993, 2004.

[82] Jerry B. Abrams, Lula Rosso, and Mark E. Tuckerman. Efficient and precise solvation free energies via alchemical adiabatic molecular dynamics. *J. Chem. Phys.*, 125:074115, 2006.

[83] T. P. Straatsma and J. A. McCammon. Computational alchemy. *Annu. Rev. Phys. Chem.*, 43:407–435, 1992.

[84] John G. Kirkwood. Statistical mechanics of fluid mixtures. *J. Chem. Phys.*, 3:300–313, 1935.

[85] Mihaly Mezei, S. Swaminathan, and David L. Beveridge. Ab Initio calculation of the Free Energy of Liquid Water. *J. Am. Chem. Soc.*, 100:3255–3256, 1978.

[86] Peter H. Berens, Donald H. J. Mackay, Gary M. White, and Kent R. Wilson. Thermodynamics and quantum corrections from molecular dynamics for liquid water. *J. Chem. Phys.*, 79:2375–2389, 1983.

[87] T. P. Straatsma and H. J.C. Berendsen. Free energy of ionic hydration: Analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. *J. Chem. Phys.*, 89:5876–5886, 1988.

[88] Donald A. McQuarrie. *Statistical Mechanics*. University Science Books, Mill Valley, CA, 1973.

[89] Christophe Chipot, Peter A. Kollman, and David A. Pearlman. Alternative approaches to potential of mean force calculations: Free energy perturbation versus thermodynamic integration. Case study of some representative nonpolar interactions. *J. Comput. Chem.*, 17:1112–1131, 1996.

[90] Gerhard Hummer. Fast-growth thermodynamic integration: Error and efficiency analysis. *J. Chem. Phys.*, 114:7330–7337, 2001.

[91] Tiziana Z. Mordasini and J. Andrew McCammon. Calculations of Relative Hydration Free Energies: A Comparative Study Using Thermodynamic Integration and an Extrapolation Method Based on a Single Reference State. *J. Phys. Chem. B*, 104:360–367, 2000.

[92] Nilanjan Ghosh and Qiang Cui. pKa of Residue 66 in Staphylococal nuclease. I. Insights from QM/MM Simulations with Conventional Sampling. *J. Phys. Chem. B*, 112:8387–8397, 2008.

[93] Volkhard Helms and Rebecca C. Wade. Free energies of Hydration from Thermodynamic Integration: Comparison of Molecular Mechanics Force Fields and Evaluation of Calculation Accuracy. *J. Comput. Chem.*, 18:449–462, 1997.

[94] Hiroaki Gouda, Irwin D. Kuntz, David A. Case, and Peter A. Kollman. Free energy calculations for theophylline binding to an RNA aptamer: Comparison of MM-PBSA and thermodynamic integration methods. *Biopolymers*, 68:16–34, 2003.

[95] Matthew D. Liptak and George C. Shields. Experimentation with different thermodynamic cycles used for pK_a calculations on carboxylic acids using complete basis set and Gaussian-n models combined with CPCM continuum solvation methods. *Int. J. Quantum Chem.*, 85:727–741, 2001.

[96] Vyacheslav S. Bryantsev, Mamadou S. Diallo, and William A. Goddard III. Calculation of Solvation Free Energies of Charged Solutes Using Mixed Cluster/Continuum Models. *J. Phys. Chem. B*, 112:9709–9719, 2008.

[97] Peter A. Kollman. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.*, 93:2395–2417, 1993.

[98] A. K. Soper. The radial distribution functions of water as derived from radiation total scattering experiments: Is there anything we can say for sure? *ISRN Phys. Chem.*, 2013:279463, 2013.

[99] Yves A. Mantz, Bin Chen, and Glenn J. Martyna. Temperature-dependent water structure: Ab initio and empirical model predictions. *Chem. Phys. Lett.*, 405:294–299, 2005.

[100] A. K. Soper. The quest for the structure of water and aqueous solutions. *J. Phys. Condens. Matter*, 9:2717–2730, 1997.

[101] Feng Wang, Sergei Izvekov, and Gregory A. Voth. Unusual "Amphiphilic" Association of Hydrated Protons in Strong Acid Solution. *J. Am. Chem. Soc.*, 130:3120–3126, 2008.

[102] N. Jiten Singh, Mina Park, Seung Kyu Min, Seung Bum Suh, and Kwang S. Kim. Magic and antimagic protonated water clusters: exotic structures with unusual dynamic effects. *Angew. Chem. Int. Ed.*, 45:3795–3800, 2006.

[103] Aron J. Cohen, Paula Mori-Sánchez, and Weitao Yang. Challenges for density functional theory. *Chem. Rev.*, 112:289–320, 2012.

[104] Kieron Burke. Perspective on density functional theory. *J. Chem. Phys.*, 136:150901, 2012.

[105] Kirk A. Peterson, David Feller, and David A. Dixon. Chemical accuracy in ab initio thermochemistry and spectroscopy: current strategies and future challenges. *Theor. Chem. Acc.*, 131:1079, 2012.

[106] Hans Matrin Senn and Walter Thiel. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Edit.*, 48:1198–1229, 2009.

[107] Marc W. van der Kamp and Adrian J. Mulholland. Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry*, 52:2708–2728, 2013.

[108] Orlando Acevedo and William L. Jorgensen. Advances in quantum and molecular mechanical (QM/MM) simulations for organic and enzymatic reactions. *Acc. Chem. Res.*, 43:142–151, 2010.

[109] Richard A. Friesner and victor Guallar. Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis. *Annu. Rev. Phys. Chem.*, 56:389–427, 2005.

[110] I. Rossi and D. G. Truhlar. Parameterization of NDDO wavefunctions using genetic algorithms. An evolutionary approach to parameterizing potential energy surfaces and direct dynamics calculations for organic reactions. *Chem. Phys. Lett.*, 233:231–236, 1995.

[111] Kwangho Nam, Qiang Cui, Jiali Gao, and Darrin M. York. Specific reaction parametrization of the AM1/d Hamiltonian for phosphoryl transfer reactions: H, O, and P atoms. *J. Chem. Theory Comput.*, 3:486–504, 2007.

[112] Ismael Tejero, Àngels González-Lafont, and José M. LLuch. A PM3/s specific reaction parameterization for iron atom in the hydrogen abraction catalyzed by soybean lipoxygenase-1. *J. Comput. Chem.*, 28:997–1005, 2007.

[113] Joshua P. Layfield, Matthew D. Owens, and Diego Troya. Theoretical study of the dynamics of the H + CH_4 and H + C_2H_6 reactions usig a specific-reaction-parameter semiempirical Hamiltonian. *J. Chem. Phys.*, 128:194302, 2008.

[114] Michael J. S. Dewar, Eve Zoebisch, Eamonn F. Healy, and James J. P. Stewart. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.*, 107:3902–3909, 1985.

[115] DeLos F. DeTar. Effects of alkyl groups on rates of S_N2 reactions. *J. Org. Chem.*, 45:5174–5176, 1980.

[116] Xin Gui Zhao, Susan C. Tucker, and Donald G. Truhlar. Solvent and secondary kinetic isotope effects for the mircohydrated S_N2 reaction of $Cl^-(H_2O)_n$ with CH_3Cl. *J. Am. Chem. Soc.*, 113:826–832, 1991.

[117] Wei-Ping Hu and Donald G. Truhlar. Deuterium kinetic isotope effects and their temperature dependence in the gas-phse S_N2 reactions X^- + CH_3Y -¿ CH_3 + Y^- (X,Y = Cl, Br, I). *J. Am. Chem. Soc.*, 117:10726–10734, 1995.

[118] Haobin Wang William L. Hase. Kinetics of F^- + CH_3Cl S_N4 nucleophilic substitution. *J. Am. Chem. Soc.*, 119:3093–3102, 1997.

[119] Ronald L. Yates, Nicolaos D. Epiotis, and Fernando Bernardi. The importance of nonbonded attraction in the stereochemistry of the S_N2' reaction. *J. Am. Chem. Soc.*, 97:6615–6621, 1975.

[120] Clemenica Rodriguez, Kathryn Linge, Palenque, Blair, Francesco Busetti, Brian Devine, Paul Van Buynder, Philip Weinstein, and Angus Cook. Recycled water: Potential health risks from volatile organic compounds and use of 1,4-dichlorobenzene as treatment performance indicator. *Water Res.*, 46:93–106, 2012.

[121] M. O. Andreae, E. Atlas, G.W. Harris, G. Helas, A. de Kock, R. Koppmann, W. Maenhaut, S. Manø, W. H. Pollock, J. Rudolph, D. Scharffe, G.Schebeske, and M. Welling. Methyl halide emissions from savanna fires in southern Africa. *J. Geophys. Res.*, 101:23603–23613, 1996.

[122] F. Keppler, R. Eiden adn V. Niedan, J. Pracht, and H. F. Schöler. Halocarbos produced by natural oxidation processes during degradation of organic matter. *Nature*, 403:298–301, 2000.

[123] Robert C. Rhew, Benjamin R. Miller, and Ray F. Weiss. Natural methyl bromide and methyl chloride emissions from coastal salt marshes. *Nature*, 403:292–295, 2000.

[124] Y. Yokouchi, Y. Noijiri, L.A. Barrie, D. Toom-Sauntry, T. Machida, Y. Inuzuka, H. Akimoto, H.-J. Li, Y. Fujinuma, and S. Aold. A strong source of methyl chloride to the atmosphere from tropical coastal land. *Nature*, 403:295–298, 2000.

[125] Vincent F. DeTuri, Paul A. Hintz, and Kent M. Ervin. Translational activation of the S_N2 nucleophilic displacement reactions Cl^- + CH_3Cl (CD_3Cl) -¿CH_3Cl (CD_3Cl) + Cl^-: A guided ion beam study. *J. Phys. Chem. A*, 101:5969–5986, 1997.

[126] Scott Gronert, Charles H. DePuy, and Veronica M. Bierbaum. Deuterium isotope effects in gas-phase reactions of alkyl halides: distinguishing E2 and S_N2 pathways. *J. Am. Chem. Soc.*, 113:4009–4010, 1991.

[127] Jon K. Laerdahl and einar Uggerud. Gas phase nucleophilic substitution. *Int. J. Mass. Spectrom.*, 214:277–314, 2002.

[128] William N. Olmstead and John I. Brauman. Gas-phase Nuleophilic displacement reactions. *J. Am. Chem. Soc.*, 99:4219–4228, 1977.

[129] Michael L. Chabinyc, Stephen L. Craig, Colleen K. Regan, and John I. Brauman. Gas-phase ionic reactions: Dynamics and mechanism of nucleopphilic displacements. *Science*, 279:1882–1886, 1998.

[130] William L. Hase. Simulations of gas phase reactions: Applications to S_N2 Nucleophilic Substitution. *Science*, 266:998–1002, 1994.

[131] P. A. Bash, M. J. Field, and M. Karplus. Free energy perturbation method for chemical reactions in the condensed phase: A dynamic approach based on

a combined quantum and molecular mechanics potential. *J. Am. Chem. Soc.*, 109:8092–8094, 1987.

[132] Yirong Mo and Jiali Gao. *AbInitio* QM/MM Simulations with a Molecular Orbital-Valence Bond (MOVB) Method: Application to an S_N2 Reaction in Water. *J. Comput. Chem.*, 21(16):1458–1469, 2000.

[133] Grigorly Vayner, K. N. Houk, William L. Jorgensen, and John I. Brauman. Steric Retardation of SN2 Reactions in the Gas Phase and Solution. *J. Am. Chem. Soc.*, 126:9054–9058, 2004.

[134] Srinivasan Parthiban, Glênisson de Oliveira, and Jan M. L. Martin. Benchmark ab Inition Energy Profiles for the Gas-Phase S_N2 Reactions $Y^- + CH_3X \rightarrow CH_3Y + X^-$ (X,Y = F,Cl,Br). Validation of Hybrid DFT Methods. *J. Phys. Chem. A*, 105:895–904, 2001.

[135] Zhengyu Zhou, Xinming Zhou, Hui Fu, and Laijin Tian. Vibrational spectroscopic study of electron transfer for the $Cl^- + CH_3Cl$ S_N2 reaction in the gas phase. *Spectrochimica Acta Part A*, 58:2061–2067, 2002.

[136] Benjamin J. Lynch, Patton L. Fast, Maegan Harris, and Donald G. Truhlar. Adiabatic Connection for Kinetics. *J. Phys. Chem. A*, 104:4811–4815, 2000.

[137] Ligchun Song, Wei Wu, Philippe C. Hiberty, and Sason Shaik. Identity S_N2 reactions $X^- + CH_3X$ to $XCH_3 + X^-$ (X=F, Cl, Br, and I) in vacuum and in aqueous solution: A valence bond study. *Chem. Eur. J.*, 12:7458–7466, 2006.

[138] Sason S. Shaik. Solvent effect on reaction barriers. The S_N2 reaction. 1. Application to the identity exchange. *J. Am. Chem. Soc.*, 106:1227–1232, 1984.

[139] Susan C. Tucker and Donald G. Truhlar. Ab Inition calculations of the transition-state geogmetry and vibrational frequencies of the S_N2 reaction of Cl^- with CH_3Cl. *J. Phys. Chem.*, 93:8138–8142, 1989.

[140] Steven R. Davis. Ab Initio study of the insertion reaction of Mg into the carbon-halogen bon of the fluoro- and chloromethane. *J. Am. Chem. Soc.*, 113:4145–4150, 1991.

[141] M. V. Basilevsky, G. E. Chudinov, and D.V. Napolov. Calculated rate constant for the raction Cl^- + CH_3Cl + Cl^- in the framework of continuum medium model. *J. Phys. Chem.*, 97:3270–3277, 1993.

[142] Eleonora Echegaray and Alejandro Toro-Labbé. Reaction electronic fluxx: a new concept to get insights into reaction mechanisms. Study of model symmetric nucleophilic substitutions. *J. Phys. Chem. A*, 112:11801–11807, 2008.

[143] Shaui Liang and Adrian E. Roitberg. AM1 Specific Reaction Parameters for Reactions of Hydroxide Ion with Halomethanes in Complex Environments: Development and Testing. *J. Chem. Theory Comput.*, 9:4470–4480, 2013.

[144] Chr. Møller and M. S. Plesset. Note on an approximation treatment for many-electron systems. *Phys. Rev.*, 46:618–622, 1934.

[145] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A.*, 38:3098–3100, 1988.

[146] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.*, 98(45):11623–116237, 1994.

[147] Jeng-Da Chai and Martin Head-Gordon. Long-range corrected hybid density functionals with atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.*, 10:6615–6620, 2008.

[148] Yan Zhao and Donald G. Truhlar. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.*, 120:215–241, 2008.

[149] Axe1 D. Becke. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.*, 98(2):1372–1377, 1993.

[150] Oleg V. Shishkin, Leonid Gorb, Anatoly V. Luzanov, Marcus Elstner, Sándor Suhai, and Jerzy Leszczynski. Structure and conformational flexibility of uracil: a

comprehensive study of performance of the MP2, B3LYP and SCC-DFTB methods. *J. Mol. Struct. (Theochem)*, 625:295–303, 2003.

[151] Stefan Grimme. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.*, 25:1463–1473, 2004.

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

[153] Roy Dennington II, Todd Keith, John Millam, Ken Eppinnett, W. Lee Hovell, and Ray Gilliland. Gaussview, Version 3.09. Semichem, Inc., Shawnee Mission, KS, 2003.

[154] Jan M. L. Martin. Heat of atomization of sulfur trioxide, SO_3: a benchmark for computational thermochemistry. *Chem. Phys. Lett.*, 310:271–276, 1999.

[155] Jan M. L. Martin and Glênisson de Oliveira. Towards standard methods for benchmark quality ab initio thermochemistry – W1 and W2 theory. *J. Chem. Phys.*, 111:1843–1856, 1999.

[156] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and W. P. Flannery. *Numerical Recipes in Fortran*. Cambridge University Press, Cambridge, 2^{nd} edition, 1992.

[157] Walter Thiel. *MNDO97, version 5.0.* University of Zurich, 1998.

[158] D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Götz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. A. Kollman. *AMBER 12.* University of California, San Francisco, San Francisco, CA, 2012.

[159] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.

[160] Hans W. Horn, William C. Swope, Jed W. Pitera, Jeffry D. Madura, Thomas J. Dick, Greg L. Hura, and Teresa Head-Gordon. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.*, 120(20):9665–9678, 2004.

[161] In Suk Joung and Thomas E. Cheatham III. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*, 112:9020–9041, 2008.

[162] Stephan E. Barlow, Jane M. Van Doren, and Veronica M. Bierbaum. The gas-phase displacement reaction of chloride ion with methyl chloride as a function of kinetic energy. *J. Am. Chem. Soc.*, 110:7240–7242, 1988.

[163] Charles H. DePuy, Scott Gronert, Amy Mullin, and Veronica M. Bierbaum. Gas-phase S_N2 and E2 reactions of alkyl halides. *J. Am. Chem. Soc.*, 112:8650–8655, 1990.

[164] Chun Li, Philip Ross, Jan E. Szulejko, and Terrance B. Mcmahon. High-pressure mass spectrometric investigations of the potential energy surfaces of gas-phase S_N2 reactions. *J. Am. Chem. Soc.*, 118:9360–9367, 1996.

[165] Berzinsh U, M. Gustafsson, D. Hanstorp, A. Klinkmüller, U. Ljungblad, and A.-M. Mårtensson-Pendrill. Isotope shift in the electron affinity of chlorine. *Phys. Rev. A*, 51:231–238, 1995.

[166] Sigid D. Peyeerimhoff and Robert J. Buenker. Electronically excited and ionized stated of the chlorine molecule. *Chem. Phys.*, 57:279–296, 1981.

[167] Per Jensen and Svend Brodersen. Determination of A_0 for the $CH_3{}^{35}Cl$ and $CH_3{}^{37}Cl$ from the v_4 infrared and Raman bands. *J. Mol. Spectrosc.*, 88:378–393, 1981.

[168] Jon Applequist and Clifford E. Felder. Electron transfer energies and dipole moments of alkyl halides and amines from an electrostatic model. *J. Chem. Phys.*, 75:1863–1868, 1981.

[169] Yonezo Morino, Yasushi Nakamura, and Takao Iijima. Mean square amplitudes and force constants of tetrahedral molecules. I. Carbon tetrachloride and germanium tetrachloride. *J. Chem. Phys.*, 32:643–652, 1960.

[170] W. John Albery and Maurice M. Kreevoy. Methyl Transfer Reactions. *Adv. Phys. Org. Chem.*, 16:87–157, 1978.

[171] Cui Zhang, Tuan Anh Pham, François Gygi, and Giulia Galli. Communication: Electronic structure of solvated chloride anion from first principles molecular dynamics. *J. Chem. Phys.*, 138:181102, 2013.

[172] Timothy J. Giese and Darrin M. York. Charge-dependent model for many-body polarization, exchange, and dispersion interactions in hybrid quantum mechanical/molecular mechanical calculations. *J. Chem. Phys.*, 127(19):194101, 2007.

[173] Timothy J. Giese and Darrin M. York. Density-functional expansion methods: grand challenges. *Theor. Chem. Acc.*, 131(3):1145, 2012.

[174] D. A. Case, V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D.S Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Götz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, Tai-Sung Lee, S. Le Grand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe,

A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, and P. A. Kollman. *AMBER 14*. University of California, San Francisco, San Francisco, CA, 2014.

[175] Adri C.T. van Duin, Siddharth Dasgupta, Francois Lorant, and Willliam A. Goddard III. ReaxFF: a reactive force field for hydrocarbons. *J. Phys. Chem. A*, 105(41):9396–9409, 2001.

[176] W. Xie and J. Gao. Design of a next generation force field: The x-pol potential. *J. Chem. Theory. Comput.*, 3(6):1890–1900, 2007.

[177] Timothy J Giese, Ming Huang, Haoyuan Chen, and Darrin M York. Recent Advances toward a General Purpose Linear-Scaling Quantum Force Field. *Acc. Chem. Res.*, 47(9):2812–20, September 2014.

[178] Timothy J. Giese, Maria T. Panteva, Haoyuan Chen, and Darrin M. York. Multipolar Ewald methods, 2: Applications using a quantum mechanical force field. *J. Chem. Theory Comput.*, 11:451–461, 2015.

[179] Hans Martin Senn and Walter Thiel. QM/MM studies of enzymes. *Curr. Opin. Chem. Biol.*, 11:182–187, 2007.

[180] Igor A. Topol, Raul E. Cachau, Alexander V. Nemukhin, Bella L. Grigorenko, and Stanley K. Burt. Quantum chemical modeling of the GTP hydrolysis by the RASGAP protein complex. *Biochim. Biophys. Acta*, 1700:125–136, 2004.

[181] L. S. Devi-Kesavan, M. Garcia-Viloca, and J. Gao. Semiempirical QM/MM potential with simple valence bond (SVB) for enzyme reactions. Application to the nucleophilic addition reaction in haloalkane dehalogenase. *Theor. Chem. Acc.*, 109:133–139, 2003.

[182] Tai-Sung Lee, Carlos Silva Lopez, George M. Giambaşu, Monika Martick, William G. Scott, and Darrin M. York. Role of Mg^{2+} in hammerhead ribozyme catalysis from molecular simulation. *J. Am. Chem. Soc.*, 130(10):3053–3064, 2008.

[183] Kin-Yiu Wong, Tai-Sung Lee, and Darrin M. York. Active participation of the Mg^{2+} ion in the reaction coordinate of RNA self-cleavage catalyzed by the hammerhead ribozyme. *J. Chem. Theory Comput.*, 7(1):1–3, 2011.

[184] Pavel Banáš, Petr Jurečka, Nils G. Walter, Jiří Šponer, and Michal Otyepka. Theoretical studies of rna catalysis: hybrid qm/mm methods and their comparison with md and qm. *Methods*, 49(2):202–216, 2009.

[185] Art E. Cho, Victor Guallar, Bruce Berne, and Richard Friesner. Importance of Accurate Charges in Molecular Docking: Quantum Mechanical/Molecular Mechanical (QM/MM) Approach. *J. Comput. Chem.*, 26:915–931, 2005.

[186] William L. Jorgensen and Julian Tirado-Rives. Molecular Modeling of Organic and Biomolecular Systems Using BOSS and MCPRO. *J. Comput. Chem.*, 26:1689–1700, 2005.

[187] Jens Antony, Stefan Grimme, Dimitrios G. Liakos, and Frank Neese. Protein-Ligand Interaction Energies with Dispersion Corrected Density Functional Theory and High-Level Wave Function Based Methods. *J. Phys. Chem. A*, 115:11210–11220, 2011.

[188] Shina C. L. Kamerlin, Maciej Haranczyk, and Arieh Warshel. Progress in Ab Initio QM/MM Free-Energy Simulations of Electrostatic Energies in Proteins: Accelerated QM/MM Studies of pKa, Redox Reactions and Solvation Free Energies. *J. Phys. Chem. B*, 113:1253–1272, 2009.

[189] Jiali Gao, Naiqi Li, and Marek Freindorf. Hybrid qm/mm simulations yield the ground and excited state pka difference: phenol in aqueous solution. *J. Am. Chem. Soc.*, 118:4912–4913, 1996.

[190] Brent A. Gregersen, Xabier Lopez, and Darrin M. York. Hybrid QM/MM Study of Thio Effects in Transphosphorylation Reactions: The Role of Solvation. *J. Am. Chem. Soc.*, 126:7504–7513, 2004.

[191] Xabier Lopez, Darrin M. York, Annick Dejaegere, and Martin Karplus. Theoretical Studies on the Hydrolysis of Phosphate Diesters in the Gas Phase, Solution, and RNase A. *Int. J. Quantum Chem.*, 86:10–26, 2002.

[192] Marc W. van der Kamp and Adrian J. Mulholland. Computational enzymology: insight into biological catalysts from modelling, 2008.

[193] Brian K. Radak, Michael E. Harris, and Darrin M. York. Molecular simulations of RNA 2'-O-transesterification reaction models in solution. *J. Phys. Chem. B*, 117:94–103, 2013.

[194] Nathalie Reuter, Annick Dejaegere, Bernard Maigret, and Martin Karplus. Frontier Bonds in QM/MM Methods: A Comparison of Different Approaches. *J. Phys. Chem. A*, 104:1720–1735, 2000.

[195] Zhiqiang Zhang, Liancai Xu, Gyoqing Wang, Chen Shao, and Ling Jiang. The development and validation of a pseudoatoms approach for combined quantum mechanics and molecular mechanics calculations of polypeptide. *J. Mol. Struct.*, 1043:38–46, 2014.

[196] Jingzhi Pu, Prof. Jiali Gao, and Prf. Donald G. Truhlar. Generalized hybrid-orbital method for combining density functional theory with molecular mechanicals. *Chem. Phys. Chem.*, 6:1853–1865, 2005.

[197] Jingzhi Pu, Jiali Gao, and Donald G. Truhlar. Generalized hybrid orbital (GHO) method for combining ab initio Hartree-Fock wave functions with molecular mechanics. *J. Phys. Chem. A*, 108:632–650, 2004.

[198] Hai Lin and Donald G. Truhlar. QM/MM: what have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.*, 117:185–199, 2007.

[199] Jiali Gao. Energy components of aqueous solution: Insight from hybrid QM/MM simulations using a polarizable solvent model. *J. Comput. Chem.*, 18:1061–1071, 1997.

[200] Yan Zhang, Hai Lin, and Donald G. Truhlar. Self-consistent polarization of the boundary in the redistributed charge and dipole scheme for combined quantum-mechanical and molecular-mechanical calculations. *J. Chem. Theory Comput.*, 3:1378–1398, 2007.

[201] J. E. Lennard-Jones. Cohesion. *Proc. Phys. Soc.*, 43:461–482, 1931.

[202] M. Freindorf and J. Gao. Optimization of the Lennard-Jones parameters for a combined ab initio quantum mechanical and molecular mechanical potential using the 3-21G basis set. *J. Comput. Chem.*, 17:386–395, 1996.

[203] Marek Freindorf, Yihan Shao, Thomas R. Furlani, and Jing Kong. Lennard-Jones Parameters for the Combined QM/MM Method Using the B3LYP/6-31 G*/AMBER Potential. *J. Comput. Chem.*, 26:1270–1278, 2005.

[204] Rosa E. Bulo, Bernd Ensing, Jetze Sikkema, and Lucas Visscher. Toward a Practical Method for Adaptive QM/MM Simulations. *J. Chem. Theory Comput.*, 5(9):2212–2221, 2009.

[205] Steven O. Nielsen, Rosa E. Bulo, Preston B. Moore, and Bernd Ensing. Recent progress in adaptive multiscale molecular dynamics simulations of soft matter. *Phys. Chem. Chem. Phys.*, 12(39):12401–12414, 2010.

[206] Kyoyeon Park, Andreas W. Götz, Ross C. Walker, and Francesco Paesani. Application of Adaptive QM/MM Methods to Molecular Dynamics Simulations of Aqueous Systems. *J. Chem. Theory Comput.*, 8(8):2868–2877, 2012.

[207] Guanhua Hou, Xiao Zhu, Marcus Elstner, and Qiang Cui. A Modified QM/MM Hamiltonian with the Self-Consistent-Charge Density-Functional-Tight-Binding Theory for Highly Charged QM Regions. *J. Chem. Theory Comput.*, 8:4293–4304, 2012.

[208] Demian Riccardi, Patricia Schaefer, Yang Yang, Haibo Yu, Nilanjan Ghosh, Xavier Prat-Resina, Peter König, Guohui Li, Dingguo Xu, Hua Guo, Marcus Elstner, and Qiang Cui. Development of effective quantum mechanical/molecular mechanical (QM/MM) methods for complex biological processes. *J. Phys. Chem. B*, 110:6458–6469, 2006.

[209] Erich R. Kuechler and Darrin M. York. Quantum menchanical study of solvent effects in a prototype S_N2 reaction in solution: Cl^- attack on CH_3Cl. *J. Chem. Phys.*, 140:054109, 2014.

[210] Bradley J. Gertner, Robert M. Whitnell, Kent R. Wilson, and James T. Hynes. Activation to the transition state: reactant and solvent energy flow for a model SN2 reaction in water. *J. Am. Chem. Soc.*, 113:74–87, 1991.

[211] Marc A. van Bochove and F. Matthias Bickelhaupt. Nucleophilic substitution at C, Si, and P: How solvation affects the shape of reaction profiles. *Eur. J. Org. Chem.*, 4:649–654, 2008.

[212] Junmei Wang, Piotr Cieplak, and Peter A. Kollman. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic biological molecules. *J. Comput. Chem.*, 21(12):1049–1074, 2000.

[213] John D. Weeks, David Chandler, and Hans C. Andersen. Role of repulsive forces in determining the equilibrium structure of simple liquids. *J. Chem. Phys.*, 54:5237–5247, 1971.

[214] Loup Verlet. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, 159(1):98–103, July 1967.

[215] S. Kita, K. Noda, and H. Inouye. Repulsion potentials for Cl^--R and Br^--R (R=He, Ne, and Ar) derived from beam experiments. *J. Chem. Phys.*, 64(8):3446–3449, 1976.

[216] Richard J. Wheatley and Sarah L. Price. An overlap model for estimating the anisotropy of repulsion. *Mol. Phys.*, 69(3):507–533, 1990.

[217] F. Wang and K. D. Jordan. A Drude-model approach to dispersion interactions in dipole-bound anions. *J. Chem. Phys.*, 114(24):10717–10724, 2001.

[218] Timothy J. Giese and Darrin M. York. Improvement of semiempirical response properties with charge-dependent response density. *J. Chem. Phys.*, 123(16):164108, 2005.

[219] David R. Lide, editor. *CRC handbook of chemistry and physics.* CRC Press LLC, Boca Raton, FL, 90 edition, 2010.

[220] W. Kohn and L. Sham. Self-consistent equations including exchange and correlation effcts. *Phys. Rev. A*, 140:A1133–A1138, 1965.

[221] Felix Sedlmeier, Dominik Horinek, and Roland R. Netz. Entropy and ethalpy convergence of hydrophobic solvation beyond the hard-sphere limit. *J. Chem. Phys.*, 134:055105, 2011.

[222] Jean-Philip Piquemal, G. Andrés Cisneros, Peter Reinhardt, Nohad Gresh, and Thomas A. Darden. Towards a force field based on density fitting. *J. Chem. Phys.*, 124(10):104101, 2006.

[223] K. T. Tang and J. Peter Toennies. The damping function of the van der Waals attraction in the potential between rare gas atoms and metal surfaces. *Surf. Sci. Lett.*, 279(3):L203–L206, 1992.

[224] R. J.-M. Pellenq and D. Nicholson. A simple method for calculating dispersion coefficients for isolated and condensed-phase species. *Mol. Phys.*, 95(3):549–570, 1998.

[225] R. J.-M. Pellenq and D. Nicholson. ERRATUM: A simple method for calculating dispersion coefficients for isolated and condensed-phase species. *Mol. Phys.*, 96(6):1001, 1999.

[226] Jana Khandogin and Darrin M. York. Quantum Descriptors for Biological Macromolecules from Linear-Scaling Electronic Structure Methods. *Proteins*, 56:724–737, 2004.

[227] David A. Pearlman. Determining the contributions of constraints in free energy calculations: Development, characterization, and recommendations. *J. Chem. Phys.*, 98:8946–8957, 1993.

[228] Michael J. Mitchell and J. Andrew McCammon. Free Energy Difference Calculations by Thermodynamic Integration: Difficulties in Obtaining a Precise Value. *J. Comput. Chem.*, 12:271–275, 1991.

[229] Thomas Steinbrecher, David L. Mobley, and David A. Case. Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations. *J. Chem. Phys.*, 127:214108, 2007.

[230] John P. Bergsma, Bradley J. Gertner, Kent R. Wilson, and James T. Hynes. Molecular dynamics of a model S_N2 reaction in water. *J. Chem. Phys.*, 86:1356–1376, 1987.

[231] Roland Schmid, Arzu M. Miah, and Valentin N. Sapunov. A new table of the thermodynamic quantities of ionic hydration: values and some applications (enthalpy-entropy compensation and Born radii). *Phys. Chem. Chem. Phys.*, 2:97–102, 2000.

[232] Adam C. Chamberlin, Christopher J. Cramer, and Donald G. Truhlar. Exension of a temperature-dependent aqueous solvation model to compounds containing nitrogen, fluorine, chlorine, bromine and sulfur. *J. Phys. Chem. B*, 112:3024–3039, 2008.

[233] Donald Mackay and Wan Wing Shiu. A critical review of Henry's law constants for chemicals of environmental interest. *J. Phys. Chem. Ref. Data*, 10:1175–1199, 1981.

[234] R. Sander. Compilation of Henry's law constants version 3.99. *Atmos. Chem. Phys. Discuss*, 14:29615–30521, 2014.

[235] Michael H. Hiatt. Determination of Henry's law constants using internal standards with benchmark values. *J. Chem. Eng. Data*, 58:902–908, 2013.

[236] Hongsuk Kang, Hwanho Choi, and Hwangseo Park. Prediction of molecular solvation free energy based on the optimization of atomic solvation parameters with genetic algorithm. *J. Chem. Inf. Model.*, 47:509–514, 2007.

[237] Steven J. Stuart and B. J. Berne. Effects of Polarizability of the Hydration of the Chloride Ion. *J. Phys. Chem.*, 100:11934–11943, 1996.

[238] Willian L. Jorgensen and Daniel L. Severance. Limited effects of polarization for Cl-$(H_2O)_n$ and $Na^+(H_2O)_n$ cluster. *J. Chem. Phys.*, 99:4233–4235, 1993.

[239] Lalith Perera and Max L. Berkowitz. Many-body effects in molecular dynamics simulations of $Na^+(H_2O)_n$ and $Cl^-(H_2O)_n$ clusters. *J. Chem. Phys.*, 95:1954–1963, 1991.

[240] Pavel Jungwirth and Douglas J. Tobias. Chloride anion on aqueous clusters, at the air–water interface, and in liduid water: Solcaent effects on Cl^- polarizability. *J. Phys. Chem. A*, 106:379–383, 2002.

[241] Douglas Tobias, Pavel Jungwirth, and Michele Parrinello. Surface solvation of halogen anions in water clusters: An ab initio molecular dynamics study of the $Cl^-(H_2O)_6$. *J. Chem. Phys.*, 114:7036–7044, 2001.

[242] Andreas Serr and Roland R. Netz. Polarizabilities of hydrated and free ions derived from DFT calculations. *Int. J. Quantum Chem.*, 106:2960–2974, 2006.

[243] Steve Kaminski, Timothy J. Giese, Michael Gaus, Darrin M. York, and Marcus Elstner. Extended Polarization in Third-Order SCC-DFTB from Chemical-Potential Equalization. *J. Phys. Chem. A*, 116(36):9131–9141, 2012.

[244] Darrin M. York and Weitao Yang. A chemical potential equalization method for molecular simulations. *J. Chem. Phys.*, 104:159–172, 1996.

[245] Timothy J. Giese, Haoyuan Chen, Ming Huang, and Darrin M. York. Parametrization of an orbital-based linear-scaling quantum force field for noncovalent interactions. *J. Chem. Theory Comput.*, 10:1086–1098, 2014.

[246] Vitor H. Teixeira, Cláudio M. Soares, and António M. Baptista. Proton pathways in a [NiFe]-hydrogenase: A theoretical study. *Proteins*, 70(3):1010–1022, 2008.

[247] Anna Pabis, Inacrist Geronimo, Darrin M. York, and Piotr Paneth. Molecular Dynamics Simulation of Nitrobenzene Dioxygenase Using AMBER Force Field. *J. Chem. Theory Comput.*, 10:2246–2254, 2014.

[248] Yun Liu, Brent A. Gregersen, Alvan Hengge, and Darrin M. York. Transesterification Thio Effects of Phosphate Diesters: Free Energy Barriers and Kinetic and Equilibrium Isotope Effects from Density-Functional Theory. *Biochemistry*, 45:10043–10053, 2006.

[249] Kevin Range, Matthew J. McGrath, Xabier Lopez, and Darrin M. York. The Structure and Stability of Biological Metaphosphate, Phosphate, and Phosphorane Compounds in the Gas Phase and in Solution. *J. Am. Chem. Soc.*, 126:1654–1665, 2004.

[250] Haoyuan Chen, Timothy J. Giese, Ming Huang, Kin-Yiu Wong, Michael E. Harris, and Darrin M. York. Mechanistic Insights into RNA Transphosphorylation from Kinetic Isotope Effects and Linear Free Energy Relationships of Model Reactions. *Chem. Eur. J.*, 20:14336–14343, 2014.

[251] Hugh Heldenbrand, Pawel A. Janowski, George Giambaşu, Timothy J. Giese, Joseph E. Wedekind, and Darrin M. York. Evidence for the role of active site residues in the hairpin ribozyme from molecular simulations along the reaction path. *J. Am. Chem. Soc.*, 136:7789–7792, 2014.

[252] Jérémie Mortier, Christin Rakers, Marcel Bermudez, Manuela S. Murgueitio, Sereina Riniker, and Gerhard Wolber. The impact of molecular dynamics on drug design: applications for the characterization of ligand-macromolecule complexes. *Drug Discov. Today*, 0:0, January 2015.

[253] Lingle Wang, Yujie Wu, Yuqing Deng, Byungchan Kim, Levi Pierce, Goran Krilov, Dmitry Lupyan, Shaughnessy Robinson, Markus K. Dahlgren, Jeremy Greenwood, Donna L. Romero, Craig Masse, Jennifer L. Knight, Thomas Steinbrecher, Thijs Beuming, Wolfgang Damm, Ed Harder, Woody Sherman, Mark Brewer, Ron Wester, Mark Murcko, Leah Frye, Ramy Farid, Teng Lin, David L. Mobley, William L. Jorgensen, Bruce J. Berne, Richard A. Friesner, and Robert Abel. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.*, 0(0):0–1, 2015, http://dx.doi.org/10.1021/ja512751q. PMID: 25625324.

[254] Christopher J. Cramer and Donald G. Truhlar. Implicit solvation models: equilibria, structure, spectra, and dynamics. *Chem. Rev.*, 99(8):2161–2200, 1999.

[255] Benoît Roux and Thomas Simonson. Implicit solvent models. *Biophys. Chem.*, 78:1–20, 1999.

[256] Thanh N. Truong. Quantum modeling of reactions in solution: an overview of the dielectric continuum methodology. *Int. Rev. Phys. Chem.*, 17:525–546, 1998.

[257] Wonpil Im, Jianhan Chen, and Charles L. Brooks III. Peptide and protein folding and conformational equilibria: theoretical treatment of electrostatics and hydrogen bonding with implicit solvent models. *Adv. Protein Chem.*, 72:173–198, 2006.

[258] Frank Eckert and Andreas Klamt. Accurate Prediction of Basicity in Aqueous Solution with COSMO-RS. *J. Comput. Chem.*, 27:11–19, 2006.

[259] Frank Eckert and Andreas Klamt. Fast Solvent Screening via Quantum Chemistry: COSMO-RS Approach . *AIChEJ*, 48:369–385, 2002.

[260] Jiali Gao and Donald G. Truhlar. Quantum Mechanical Methods for Enzyme Kinetics. *Annu. Rev. Phys. Chem.*, 53:467–505, 2002.

[261] M. Elstner, Th. Frauenheim, E. Kaxiras, G. Seifert, and S. Suhai. A Self-Consistent Charge Density-Functional Based Tight-Binding Scheme for Large Biomolecules. *Phys. Status Solidi. B*, 217:357–376, 2000.

[262] A. K. McMahan. Two-center s-f Slater-Koster integrals. *Phys. Rev. B*, 58(8):4293–4303, 1998.

[263] Michael J. S. Dewar and Kenneth M. Dieter. Evaluation of AM1 calculated proton affinities and deprotonation enthalpies. *J. Am. Chem. Soc.*, 108(25):8075–8086, 1986.

[264] J. J. P. Stewart. Optimization of parameters for semiempirical methods II. applications. *J. Comput. Chem.*, 10:221–264, 1989.

[265] James J. P. Stewart. Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.*, 10:209–220, 1989.

[266] Thomas Krüger and Marcus Elstner. Validation of the density-functional based tight-binding approximation method for the calculation of reaction energies and other data. *J. Chem. Phys.*, 122:114110, 2005.

[267] Nikolaj Otte, Mirjam Scholten, and Walter Thiel. Looking at Self-Consistent-Charge Density Functional Tight Binding from a Semiempirical Perspective. *J. Phys. Chem. A*, 111:5751–5755, 2007.

[268] Michael Gaus, Xiya Lu, Marcus Elstner, and Qiang Cui. Parameterization of DFTB3/3OB for Sulfur and Phosphorus for Chemical and Biological Applications. *J. Chem. Theory Comput.*, 10(4):1518–1537, 2014.

[269] Guanhua Hou, Xiao Zhu, and Qiang Cui. An implicit solvent model for SCC-DFTB with charge-dependent radii. *J. Chem. Theory Comput.*, 6(8):2303–2314, 2010.

[270] Bojana Ginovska, Donald M. Camaioni, Michel Dupuis, Christine A. Schwerdtfeger, and Quinn Gil. Charge-Dependent Cavity Radii for an Accurate Dielectric Continuum Model of Solvation with Emphasis on Ions: Aqueous Solutes with Oxo, Hydroxo, Amino, Methyl, Chloro, Bromo, and Fluoro Functionalities. *J. Phys. Chem. A*, 112:10604–10613, 2008.

[271] M. Aguilar, M. Martin, S. Tolosa, and F. Olivares Del Valle. Solute-solvent interaction. adaptation of the cavity size as a function of the basis set. *J. Mol. Struct. (Theochem)*, 166:313–318, 1988.

[272] M. Aguilar and F. Olivares Del Valle. Solute-Solvent interactions. A simple procedure for constructing the solvent cavity for retaining a molecular solute. *Chem. Phys.*, 129:439–450, 1989.

[273] Michael Gaus, Qiang Cui, and Marcus Elstner. DFTB3: Extension of the seld-consistent-charge density-functional tight-binding method (SCC-DFTB). *J. Chem. Theory Comput.*, 7(4):931–948, 2011.

[274] Michael Gaus, Albrecht Goez, and Marcus Elstner. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.*, 9:338–354, 2013.

[275] Jana Khandogin, Brent A. Gregersen, Walter Thiel, and Darrin M. York. Smooth Solvation Method for d-Orbital Semiempirical Calculations of Biological Reactions. 1. Implementation. *J. Phys. Chem. B*, 109:9799–9809, 2005.

[276] Brent A. Gregersen, Jana Khandogin, Walter Thiel, and Darrin M. York. Smooth Solvation Method for d-Orbital Semiempirical Calculations of Biological Reactions. 2. Application to Transphosphorylation Thio Effects in Solution. *J. Phys. Chem. B*, 109:9810–9817, 2005.

[277] V. I. Lebedev. Quadratures on the sphere. *Ž. Vyčisl. Mat. i Mat. Fiz.*, 16:293–306, 1976.

[278] V. I. Lebedev. Quadrature formulas for the sphere of 25th to 29th order accuracy. *Sibirsk. Mat. Ž.*, 18:132–142, 1977.

[279] Brent A. Gregersen and Darrin M. York. High-order discretization schemes for biochemical applications of boundary element solvation and variational electrostatic projection methods. *J. Chem. Phys.*, 122:194110, 2005.

[280] Johannes Kästner, Joanne M. Carr, Thomas W. Keal, Walter Thiel, Adrian Wander, and Paul Sherwood. DL-FIND: An open-source geomtery optimizer for atomistic simulations. *J. Phys. Chem. A*, 113:11856–11865, 2009.

[281] M. Elstner. SCC-DFTB: What Is the Proper Degree of Self-Consistency? *J. Phys. Chem. A*, 111:5614–5621, 2007.

[282] Yang Yang, Haibo Yu, Darrin M. York, Qiang Cui, and Marcus Elstner. Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method: Third-Order Expansion of the Density Functional Theory Total Energy and Introduction of a Modified Effective Coulomb Interaction. *J. Phys. Chem. A*, 111(42):10861–10873, 2007.

[283] Elena Formoso, Jon M Matxain, Xabier Lopez, and Darrin M York. Molecular dynamics simulation of bovine pancreatic ribonuclease A-CpA and transition state-like complexes. *J. Phys. Chem. B*, 114(21):7371–7382, Jun 2010.

[284] Hong Gu, Shuming Zhang, Kin-Yiu Wong, Brian K. Radak, Thakshila Dissanayake, Daniel L. Kellerman, Qing Dai, Masaru Miyagi, Vernon E. Anderson, Darrin M. York, Joseph A. Piccirilli, and Michael E. Harris. Experimental and computational analysis of the transition state for ribonuclease A-catalyzed RNA 2'-*O*-transphosphorylation. *Proc. Natl. Acad. Sci. USA*, 110:13002–13007, 2013.

[285] George M. Giambaşu, Darrin M. York, and David A. Case. Structural fidelity and NMR relaxation analysis in a prototype RNA hairpin. *RNA*, 21:963–974, Mar 2015.

[286] William G. Scott. *What can the New Hammerhead Ribozyme Structures Teach us About Design?*, pages 305–323. Springer-Verlag, Berlin Heidelberg, 2010.

[287] William G. Scott. Morphing the minimal and full-length hammerhead ribozymes: implications for the cleavage mechanism. *Biol. Chem.*, 388:727–735, 2007.

[288] William G. Scott, James B. Murray, John R. P. Arnold, Barry L. Stoddard, and Aaron Klug. Capturing the structure of a catalytic RNA intermediate: The Hammerhead Ribozyme. *Science*, 274:2065–2069, 1996.

[289] Sixue Zhang, Abir Ganguly, Puja Goyal, Jamie L. Bingaman, Philip C. Bevilacqua, and Sharon Hammes-Schiffer. Role of the active site guanine in the *glmS* ribozyme self-cleavage mechanism: Quantum mechanical/molecular mechanical free energy simulations. *J. Am. Chem. Soc.*, 137:784–798, 2015.

[290] Júlia Viladoms and Martha J. Fedor. The *glmS* Ribozyme Cofactor is a General Acid-Base Catalyst. *J. Am. Chem. Soc.*, 134:19043–19049, 2012.

[291] Pallavi Thaplyal, Abir Ganguly, Sharon Hammes-Schiffer, and Philip C. Bevilacqua. Inverse Thio Effects in the Hepatitis Delta Virus Ribozyme Reveal that the Reaction Pathway Is Controlled by Metal Ion Charge Density. *Biochemistry*, 54:2160–2175, March 2015.

[292] Adrian R. Ferré-D'Amaré and William G. Scott. Small Self-cleaving Ribozymes. *Cold Spring Harb Perspect Biol*, 2(10):a003574, Oct 2010.

[293] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, Th. Frauenheim, S. Suhai, and G. Seifert. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B*, 58(11):7260–7268, 1998.

[294] Ming Huang and Darrin M. York. Linear free energy relationships in RNA transesterification: theoretical models to aid experimental interpretations. *Phys. Chem. Chem. Phys.*, 16:15846–15855, 2014.

[295] D. M. Brown and D. A. Usher. Hydrolysis of Hydroxyalkyl Phosphate Esters: Effect of Changing Ester Group. *J. Chem. Soc.*, 87:6558–6564, 1965.

[296] Markus Kosonen, Esmail Youseti-Salakdeh, Roger Strömberg, and Harri Lönnberg. Mutual isomerization of uridine 2-and 3-alkylphosphates and cleavage to a 2 3-cyclic phosphate: the effect of the alkyl group on the hydronium- and hydroxide-ion-catalyzed reactions. *J. Chem. Soc. Perkin Trans. 2*, pages 2611–2666, 1997.

[297] P. W. Barnard, C. A. Bunton, D. R. Llewellyn, C. A. Vernon, and V. A. Welch. The reactions of organic phosphates. part V.the hydrolysis of triphenyl and trimethyl phosphates. *J. Chem. Soc.*, pages 2670–2676, 1961.

Appendix A

List of Abbreviations

AM1 Austin model 1
AMBER Assisted Model Building with Energy Refinement
AO atomic orbital
COSMO conductor-like screening model
DFT density functional theory
DFTB density functional tight-binding
DFTB3 third-order self-consistent density functional tight-binding
DFTB3/3OB a third-order self-consistent density functional tight-binding specific reaction aparameterization for biological chemistry
DNA dioxyriboseneucleic acid
DY Darrin York
FEP free energy perturbation
GB general Borne solvation model
GGA general gradient approximations
HF Hartree-Fock
IDC ion-dipole complex
IRC intrinsic reaction coordinate
LC long-range corrected
LJ Lennard-Jones
LSDA local spin density approximation
MBAR multistate Bennett acceptance ratio estimator
MD molecular dynamics
MM molecular mechanics or molecular mechanical
MNDO modified neglect of differential overlap
MO molecular orbital
MP2 Second-order Møller-Plesset
MOR methyloxyribose
MSE mean signed error
MSI Minnesota Supercomputing Institute
MUE mean unsigned error
NDDO neglect of diatomic differential overlap
PB Poisson-Boltzmann solvation model
PCM polarizable continuum model
PM3 parameterized model number 3
PMF potential of mean force
QM quantum mechanics or quantum mechanical
QM/MM hybrid quatum mechanical/molecular mechanical
QMFF quantum mechanical force field
QXD charge-dependent exchange and dispersion
RDF radial distribution function
RMSE root mean signed error
RNA riboneucleic acid
RxC reaction coordinate
SAS solvent accessible surface
SCC-DFTB second-order self-consistent density functional tight-binding
SCF self-consistent field
SCOSMO smooth conductor-like screening

model
SES solvent excluded surface
S_N2 bimolecular nucleophilic substitution
SRP specific reaction parameterization
TI thermodynamic integration
TMP trimethylphosphate
TS transition state
vFEP variational free energy profile method
VRSCOSMO variable radii smooth conductor-like screening model
W1 Weizmann-1
WHAM weighted histogram method
XSEDE Extreme Science and Engineering Discovery Environment

Appendix B

Parameterization Weights and MNDO97 Output for the SRP Parameterization of the AM1 Hamiltonian

Weights are resported as σ values; e.g. $1/w$ where w is the weight with the exception when $\sigma = 0$ then $w = 0$.

```
#experimental geometry
#MOLECULE CL2
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0     0.12219         0.00000         0.12219     1.000   0.015  Expt HOF
#  1   0   0   0   0     0.12219   -577536.76813    577536.89033     0.000   0.000  Cl2 Energy
#  2   1   2   0   0     1.93864         1.98800        -0.04936 40000.000  97.452  Expt Cl-Cl Bond
#experimental geometry
#MOLECULE HCL
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -35.57492       -22.06000       -13.51492     1.000 182.653  Expt HOF
#  2   1   2   0   0     1.30915         1.27500         0.03415 40000.000  46.640  Expt H-Cl Bond
#  7   0   0   0   0     1.67403         1.08000         0.59403   400.000 141.148  Expt Dipole
# 17   1   0   0   0    -0.22031        -0.40306         0.18276     0.000   0.000  Cl
# 17   2   0   0   0     0.22031        -0.18759         0.40789     0.000   0.000  H
```

```
#geometry from g03 output
#MOLECULE CCL4
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -21.56027       -23.90000         2.33973     1.000   5.474  Expt HOF
#  7   0   0   0   0     0.00386         0.00000         0.00386   400.000   0.006  Expt Dipole
#  2   1   2   0   0     1.79558         1.76700         0.02858 40000.000  32.679  C-Cl Bond
#  2   1   3   0   0     1.79571         1.76700         0.02871 40000.000  32.980  C-Cl Bond
#  2   1   4   0   0     1.79602         1.76700         0.02902 40000.000  33.682  C-Cl Bond
#  2   1   5   0   0     1.79533         1.76700         0.02833 40000.000  32.103  C-Cl Bond
#  3   2   1   3   0   109.45561       109.47120        -0.01559     4.000   0.001  Cl-C-Cl Angle
#  3   2   1   4   0   109.42982       109.47120        -0.04138     4.000   0.007  Cl-C-Cl Angle
#  3   2   1   5   0   109.47901       109.47120         0.00781     4.000   0.000  Cl-C-Cl Angle
#  3   3   1   4   0   109.46681       109.47124        -0.00443     4.000   0.000  Cl-C-Cl Angle
#  3   3   1   5   0   109.49222       109.47124         0.02098     4.000   0.002  Cl-C-Cl Angle
#  3   4   1   5   0   109.50382       109.47124         0.03258     4.000   0.004  Cl-C-Cl Angle
#geometry from g03 output
#MOLECULE MECL
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -16.84287       -19.60000         2.75713    16.000 121.628  Expt HOF
#  1   0   0   0   0   -16.84287   -313817.27717    313800.43429     0.000   0.000  MeCl Energy
# 17   1   0   0   0    -0.13223        -0.40306         0.27084     0.000   0.000  C
# 17   5   0   0   0    -0.16270        -0.18759         0.02489     0.000   0.000  Cl
#  7   0   0   0   0     1.81647         1.93060        -0.11413     0.000   0.000  Dipole
#  7   0   0   0   0     1.81647         1.87000        -0.05353 10000.000  28.656  Expt Dipole
#  2   1   2   0   0     1.11359         1.08421         0.02938 40000.000  34.536  H-C Bond
#  2   1   3   0   0     1.11351         1.08421         0.02930 40000.000  34.347  H-C Bond
#  2   1   4   0   0     1.11342         1.08421         0.02921 40000.000  34.134  H-C Bond
#  2   1   5   0   0     1.77083         1.77852        -0.00768 40000.000   2.360  C-Cl Bond
#  3   2   1   3   0   110.44136       110.51304        -0.07169     4.000   0.021  H-C-H Angle
#  3   2   1   4   0   110.48120       110.51304        -0.03185     4.000   0.004  H-C-H Angle
#  3   2   1   5   0   108.44973       108.40818         0.04156     4.000   0.007  H-C-Cl Angle
#  3   3   1   4   0   110.57514       110.51299         0.06215     4.000   0.015  H-C-H Angle
#  3   3   1   5   0   108.35955       108.40817        -0.04862     4.000   0.009  H-C-Cl Angle
#  3   4   1   5   0   108.45831       108.40817         0.05013     4.000   0.010  H-C-Cl Angle
#geometry from g03 output
#MOLECULE MECL_COMPLEX
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -83.15121   -602651.82426    602568.67305     0.000   0.000  Energy
# 17   1   0   0   0    -0.05638        -0.32591         0.26953     0.000   0.000  C
# 17   5   0   0   0    -0.30638        -0.35764         0.05126     0.000   0.000  Cl
```

```
# 17   6   0   0   0    -0.99086        -0.94227        -0.04858     0.000   0.000  Cl
#  7   0   0   0   0     8.99214         8.23910         0.75304     0.000   0.000  Dipole
#  2   1   2   0   0     1.11034         1.08090         0.02943 40000.000  34.656  H-C Bond
#  2   1   3   0   0     1.11010         1.08094         0.02916 40000.000  34.014  H-C Bond
#  2   1   4   0   0     1.11062         1.08094         0.02969 40000.000  35.256  H-C Bond
#  2   1   5   0   0     1.81930         1.81718         0.00212 40000.000   0.180  C-Cl Bond
#  2   1   6   0   0     3.08817         3.10099        -0.01282 10000.000   1.643  C-Cl Bond
#  3   2   1   3   0   111.01516       110.46605         0.54911     0.250   0.075  H-C-H Angle
#  3   2   1   4   0   110.91193       110.46605         0.44588     0.250   0.050  H-C-H Angle
#  3   2   1   5   0   108.14007       108.46079        -0.32072     0.250   0.026  H-C-Cl Angle
#  3   3   1   4   0   111.03893       110.48384         0.55510     0.250   0.077  H-C-H Angle
#  3   3   1   5   0   107.82850       108.44597        -0.61747     0.250   0.095  H-C-Cl Angle
#  3   4   1   5   0   107.75375       108.44597        -0.69222     0.250   0.120  H-C-Cl Angle
#  3   5   1   6   0   179.88685       179.91479        -0.02793     0.250   0.000  Cl-C-Cl Angle
#geometry from g03 output
#MOLECULE CLMECL_TS
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -69.46556   -602638.10526    602568.63970      0.000   0.000  Energy
# 17   4   0   0   0     0.19362        -0.54397         0.73758     0.000   0.000  C
# 17   5   0   0   0    -0.77055        -0.61181        -0.15875     0.000   0.000  Cl
# 17   6   0   0   0    -0.77063        -0.61181        -0.15882     0.000   0.000  Cl
#  7   0   0   0   0     0.00312         5.09780        -5.09468     0.000   0.000  Dipole
#  2   1   4   0   0     1.09732         1.07050         0.02682 40000.000  28.775  H-C Bond
#  2   2   4   0   0     1.09732         1.07050         0.02682 40000.000  28.773  H-C Bond
#  2   3   4   0   0     1.09739         1.07045         0.02693 40000.000  29.018  H-C Bond
#  2   4   5   0   0     2.29996         2.29996         0.00000 40000.000   0.000  C-Cl Bond
#  2   4   6   0   0     2.29996         2.29996         0.00000 40000.000   0.000  C-Cl Bond
#  3   1   4   2   0   119.94948       120.00174        -0.05226     0.250   0.001  H-C-H Angle
#  3   1   4   3   0   119.91053       119.99912        -0.08859     0.250   0.002  H-C-H Angle
#  3   2   4   3   0   120.13998       119.99912         0.14085     0.250   0.005  H-C-H Angle
#  3   5   4   6   0   179.95434       179.96911        -0.01477     0.250   0.000  Cl-C-Cl Angle
# 22   0   0   0   0     0.41707         0.00000         0.41707   100.000  17.395  CNORM
#geometry from g03 output
#MOLECULE CLMECL_TSX
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -69.46564   -602638.10526    602568.63962      0.000   0.000  /ClMeCl_1 Energy
#  2   1   2   0   0     2.30034         2.29996         0.00039 40000.000   0.006  C-Cl Bond
#  2   1   3   0   0     2.29996         2.29996         0.00000 40000.000   0.000  C-Cl Bond
# 17   1   0   0   0     0.19362        -0.54397         0.73759     4.000   2.176  C
# 17   2   0   0   0    -0.77094        -0.61181        -0.15913     4.000   0.101  Cl
# 17   3   0   0   0    -0.77033        -0.61181        -0.15851     4.000   0.101  Cl
#  7   0   0   0   0     0.00615         0.00190         0.00425     0.000   0.000  Dipole
```

```
# 22   0   0   0   0     0.40364         0.00000         0.40364   100.000  16.293  CNORM
#geometry from g03 output
#MOLECULE CLMECL_01
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -69.46575   -602638.10526    602568.63951     0.000   0.000  /ClMeCl_1 Energy
# 17   1   0   0   0     0.19362        -0.54397         0.73760     4.000   2.176  C
# 17   2   0   0   0    -0.77058        -0.61181        -0.15877     4.000   0.101  Cl
# 17   3   0   0   0    -0.77059        -0.61181        -0.15878     4.000   0.101  Cl
#  7   0   0   0   0     0.00165         0.00190        -0.00025     0.000   0.000  Dipole
# 22   0   0   0   0     0.57863         0.00000         0.57863   100.000  33.481  CNORM
#geometry from g03 output
#MOLECULE CLMECL_02
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -70.20286   -602638.57099    602568.36814     0.000   0.000  /ClMeCl_2 Energy
# 17   1   0   0   0     0.15629        -0.53747         0.69376     4.000   1.925  C
# 17   2   0   0   0    -0.85452        -0.68310        -0.17143     4.000   0.118  Cl
# 17   3   0   0   0    -0.63484        -0.53841        -0.09643     4.000   0.037  Cl
#  7   0   0   0   0     2.61836         1.91130         0.70706     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_03
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -71.61165   -602639.78943    602568.17778     0.000   0.000  /ClMeCl_3 Energy
# 17   1   0   0   0     0.12472        -0.49518         0.61991     4.000   1.537  C
# 17   2   0   0   0    -0.89033        -0.74215        -0.14818     4.000   0.088  Cl
# 17   3   0   0   0    -0.56671        -0.49760        -0.06912     4.000   0.019  Cl
#  7   0   0   0   0     3.80808         3.35730         0.45078     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_04
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -73.19716   -602641.39548    602568.19831     0.000   0.000  /ClMeCl_4 Energy
# 17   1   0   0   0     0.09891        -0.45105         0.54995     4.000   1.210  C
# 17   2   0   0   0    -0.91277        -0.78493        -0.12784     4.000   0.065  Cl
# 17   3   0   0   0    -0.52064        -0.47473        -0.04591     4.000   0.008  Cl
#  7   0   0   0   0     4.60645         4.29090         0.31555     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_05
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
```

```
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -74.78603   -602643.08982    602568.30379     0.000   0.000  /ClMeCl_5 Energy
# 17   1   0   0   0     0.07478        -0.41549         0.49027     4.000   0.961  C
# 17   2   0   0   0    -0.92947        -0.81688        -0.11260     4.000   0.051  Cl
# 17   3   0   0   0    -0.48279        -0.45830        -0.02449     4.000   0.002  Cl
#  7   0   0   0   0     5.25281         4.93850         0.31431     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_06
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -76.29437   -602644.70672    602568.41236     0.000   0.000  /ClMeCl_6 Energy
# 17   1   0   0   0     0.05605        -0.38907         0.44512     4.000   0.793  C
# 17   2   0   0   0    -0.94162        -0.84118        -0.10045     4.000   0.040  Cl
# 17   3   0   0   0    -0.45493        -0.44510        -0.00983     4.000   0.000  Cl
#  7   0   0   0   0     5.75875         5.41600         0.34275     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_07
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -77.67440   -602646.16487    602568.49047     0.000   0.000  /ClMeCl_7 Energy
# 17   1   0   0   0     0.03770        -0.36975         0.40744     4.000   0.664  C
# 17   2   0   0   0    -0.95158        -0.86037        -0.09121     4.000   0.033  Cl
# 17   3   0   0   0    -0.42938        -0.43351         0.00414     4.000   0.000  Cl
#  7   0   0   0   0     6.21549         5.80360         0.41189     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_08
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -78.90117   -602647.43740    602568.53623     0.000   0.000  /ClMeCl_8 Energy
# 17   1   0   0   0     0.02434        -0.35530         0.37964     4.000   0.577  C
# 17   2   0   0   0    -0.95927        -0.87614        -0.08313     4.000   0.028  Cl
# 17   3   0   0   0    -0.41049        -0.42312         0.01262     4.000   0.001  Cl
#  7   0   0   0   0     6.59497         6.13810         0.45687     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_09
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -79.95976   -602648.52286    602568.56311     0.000   0.000  /ClMeCl_9 Energy
# 17   1   0   0   0     0.01067        -0.34486         0.35553     4.000   0.506  C
# 17   2   0   0   0    -0.96580        -0.88914        -0.07666     4.000   0.024  Cl
# 17   3   0   0   0    -0.39235        -0.41364         0.02129     4.000   0.002  Cl
#  7   0   0   0   0     6.95182         6.43480         0.51702     0.000   0.000  Dipole
```

```
#geometry from g03 output
#MOLECULE CLMECL_10
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -80.84773   -602649.42466    602568.57693     0.000   0.000  /ClMeCl_10 Energy
# 17   1   0   0   0    -0.00153        -0.33748         0.33595     4.000   0.451  C
# 17   2   0   0   0    -0.97121        -0.89993        -0.07128     4.000   0.020  Cl
# 17   3   0   0   0    -0.37647        -0.40491         0.02844     4.000   0.003  Cl
#  7   0   0   0   0     7.27740         6.70310         0.57430     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_11
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -81.56932   -602650.15238    602568.58306     0.000   0.000  /ClMeCl_11 Energy
# 17   1   0   0   0    -0.01240        -0.33238         0.31998     4.000   0.410  C
# 17   2   0   0   0    -0.97572        -0.90904        -0.06668     4.000   0.018  Cl
# 17   3   0   0   0    -0.36248        -0.39683         0.03436     4.000   0.005  Cl
#  7   0   0   0   0     7.57767         6.95170         0.62597     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_12
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -82.13428   -602650.72505    602568.59077     0.000   0.000  /ClMeCl_12 Energy
# 17   1   0   0   0    -0.02203        -0.32893         0.30690     4.000   0.377  C
# 17   2   0   0   0    -0.97951        -0.91687        -0.06264     4.000   0.016  Cl
# 17   3   0   0   0    -0.35016        -0.38933         0.03917     4.000   0.006  Cl
#  7   0   0   0   0     7.85662         7.18660         0.67002     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_13
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -82.55643   -602651.16023    602568.60379     0.000   0.000  /ClMeCl_13 Energy
# 17   1   0   0   0    -0.03063        -0.32680         0.29617     4.000   0.351  C
# 17   2   0   0   0    -0.98270        -0.92357        -0.05913     4.000   0.014  Cl
# 17   3   0   0   0    -0.33919        -0.38229         0.04310     4.000   0.007  Cl
#  7   0   0   0   0     8.11850         7.41040         0.70810     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_14
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -82.85209   -602651.47216    602568.62007     0.000   0.000  /ClMeCl_14 Energy
```

```
# 17   1   0   0   0    -0.03822        -0.32566         0.28744     4.000   0.330  C
# 17   2   0   0   0    -0.98540        -0.92931        -0.05608     4.000   0.013  Cl
# 17   3   0   0   0    -0.32945        -0.37564         0.04619     4.000   0.009  Cl
#  7   0   0   0   0     8.36553         7.62500         0.74053     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_15
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -83.03882   -602651.67616    602568.63734     0.000   0.000  /ClMeCl_15 Energy
# 17   1   0   0   0    -0.04491        -0.32526         0.28035     4.000   0.314  C
# 17   2   0   0   0    -0.98768        -0.93425        -0.05342     4.000   0.011  Cl
# 17   3   0   0   0    -0.32080        -0.36937         0.04856     4.000   0.009  Cl
#  7   0   0   0   0     8.60000         7.83260         0.76740     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_16
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -83.13405   -602651.78843    602568.65438     0.000   0.000  /ClMeCl_16 Energy
# 17   1   0   0   0    -0.05047        -0.32552         0.27505     4.000   0.303  C
# 17   2   0   0   0    -0.98961        -0.93849        -0.05112     4.000   0.010  Cl
# 17   3   0   0   0    -0.31343        -0.36344         0.05001     4.000   0.010  Cl
#  7   0   0   0   0     8.82202         8.03580         0.78622     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_17
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -83.15476   -602651.82394    602568.66918     0.000   0.000  /ClMeCl_17 Energy
# 17   1   0   0   0    -0.05562        -0.32590         0.27028     4.000   0.292  C
# 17   2   0   0   0    -0.99125        -0.94220        -0.04905     4.000   0.010  Cl
# 17   3   0   0   0    -0.30657        -0.35780         0.05124     4.000   0.011  Cl
#  7   0   0   0   0     9.03663         8.23520         0.80143     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_18
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -83.11583   -602651.79495    602568.67912     0.000   0.000  /ClMeCl_18 Energy
# 17   1   0   0   0    -0.06014        -0.32627         0.26613     4.000   0.283  C
# 17   2   0   0   0    -0.99264        -0.94544        -0.04720     4.000   0.009  Cl
# 17   3   0   0   0    -0.30043        -0.35245         0.05201     4.000   0.011  Cl
#  7   0   0   0   0     9.24332         8.43080         0.81252     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_19
```

```
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -83.03018   -602651.71130    602568.68113     0.000   0.000  /ClMeCl_19 Energy
# 17   1   0   0   0    -0.06412        -0.32688         0.26276     4.000   0.276  C
# 17   2   0   0   0    -0.99382        -0.94820        -0.04563     4.000   0.008  Cl
# 17   3   0   0   0    -0.29493        -0.34734         0.05241     4.000   0.011  Cl
#  7   0   0   0   0     9.44322         8.62360         0.81962     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_20
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -82.90842   -602651.58235    602568.67393     0.000   0.000  /ClMeCl_20 Energy
# 17   1   0   0   0    -0.06764        -0.32756         0.25992     4.000   0.270  C
# 17   2   0   0   0    -0.99482        -0.95055        -0.04427     4.000   0.008  Cl
# 17   3   0   0   0    -0.28995        -0.34243         0.05248     4.000   0.011  Cl
#  7   0   0   0   0     9.63731         8.81390         0.82341     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_21
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -82.75890   -602651.41876    602568.65986     0.000   0.000  /ClMeCl_21 Energy
# 17   1   0   0   0    -0.07078        -0.32823         0.25745     4.000   0.265  C
# 17   2   0   0   0    -0.99567        -0.95259        -0.04308     4.000   0.007  Cl
# 17   3   0   0   0    -0.28544        -0.33773         0.05229     4.000   0.011  Cl
#  7   0   0   0   0     9.82631         9.00300         0.82331     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_22
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -82.58788   -602651.22894    602568.64105     0.000   0.000  ClMeCl_22 Energy
# 17   1   0   0   0    -0.07361        -0.32838         0.25477     4.000   0.260  C
# 17   2   0   0   0    -0.99638        -0.95443        -0.04195     4.000   0.007  Cl
# 17   3   0   0   0    -0.28129        -0.33327         0.05197     4.000   0.011  Cl
#  7   0   0   0   0    10.01108         9.19050         0.82058     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_23
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -82.39972   -602651.01878    602568.61906     0.000   0.000  ClMeCl_23 Energy
# 17   1   0   0   0    -0.07615        -0.32816         0.25201     4.000   0.254  C
# 17   2   0   0   0    -0.99698        -0.95605        -0.04093     4.000   0.007  Cl
```

```
# 17   3   0   0   0    -0.27750        -0.32902         0.05152     4.000   0.011  Cl
#  7   0   0   0   0    10.19190         9.37630         0.81560     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_24
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -82.19752   -602650.79269    602568.59518     0.000   0.000  ClMeCl_24 Energy
# 17   1   0   0   0    -0.07849        -0.32789         0.24940     4.000   0.249  C
# 17   2   0   0   0    -0.99748        -0.95745        -0.04003     4.000   0.006  Cl
# 17   3   0   0   0    -0.27398        -0.32495         0.05097     4.000   0.010  Cl
#  7   0   0   0   0    10.36948         9.56070         0.80878     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_25
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -81.98329   -602650.55549    602568.57221     0.000   0.000  ClMeCl_25 Energy
# 17   1   0   0   0    -0.08065        -0.32770         0.24704     4.000   0.244  C
# 17   2   0   0   0    -0.99790        -0.95867        -0.03923     4.000   0.006  Cl
# 17   3   0   0   0    -0.27069        -0.32104         0.05035     4.000   0.010  Cl
#  7   0   0   0   0    10.54421         9.74450         0.79971     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_26
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -81.75841   -602650.31258    602568.55418     0.000   0.000  ClMeCl_26 Energy
# 17   1   0   0   0    -0.08268        -0.32698         0.24430     4.000   0.239  C
# 17   2   0   0   0    -0.99826        -0.95982        -0.03843     4.000   0.006  Cl
# 17   3   0   0   0    -0.26760        -0.31732         0.04973     4.000   0.010  Cl
#  7   0   0   0   0    10.71645         9.92780         0.78865     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_27
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -81.52394   -602650.06817    602568.54423     0.000   0.000  ClMeCl_27 Energy
# 17   1   0   0   0    -0.08460        -0.32581         0.24121     4.000   0.233  C
# 17   2   0   0   0    -0.99855        -0.96093        -0.03763     4.000   0.006  Cl
# 17   3   0   0   0    -0.26465        -0.31379         0.04914     4.000   0.010  Cl
#  7   0   0   0   0    10.88663        10.11030         0.77633     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_28
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
```

```
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -81.28088   -602649.82419    602568.54331     0.000   0.000  ClMeCl_28 Energy
# 17   1   0   0   0    -0.08641        -0.32406         0.23765     4.000   0.226  C
# 17   2   0   0   0    -0.99880        -0.96198        -0.03682     4.000   0.005  Cl
# 17   3   0   0   0    -0.26187        -0.31046         0.04859     4.000   0.009  Cl
#  7   0   0   0   0    11.05477        10.29100         0.76377     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_29
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -81.03031   -602649.58097    602568.55066     0.000   0.000  ClMeCl_29 Energy
# 17   1   0   0   0    -0.08811        -0.32245         0.23434     4.000   0.220  C
# 17   2   0   0   0    -0.99900        -0.96294        -0.03606     4.000   0.005  Cl
# 17   3   0   0   0    -0.25923        -0.30724         0.04801     4.000   0.009  Cl
#  7   0   0   0   0    11.22112        10.47060         0.75052     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_30
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -80.77356   -602649.34070    602568.56714     0.000   0.000  ClMeCl_30 Energy
# 17   1   0   0   0    -0.08969        -0.32118         0.23149     4.000   0.214  C
# 17   2   0   0   0    -0.99917        -0.96385        -0.03532     4.000   0.005  Cl
# 17   3   0   0   0    -0.25675        -0.30412         0.04738     4.000   0.009  Cl
#  7   0   0   0   0    11.38578        10.65020         0.73558     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_31
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -80.51224   -602649.10582    602568.59358     0.000   0.000  ClMeCl_31 Energy
# 17   1   0   0   0    -0.09116        -0.31999         0.22883     4.000   0.209  C
# 17   2   0   0   0    -0.99931        -0.96476        -0.03456     4.000   0.005  Cl
# 17   3   0   0   0    -0.25441        -0.30113         0.04672     4.000   0.009  Cl
#  7   0   0   0   0    11.54889        10.83000         0.71889     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_32
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -80.24829   -602648.87872    602568.63043     0.000   0.000  ClMeCl_32 Energy
# 17   1   0   0   0    -0.09253        -0.31818         0.22565     4.000   0.204  C
# 17   2   0   0   0    -0.99943        -0.96567        -0.03376     4.000   0.005  Cl
# 17   3   0   0   0    -0.25221        -0.29835         0.04615     4.000   0.009  Cl
#  7   0   0   0   0    11.71058        11.00900         0.70158     0.000   0.000  Dipole
```

```
#geometry from g03 output
#MOLECULE CLMECL_33
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -79.98388   -602648.65834    602568.67446     0.000   0.000  ClMeCl_33 Energy
# 17   1   0   0   0    -0.09380        -0.31577         0.22197     4.000   0.197  C
# 17   2   0   0   0    -0.99953        -0.96657        -0.03296     4.000   0.004  Cl
# 17   3   0   0   0    -0.25014        -0.29578         0.04564     4.000   0.008  Cl
#  7   0   0   0   0    11.87097        11.18600         0.68497     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_34
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -79.72142   -602648.44380    602568.72238     0.000   0.000  ClMeCl_34 Energy
# 17   1   0   0   0    -0.09496        -0.31327         0.21831     4.000   0.191  C
# 17   2   0   0   0    -0.99961        -0.96743        -0.03218     4.000   0.004  Cl
# 17   3   0   0   0    -0.24820        -0.29333         0.04512     4.000   0.008  Cl
#  7   0   0   0   0    12.03013        11.36110         0.66903     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_35
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -79.46336   -602648.23452    602568.77116     0.000   0.000  ClMeCl_35 Energy
# 17   1   0   0   0    -0.09602        -0.31151         0.21549     4.000   0.186  C
# 17   2   0   0   0    -0.99968        -0.96829        -0.03139     4.000   0.004  Cl
# 17   3   0   0   0    -0.24640        -0.29084         0.04444     4.000   0.008  Cl
#  7   0   0   0   0    12.18813        11.53590         0.65223     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_36
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -79.21219   -602648.03196    602568.81977     0.000   0.000  ClMeCl_36 Energy
# 17   1   0   0   0    -0.09698        -0.31041         0.21343     4.000   0.182  C
# 17   2   0   0   0    -0.99973        -0.96917        -0.03057     4.000   0.004  Cl
# 17   3   0   0   0    -0.24473        -0.28835         0.04362     4.000   0.008  Cl
#  7   0   0   0   0    12.34503        11.71090         0.63413     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_37
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -78.97029   -602647.83718    602568.86689     0.000   0.000  ClMeCl_37 Energy
```

```
# 17   1   0   0   0    -0.09784        -0.30963         0.21179     4.000   0.179  C
# 17   2   0   0   0    -0.99978        -0.97007        -0.02971     4.000   0.004  Cl
# 17   3   0   0   0    -0.24317        -0.28591         0.04273     4.000   0.007  Cl
#  7   0   0   0   0    12.50095        11.88580         0.61515     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_38
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -78.73982   -602647.65044    602568.91062     0.000   0.000  ClMeCl_38 Energy
# 17   1   0   0   0    -0.09862        -0.30871         0.21010     4.000   0.177  C
# 17   2   0   0   0    -0.99982        -0.97094        -0.02888     4.000   0.003  Cl
# 17   3   0   0   0    -0.24173        -0.28364         0.04191     4.000   0.007  Cl
#  7   0   0   0   0    12.65600        12.05960         0.59640     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_39
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -78.52190   -602647.47028    602568.94838     0.000   0.000  ClMeCl_39 Energy
# 17   1   0   0   0    -0.09937        -0.30741         0.20804     4.000   0.173  C
# 17   2   0   0   0    -0.99985        -0.97173        -0.02812     4.000   0.003  Cl
# 17   3   0   0   0    -0.24049        -0.28159         0.04110     4.000   0.007  Cl
#  7   0   0   0   0    12.80992        12.23100         0.57892     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_40
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -78.32022   -602647.29489    602568.97467     0.000   0.000  ClMeCl_40 Energy
# 17   1   0   0   0    -0.09996        -0.30601         0.20605     4.000   0.170  C
# 17   2   0   0   0    -0.99988        -0.97246        -0.02742     4.000   0.003  Cl
# 17   3   0   0   0    -0.23911        -0.27970         0.04059     4.000   0.007  Cl
#  7   0   0   0   0    12.96381        12.40060         0.56321     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_41
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -78.13362   -602647.12389    602568.99028     0.000   0.000  ClMeCl_41 Energy
# 17   1   0   0   0    -0.10054        -0.30526         0.20472     4.000   0.168  C
# 17   2   0   0   0    -0.99990        -0.97321        -0.02668     4.000   0.003  Cl
# 17   3   0   0   0    -0.23792        -0.27777         0.03985     4.000   0.006  Cl
#  7   0   0   0   0    13.11675        12.57040         0.54635     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_42
```

```
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -77.96343   -602646.95854    602568.99511     0.000   0.000  ClMeCl_42 Energy
# 17   1   0   0   0    -0.10109        -0.30482         0.20374     4.000   0.166  C
# 17   2   0   0   0    -0.99992        -0.97400        -0.02592     4.000   0.003  Cl
# 17   3   0   0   0    -0.23678        -0.27585         0.03907     4.000   0.006  Cl
#  7   0   0   0   0    13.26915        12.74030         0.52885     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_43
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -77.80982   -602646.79941    602568.98959     0.000   0.000  ClMeCl_43 Energy
# 17   1   0   0   0    -0.10159        -0.30495         0.20336     4.000   0.165  C
# 17   2   0   0   0    -0.99993        -0.97481        -0.02513     4.000   0.003  Cl
# 17   3   0   0   0    -0.23570        -0.27392         0.03822     4.000   0.006  Cl
#  7   0   0   0   0    13.42108        12.91020         0.51088     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_44
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -77.67251   -602646.64648    602568.97398     0.000   0.000  ClMeCl_44 Energy
# 17   1   0   0   0    -0.10207        -0.30535         0.20328     4.000   0.165  C
# 17   2   0   0   0    -0.99994        -0.97558        -0.02436     4.000   0.002  Cl
# 17   3   0   0   0    -0.23466        -0.27205         0.03739     4.000   0.006  Cl
#  7   0   0   0   0    13.57264        13.07930         0.49334     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_45
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -77.55084   -602646.49902    602568.94818     0.000   0.000  ClMeCl_45 Energy
# 17   1   0   0   0    -0.10254        -0.30581         0.20327     4.000   0.165  C
# 17   2   0   0   0    -0.99995        -0.97628        -0.02368     4.000   0.002  Cl
# 17   3   0   0   0    -0.23363        -0.27031         0.03668     4.000   0.005  Cl
#  7   0   0   0   0    13.72390        13.24670         0.47720     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_46
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -77.44383   -602646.35532    602568.91149     0.000   0.000  ClMeCl_46 Energy
# 17   1   0   0   0    -0.10300        -0.30643         0.20343     4.000   0.166  C
# 17   2   0   0   0    -0.99996        -0.97686        -0.02310     4.000   0.002  Cl
```

```
# 17   3   0   0   0    -0.23263        -0.26870         0.03606     4.000   0.005  Cl
#  7   0   0   0   0    13.87488        13.41200         0.46288     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_47
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -77.35024   -602646.21457    602568.86433     0.000   0.000  ClMeCl_47 Energy
# 17   1   0   0   0    -0.10346        -0.30756         0.20411     4.000   0.167  C
# 17   2   0   0   0    -0.99997        -0.97742        -0.02255     4.000   0.002  Cl
# 17   3   0   0   0    -0.23164        -0.26709         0.03545     4.000   0.005  Cl
#  7   0   0   0   0    14.02565        13.57680         0.44885     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_48
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -77.26861   -602646.07746    602568.80885     0.000   0.000  ClMeCl_48 Energy
# 17   1   0   0   0    -0.10391        -0.30916         0.20525     4.000   0.169  C
# 17   2   0   0   0    -0.99997        -0.97804        -0.02193     4.000   0.002  Cl
# 17   3   0   0   0    -0.23066        -0.26547         0.03481     4.000   0.005  Cl
#  7   0   0   0   0    14.17621        13.74260         0.43361     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_49
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -77.19736   -602645.94530    602568.74794     0.000   0.000  ClMeCl_49 Energy
# 17   1   0   0   0    -0.10437        -0.31073         0.20636     4.000   0.170  C
# 17   2   0   0   0    -0.99998        -0.97868        -0.02130     4.000   0.002  Cl
# 17   3   0   0   0    -0.22968        -0.26393         0.03425     4.000   0.005  Cl
#  7   0   0   0   0    14.32660        13.90860         0.41800     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_50
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -77.13484   -602645.81804    602568.68321     0.000   0.000  ClMeCl_50 Energy
# 17   1   0   0   0    -0.10483        -0.31222         0.20739     4.000   0.172  C
# 17   2   0   0   0    -0.99998        -0.97929        -0.02069     4.000   0.002  Cl
# 17   3   0   0   0    -0.22870        -0.26248         0.03378     4.000   0.005  Cl
#  7   0   0   0   0    14.47685        14.07400         0.40285     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_51
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
```

```
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -77.07936   -602645.69530    602568.61595     0.000   0.000  ClMeCl_51 Energy
# 17   1   0   0   0    -0.10529        -0.31387         0.20857     4.000   0.174  C
# 17   2   0   0   0    -0.99999        -0.97988        -0.02010     4.000   0.002  Cl
# 17   3   0   0   0    -0.22773        -0.26108         0.03335     4.000   0.004  Cl
#  7   0   0   0   0    14.62696        14.23890         0.38806     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_52
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -76.85121   -602645.13011    602568.27890     0.000   0.000  ClMeCl_5252 Energ
# 17   1   0   0   0    -0.10768        -0.32489         0.21721     4.000   0.189  C
# 17   2   0   0   0    -0.99999        -0.98234        -0.01765     4.000   0.001  Cl
# 17   3   0   0   0    -0.22285        -0.25440         0.03155     4.000   0.004  Cl
#  7   0   0   0   0    15.37560        15.05060         0.32500     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_53
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -76.69952   -602644.82802    602568.12851     0.000   0.000  ClMeCl_5353 Energ
# 17   1   0   0   0    -0.10763        -0.33278         0.22516     4.000   0.203  C
# 17   2   0   0   0    -1.00000        -0.98384        -0.01615     4.000   0.001  Cl
# 17   3   0   0   0    -0.22154        -0.25071         0.02917     4.000   0.003  Cl
#  7   0   0   0   0    15.81757        15.53530         0.28227     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_54
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -76.51236   -602644.55160    602568.03925     0.000   0.000  ClMeCl_5454 Energ
# 17   1   0   0   0    -0.11052        -0.33933         0.22880     4.000   0.209  C
# 17   2   0   0   0    -1.00000        -0.98514        -0.01486     4.000   0.001  Cl
# 17   3   0   0   0    -0.21718        -0.24743         0.03025     4.000   0.004  Cl
#  7   0   0   0   0    16.26949        16.01450         0.25499     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_55
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -76.26415   -602644.29828    602568.03413     0.000   0.000  ClMeCl_5555 Energ
# 17   1   0   0   0    -0.11187        -0.34595         0.23408     4.000   0.219  C
# 17   2   0   0   0    -1.00000        -0.98655        -0.01345     4.000   0.001  Cl
# 17   3   0   0   0    -0.21449        -0.24409         0.02960     4.000   0.004  Cl
#  7   0   0   0   0    16.71451        16.49210         0.22241     0.000   0.000  Dipole
```

```
#geometry from g03 output
#MOLECULE CLMECL_56
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -75.95952   -602644.06453    602568.10501     0.000   0.000  ClMeCl_5656 Energ
# 17   1   0   0   0    -0.11314        -0.35230         0.23915     4.000   0.229  C
# 17   2   0   0   0    -1.00000        -0.98794        -0.01206     4.000   0.001  Cl
# 17   3   0   0   0    -0.21193        -0.24087         0.02894     4.000   0.003  Cl
#  7   0   0   0   0    17.15821        16.96740         0.19081     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_57
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -75.61460   -602643.85017    602568.23558     0.000   0.000  ClMeCl_5757 Energ
# 17   1   0   0   0    -0.11454        -0.35762         0.24307     4.000   0.236  C
# 17   2   0   0   0    -1.00000        -0.98923        -0.01077     4.000   0.000  Cl
# 17   3   0   0   0    -0.20917        -0.23802         0.02886     4.000   0.003  Cl
#  7   0   0   0   0    17.60195        17.44110         0.16085     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_58
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -75.25550   -602643.65527    602568.39977     0.000   0.000  ClMeCl_5858 Energ
# 17   1   0   0   0    -0.11554        -0.36099         0.24546     4.000   0.241  C
# 17   2   0   0   0    -1.00000        -0.99047        -0.00953     4.000   0.000  Cl
# 17   3   0   0   0    -0.20702        -0.23527         0.02826     4.000   0.003  Cl
#  7   0   0   0   0    18.04282        17.91180         0.13102     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_59
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -74.90159   -602643.47536    602568.57377     0.000   0.000  ClMeCl_5959 Energ
# 17   1   0   0   0    -0.11330        -0.36511         0.25181     4.000   0.254  C
# 17   2   0   0   0    -1.00000        -0.99173        -0.00827     4.000   0.000  Cl
# 17   3   0   0   0    -0.20886        -0.23257         0.02371     4.000   0.002  Cl
#  7   0   0   0   0    18.46832        18.38310         0.08522     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_60
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -74.60452   -602643.30763    602568.70311     0.000   0.000  ClMeCl_6060 Energ
```

```
# 17   1   0   0   0    -0.11708        -0.36785         0.25077     4.000   0.252  C
# 17   2   0   0   0    -1.00000        -0.99278        -0.00722     4.000   0.000  Cl
# 17   3   0   0   0    -0.20341        -0.23010         0.02669     4.000   0.003  Cl
#  7   0   0   0   0    18.92048        18.85030         0.07018     0.000   0.000  Dipole
#geometry from g03 output
#MOLECULE CLMECL_61
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -74.34715   -602643.15320    602568.80605     0.000   0.000  ClMeCl_6161 Energ
# 17   1   0   0   0    -0.11769        -0.37030         0.25262     4.000   0.255  C
# 17   2   0   0   0    -1.00000        -0.99379        -0.00621     4.000   0.000  Cl
# 17   3   0   0   0    -0.20189        -0.22779         0.02590     4.000   0.003  Cl
#  7   0   0   0   0    19.35759        19.31760         0.03999     0.000   0.000  Dipole
#geometry not relevent
#MOLECULE CLATOM
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0    28.99000   -288738.71251    288767.70251     0.000   0.000  ClAtom Energy
#geometry from g03 output
#MOLECULE CLANION
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED       OBSERVED       DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE           VALUE         CALC-OBS      WEIGHT  CHISQ   PROPERTY
#  1   0   0   0   0   -54.66441       -54.30000        -0.36441     1.000   0.133  Expt HOF
#  1   0   0   0   0   -54.66441   -288822.85783    288768.19342     0.000   0.000  ClAnion Energy
#Dummy Calculation
#MOLECULE DUMMY H2
#COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS.
#PROPERTY DEFINITION  CALCULATED   OBSERVED     DEVIATION
#ID1 ID4 ID5 ID6 ID7    VALUE       VALUE        CALC-OBS    WEIGHT  CHISQ   PROPERTY
# 11  69  70   0   0    83.65441    83.40000      0.25441    44.444   2.877  0 EA Cl
# 11  69   1   0   0    57.85781    59.34312     -1.48531     0.000   0.000  BDE Cl2 - DFT
# 11  69   1   0   0    57.85781    58.10000     -0.24219    16.000   0.939  0 BDE Cl2
# 11   6   5   0   0    13.68565    13.71900     -0.03335   400.000   0.445  Rel. E. CMPLX -> TS
# 11   5  70   4   0   -11.64392   -11.68926      0.04533   400.000   0.822  Rel. E. R -> CMPLX
# 11   6  70   4   0     2.04172     2.02974      0.01198   400.000   0.057  Rel. E. R -> TS
# 11   8   9   0   0     0.73711     0.46574      0.27137   100.000   7.364  Rel. E. 02 -> 01  C-Cl: 2.35 A
# 11   8  10   0   0     2.14590     1.68417      0.46173   100.000  21.319  Rel. E. 03 -> 01  C-Cl: 2.40 A
# 11   8  11   0   0     3.73142     3.29022      0.44119   100.000  19.465  Rel. E. 04 -> 01  C-Cl: 2.45 A
# 11   8  12   0   0     5.32028     4.98456      0.33572   100.000  11.270  Rel. E. 05 -> 01  C-Cl: 2.50 A
# 11   8  13   0   0     6.82862     6.60147      0.22715   100.000   5.160  Rel. E. 06 -> 01  C-Cl: 2.55 A
# 11   8  14   0   0     8.20866     8.05961      0.14904   100.000   2.221  Rel. E. 07 -> 01  C-Cl: 2.60 A
# 11   8  15   0   0     9.43542     9.33214      0.10328   100.000   1.067  Rel. E. 08 -> 01  C-Cl: 2.65 A
# 11   8  16   0   0    10.49401    10.41761      0.07640   100.000   0.584  Rel. E. 09 -> 01  C-Cl: 2.70 A
```

```
# 11   8  17   0   0    11.38198    11.31940     0.06258   100.000   0.392  Rel. E. 10 -> 01  C-Cl: 2.75 A
# 11   8  18   0   0    12.10357    12.04713     0.05645   100.000   0.319  Rel. E. 11 -> 01  C-Cl: 2.80 A
# 11   8  19   0   0    12.66853    12.61979     0.04874   100.000   0.238  Rel. E. 12 -> 01  C-Cl: 2.85 A
# 11   8  20   0   0    13.09068    13.05497     0.03572   100.000   0.128  Rel. E. 13 -> 01  C-Cl: 2.90 A
# 11   8  21   0   0    13.38635    13.36690     0.01944   100.000   0.038  Rel. E. 14 -> 01  C-Cl: 2.95 A
# 11   8  22   0   0    13.57307    13.57091     0.00217   100.000   0.000  Rel. E. 15 -> 01  C-Cl: 3.00 A
# 11   8  23   0   0    13.66830    13.68317    -0.01487   100.000   0.022  Rel. E. 16 -> 01  C-Cl: 3.05 A
# 11   8  24   0   0    13.68901    13.71869    -0.02967   100.000   0.088  Rel. E. 17 -> 01  C-Cl: 3.10 A
# 11   8  25   0   0    13.65008    13.68970    -0.03961   100.000   0.157  Rel. E. 18 -> 01  C-Cl: 3.15 A
# 11   8  26   0   0    13.56443    13.60605    -0.04162   100.000   0.173  Rel. E. 19 -> 01  C-Cl: 3.20 A
# 11   8  27   0   0    13.44267    13.47710    -0.03442   100.000   0.118  Rel. E. 20 -> 01  C-Cl: 3.25 A
# 11   8  28   0   0    13.29315    13.31350    -0.02035    25.000   0.010  Rel. E. 21 -> 01  C-Cl: 3.30 A
# 11   8  29   0   0    13.12214    13.12368    -0.00154    25.000   0.000  Rel. E. 22 -> 01  C-Cl: 3.35 A
# 11   8  30   0   0    12.93397    12.91353     0.02044    25.000   0.010  Rel. E. 23 -> 01  C-Cl: 3.40 A
# 11   8  31   0   0    12.73177    12.68744     0.04433    25.000   0.049  Rel. E. 24 -> 01  C-Cl: 3.45 A
# 11   8  32   0   0    12.51754    12.45024     0.06730    25.000   0.113  Rel. E. 25 -> 01  C-Cl: 3.50 A
# 11   8  33   0   0    12.29266    12.20733     0.08533    25.000   0.182  Rel. E. 26 -> 01  C-Cl: 3.55 A
# 11   8  34   0   0    12.05820    11.96291     0.09528    25.000   0.227  Rel. E. 27 -> 01  C-Cl: 3.60 A
# 11   8  35   0   0    11.81514    11.71894     0.09620    25.000   0.231  Rel. E. 28 -> 01  C-Cl: 3.65 A
# 11   8  36   0   0    11.56457    11.47571     0.08885    25.000   0.197  Rel. E. 29 -> 01  C-Cl: 3.70 A
# 11   8  37   0   0    11.30781    11.23544     0.07237    25.000   0.131  Rel. E. 30 -> 01  C-Cl: 3.75 A
# 11   8  38   0   0    11.04649    11.00056     0.04593    25.000   0.053  Rel. E. 31 -> 01  C-Cl: 3.80 A
# 11   8  39   0   0    10.78255    10.77347     0.00908    25.000   0.002  Rel. E. 32 -> 01  C-Cl: 3.85 A
# 11   8  40   0   0    10.51814    10.55309    -0.03495    25.000   0.031  Rel. E. 33 -> 01  C-Cl: 3.90 A
# 11   8  41   0   0    10.25567    10.33854    -0.08287    25.000   0.172  Rel. E. 34 -> 01  C-Cl: 3.95 A
# 11   8  42   0   0     9.99761    10.12927    -0.13166    25.000   0.433  Rel. E. 35 -> 01  C-Cl: 4.00 A
# 11   8  43   0   0     9.74644     9.92671    -0.18026    25.000   0.812  Rel. E. 36 -> 01  C-Cl: 4.05 A
# 11   8  44   0   0     9.50454     9.73193    -0.22739    25.000   1.293  Rel. E. 37 -> 01  C-Cl: 4.10 A
# 11   8  45   0   0     9.27407     9.54518    -0.27111    25.000   1.837  Rel. E. 38 -> 01  C-Cl: 4.15 A
# 11   8  46   0   0     9.05615     9.36502    -0.30887    25.000   2.385  Rel. E. 39 -> 01  C-Cl: 4.20 A
# 11   8  47   0   0     8.85447     9.18963    -0.33516    25.000   2.808  Rel. E. 40 -> 01  C-Cl: 4.25 A
# 11   8  48   0   0     8.66787     9.01864    -0.35077    25.000   3.076  Rel. E. 41 -> 01  C-Cl: 4.30 A
# 11   8  49   0   0     8.49769     8.85329    -0.35560    25.000   3.161  Rel. E. 42 -> 01  C-Cl: 4.35 A
# 11   8  50   0   0     8.34407     8.69415    -0.35008    25.000   3.064  Rel. E. 43 -> 01  C-Cl: 4.40 A
# 11   8  51   0   0     8.20676     8.54123    -0.33447    25.000   2.797  Rel. E. 44 -> 01  C-Cl: 4.45 A
# 11   8  52   0   0     8.08509     8.39376    -0.30867    25.000   2.382  Rel. E. 45 -> 01  C-Cl: 4.50 A
# 11   8  53   0   0     7.97808     8.25006    -0.27198    25.000   1.849  Rel. E. 46 -> 01  C-Cl: 4.55 A
# 11   8  54   0   0     7.88449     8.10931    -0.22482    25.000   1.264  Rel. E. 47 -> 01  C-Cl: 4.60 A
# 11   8  55   0   0     7.80286     7.97220    -0.16934    25.000   0.717  Rel. E. 48 -> 01  C-Cl: 4.65 A
# 11   8  56   0   0     7.73162     7.84005    -0.10843    25.000   0.294  Rel. E. 49 -> 01  C-Cl: 4.70 A
# 11   8  57   0   0     7.66909     7.71279    -0.04370    25.000   0.048  Rel. E. 50 -> 01  C-Cl: 4.75 A
# 11   8  58   0   0     7.61361     7.59005     0.02356    25.000   0.014  Rel. E. 51 -> 01  C-Cl: 4.80 A
# 11   8  59   0   0     7.38546     7.02485     0.36061    25.000   3.251  Rel. E. 52 -> 01  C-Cl: 5.05 A
# 11   8  60   0   0     7.23377     6.72277     0.51100    25.000   6.528  Rel. E. 53 -> 01  C-Cl: 5.20 A
# 11   8  61   0   0     7.04661     6.44635     0.60026    25.000   9.008  Rel. E. 54 -> 01  C-Cl: 5.35 A
# 11   8  62   0   0     6.79840     6.19302     0.60538    25.000   9.162  Rel. E. 55 -> 01  C-Cl: 5.50 A
```

```
# 11   8  63   0   0     6.49377     5.95927      0.53450    25.000   7.142  Rel. E. 56 -> 01  C-Cl: 5.65 A
# 11   8  64   0   0     6.14885     5.74492      0.40393    25.000   4.079  Rel. E. 57 -> 01  C-Cl: 5.80 A
# 11   8  65   0   0     5.78975     5.55001      0.23974    25.000   1.437  Rel. E. 58 -> 01  C-Cl: 5.95 A
# 11   8  66   0   0     5.43584     5.37011      0.06574    25.000   0.108  Rel. E. 59 -> 01  C-Cl: 6.10 A
# 11   8  67   0   0     5.13877     5.20237     -0.06360    25.000   0.101  Rel. E. 60 -> 01  C-Cl: 6.25 A
# 11   8  68   0   0     4.88140     5.04794     -0.16654    25.000   0.693  Rel. E. 61 -> 01  C-Cl: 6.40 A
```

Appendix C

Detailed Derivatives for the Expressions in the QXD Model

Repulsive interaction terms

Differentiation of ζ_{ij} with respect to ρ_{nm}

Recall[172]

$$\zeta_{ij} = \frac{\zeta_i(q_i)^3\zeta_j(q_j)^3}{8\pi[\zeta_i(q_i)^2 - \zeta_j(q_j)^2]^3}$$

It should be noted that only q_i, the charge of the atom in the QM region depends on ρ_{nm}. This means all terms involving q_j can be treated as constants. For ease of writting let:

$$\zeta_i(0) = \varsigma \tag{C.1}$$

and

$$\frac{\zeta_j(q_j)^3\varsigma^3}{8\pi} = Q$$

Therefore ζ_{ij} can be stated as

$$\zeta_{ij} = Q\frac{e^{3\zeta_{q,i}q_i}}{\{\varsigma^2 e^{2\zeta_{q,i}q_i} - \zeta_j(q_j)^2\}^3}.$$

Differentiating ζ_{ij} with respect to ρ_{nm}

$$\begin{aligned}
\frac{\partial \zeta_{ij}}{\partial \rho_{nm}} &= Q\left[\frac{3\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)e^{3\zeta_{q,i}q_i}(\varsigma^2 e^{2\zeta_{q,i}q_i}-\zeta_j(q_j)^2)^3}{(\varsigma^2 e^{2\zeta_{q,i}q_i}-\zeta_j(q_j)^2)^6}\right. \\
&\quad \left.-\frac{6\varsigma^2\zeta_{q,i}(\left(\frac{\partial q_i}{\partial \rho_{nm}}\right))e^{2\zeta_{q,i}}(\varsigma^2 e^{2\zeta_{q,i}q_i}-\zeta_j(q_j)^2)^2 e^{3\zeta_{q,i}q_i}}{(\varsigma^2 e^{2\zeta_{q,i}q_i}-\zeta_j(q_j)^2)^6}\right] \\
&= Q\frac{3\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)e^{3\zeta_{q,i}q_i}(\varsigma^2 e^{2\zeta_{q,i}q_i}-\zeta_j(q_j)^2)^2\{\varsigma^2 e^{2\zeta_{q,i}q_i}-\zeta_j(q_j)^2-2\varsigma^2 e^{2\zeta_{q,i}q_i}\}}{(\varsigma^2 e^{2\zeta_{q,i}q_i}-\zeta_j(q_j)^2)^6} \\
&= Q\frac{3\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)e^{3\zeta_{q,i}q_i}\{-\zeta_j(q_j)^2-\varsigma^2 e^{2\zeta_{q,i}q_i}\}}{(\varsigma^2 e^{2\zeta_{q,i}q_i}-\zeta_j(q_j)^2)^4} \\
&= -Q\frac{3\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)e^{3\zeta_{q,i}q_i}\{\zeta_j(q_j)^2+\varsigma^2 e^{2\zeta_{q,i}q_i}\}}{(\zeta_j(q_j)^2-\varsigma^2 e^{2\zeta_{q,i}q_i})^4}
\end{aligned}$$

Now, susbstituting the defined terms above back into the expression

$$\frac{\partial \zeta_{ij}}{\partial \rho_{nm}} = \left(-\frac{\zeta_j^3(q_j)\varsigma^3}{8\pi}\right)\frac{3\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)e^{3\zeta_{q,i}q_i}\{\zeta_j^2(q_j)+\zeta_i^2(q_i)\}}{(\zeta_j^2(q_j)-\zeta_i^2(q_i))^4}$$

Then upon rearrangement

$$\frac{\partial \zeta_{ij}}{\partial \rho_{nm}} = -3\zeta_i(q_i)^3\zeta_j(q_j)^3\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)\frac{\zeta_i(q_i)^2+\zeta_j(q_j)^2}{8\pi(\zeta_j(q_j)^2-\zeta_i(q_i)^2)^4} \tag{C.2}$$

Differentiation of Δ_{ij} with respect to ρ_{nm}

Recall[172]

$$\Delta_{ij} = \frac{\zeta_j(q_j)e^{-\zeta_i(q_i)r_{ij}}}{r_{ij}}\left(4\zeta_i(q_i)+r_{ij}[\zeta_i(q_i)^2-\zeta_j(q_j)^2]\right).$$

Using Eqn. (C.1) and differentiating Δ_{ij} with respect to ρ_{nm} one can state

$$\begin{aligned}
\frac{\partial \Delta_{ij}}{\partial \rho_{nm}} &= -\zeta_j(q_j)\varsigma\{e^{-r_{ij}\varsigma e^{\zeta_{q,i}q_i}}\}\{e^{\zeta_{q,i}q_i}\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)\}[4\varsigma e^{\zeta_{q,i}q_i} + r_{ij}\varsigma^2 e^{2\zeta_{q,i}q_i} - r_{ij}\zeta_j(q_j)^2] \\
&\quad + \frac{\zeta_j(q_j)\{e^{-r_{ij}\varsigma e^{\zeta_{q,i}q_i}}\}}{r_{ij}}[4\varsigma\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)e^{\zeta_{q,i}q_i} + 2r_{ij}\varsigma^2\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)e^{\zeta_{q,i}q_i}] \\
&= \frac{\zeta_j(q_j)\varsigma e^{\zeta_{q,i}q_i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)\zeta_{q,i}\{e^{-r_{ij}\varsigma e^{\zeta_{q,i}q_i}}\}}{r_{ij}}\Big[-\{4\varsigma r_{ij}e^{\zeta_{q,i}q_i} + \varsigma^2 r_{ij}^2 e^{2\zeta_{q,i}q_i} - r_{ij}\zeta_j(q_j)^2\} \\
&\quad -\{4 + 2r_{ij}\varsigma e^{\zeta_{q,i}q_i}\}\Big] \\
&= \frac{\zeta_j(q_j)\varsigma e^{\zeta_{q,i}q_i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)\zeta_{q,i}\{e^{-r_{ij}\varsigma e^{\zeta_{q,i}q_i}}\}}{r_{ij}}[-2r_{ij}\varsigma e^{\zeta_{q,i}q_i} - r_{ij}^2\varsigma^2 e^{2\zeta_{q,i}q_i} + r_{ij}^2\zeta_j(q_j)^2 + 4] \\
&= -\frac{\zeta_j(q_j)\varsigma e^{\zeta_{q,i}q_i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)\zeta_{q,i}\{e^{-r_{ij}\varsigma e^{\zeta_{q,i}q_i}}\}}{r_{ij}}[2r_{ij}\varsigma e^{\zeta_{q,i}q_i} + r_{ij}^2\{\varsigma^2 e^{2\zeta_{q,i}q_i} - \zeta_j(q_j)^2\} - 4]
\end{aligned}$$

Substituting back in the definition of ς

$$\frac{\partial \Delta_{ij}}{\partial \rho_{nm}} = -\frac{\zeta_j(q_j)\zeta_i(q_i)\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)\zeta_{q,i}e^{-\zeta_i(q_i)r_{ij}}}{r_{ij}}[2r_{ij}\zeta_i(q_i) + r_{ij}^2\{\zeta_i(q_i)^2 - \zeta_j(q_j)^2\} - 4]$$

And then upon rearrangement

$$\begin{aligned}
\frac{\partial \Delta_{ij}}{\partial \rho_{nm}} &= -\zeta_i(q_i)\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)\left[\frac{2\zeta_j(q_j)e^{-\zeta_i(q_i)r_{ij}}}{r_{ij}}(-\zeta_i(q_i)r_{ij} - 2) + \Delta_{ij}r_{ij}\right] \\
&= \zeta_i(q_i)\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)\left[\frac{2\zeta_j(q_j)e^{-\zeta_i(q_i)r_{ij}}}{r_{ij}}(\zeta_i(q_i)r_{ij} + 2) - \Delta_{ij}r_{ij}\right] \qquad \text{(C.3)}
\end{aligned}$$

Differentiation of Δ_{ji} with respect to ρ_{nm}

Recall[172]

$$\Delta_{ji} = \frac{\zeta_i(q_i)e^{-\zeta_j(q_j)r_{ij}}}{r_{ij}}\left(4\zeta_j(q_j) + r_{ij}[\zeta_j(q_j)^2 - \zeta_i(q_i)^2]\right).$$

Now differentiating Δ_{ji} with respect to ρ_{nm} one can state, using the definition stated in Eqn. (C.1)

$$\begin{aligned}
\frac{\partial \Delta_{ji}}{\partial \rho_{nm}} &= \frac{\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)\varsigma e^{\zeta_{q,i}q_i}e^{-\zeta_j(q_j r_{ji}}}{r_{ji}}\{4\zeta_j(q_j + r_{ji}\zeta_j(q_j^2 - r_{ji}\varsigma^2 e^{2\zeta_{q,i}q_i}\} \\
&\quad - \frac{\varsigma e^{\zeta_{i,q}q_i}e^{-\zeta_j(q_j r_{ji}}}{r_{ji}}\{2r_{ji}\varsigma^2\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)e^{2\zeta_{q,i}q_i}\} \\
&= \frac{\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)\varsigma e^{\zeta_{q,i}q_i - \zeta_j(q_j r_{ji}}}{r_{ji}}\{4\zeta_j(q_j + r_{ji}\zeta_j(q_j^2 - r_{ji}\varsigma^2 e^{2\zeta_{q,i}q_i} - 2r_{ji}\varsigma^2 e^{2\zeta_{q,i}q_i}\} \\
&= \frac{\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)\varsigma e^{\zeta_{q,i}q_i - \zeta_j(q_j r_{ji}}}{r_{ji}}\{4\zeta_j(q_j + r_{ji}\zeta_j(q_j^2 - 3r_{ji}\varsigma^2 e^{2\zeta_{q,i}q_i}\}
\end{aligned}$$

Substituting back in the definition of ς

$$\frac{\partial \Delta_{ji}}{\partial \rho_{nm}} = \frac{\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)\zeta_i(q_i)e^{-\zeta_j(q_j)r_{ji}}}{r_{ji}}[4\zeta_j(q_j) + r_{ji}\{\zeta_j(q_j)^2 - 3\zeta_i(q_i)^2\}]$$

and upon rearrangement

$$\frac{\partial \Delta_{ji}}{\partial \rho_{nm}} = \zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)\left[\Delta_{ji} - 2e^{-\zeta_j(q_j)r_{ji}}\zeta_i(q_i)^3\right]. \tag{C.4}$$

Dispersion Potential Terms

Differentiation of Δ'_{ij} with respect to ρ_{nm}

Recall[172]

$$\Delta'_{ij} = \zeta_j(q_j)\frac{e^{-\zeta_i(q_i)r_{ij}}}{r_{ij}}[\zeta_i(q_i)^2 - \zeta_j(q_j)^2] - \Delta_{ij}\left(\frac{1}{r_{ij}} + \zeta_i(q_i)\right)$$

Now, differentiating Δ'_{ij} with respect to ρ_{nm}:

$$\begin{aligned}
\frac{\partial \Delta'_{ij}}{\partial \rho_{nm}} &= \frac{\zeta_j(q_j)}{r_{ij}}\left[\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)\left(-r_{ij}\zeta_i(q_i)e^{-\zeta_i(q_i)r_{ij}}(\zeta_i(q_i)^2-\zeta_j(q_j)^2)+2e^{-\zeta_i(q_i)r_{ij}}\zeta_i(q_i)^2\right)\right] \\
&\quad -\left(d\Delta_{ij}\left[\frac{1}{r_{ij}}+\zeta_i(q_i)\right]+\Delta_{ij}\left[\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)\zeta_i(q_i)\right]\right) \\
&= \frac{\zeta_i(q_i)\zeta_j(q_j)\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)e^{-\zeta_i(q_i)r_{ij}}}{r_{ij}}\left[2\zeta_i(q_i)+r_{ij}\left(\zeta_j(q_j)^2-\zeta_i(q_i)^2\right)\right] \\
&\quad -d\Delta_{ij}\left[\frac{1}{r_{ij}}+\zeta_i(q_i)\right]-\Delta_{ij}\zeta_i(q_i)\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right) \\
&= \frac{d\Delta_{ij}+4\zeta_i(q_i)\zeta_j(q_j)\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)e^{-\zeta_i(q_i)r_{ij}}}{r_{ij}}\left(\zeta_i(q_i)-\frac{1}{r_{ij}}\right)-d\Delta_{ij}\left[\frac{1}{r_{ij}}+\zeta_i(q_i)\right] \\
&= \frac{4\zeta_i(q_i)\zeta_j(q_j)\zeta_{q,i}(dq_i)e^{-\zeta_i(q_i)r_{ij}}}{r_{ij}}\left(\zeta_i(q_i)-\frac{1}{r_{ij}}\right) \\
&\quad -\zeta_i(q_i)\left(d\Delta_{ij}+\Delta_{ij}\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)\right) \qquad \text{(C.5)}
\end{aligned}$$

where $d\Delta_{ij}$ would represent $\left(\frac{\partial \Delta_{ij}}{\partial \rho_{nm}}\right)$. To find the derivation of this term please refer to earlier derivations.

Differentiation of Δ'_{ji} with respect to ρ_{nm}

Recall[172]

$$\Delta'_{ji}=\zeta_i(q_j)\frac{e^{-\zeta_j(q_j)r_{ij}}}{r_{ij}}[\zeta_j(q_j)^2-\zeta_i(q_i)^2]-\Delta_{ji}\left(\frac{1}{r_{ij}}+\zeta_j(q_j)\right)$$

Now, differentiating Δ'_{ij} with respect to ρ_{nm}:

$$\begin{aligned}
\frac{\partial \Delta'_{ji}}{\partial \rho_{nm}} &= \frac{e^{-\zeta_j(q_j)r_{ji}}}{r_{ji}}\left[\zeta_i(q_i)\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)\left(\zeta_j(q_j)^2-\zeta_i(q_i)^2\right)-\zeta_i(q_i)\left\{2\zeta_1(q_i)^2\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)\right\}\right] \\
&\quad - d\Delta_{ji}\left[\frac{1}{r_{ji}}+\zeta_j(q_j)\right] \\
&= \frac{\zeta_i(q_i)\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)e^{-\zeta_j(q_j)r_{ji}}}{r_{ji}}\left[\zeta_j(q_j)^2-3\zeta_i(q_i)^2\right]-d\Delta_{ji}\left[\frac{1}{r_{ji}}+\zeta_j(q_j)\right] \\
&= \frac{d\Delta_{ji}-4\zeta_i(q_i)\zeta_j(q_j)\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)r_{ji}^{-1}e^{-\zeta_j(q_j)r_{ji}}}{r_{ji}}-d\Delta_{ji}\left[\frac{1}{r_{ji}}+\zeta_j(q_j)\right] \\
&= \frac{-4\zeta_i(q_i)\zeta_j(q_j)\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)r_{ji}^{-1}e^{-\zeta_j(q_j)r_{ji}}}{r_{ji}}-d\Delta_{ji}\zeta_j(q_j) \\
&= -\frac{\zeta_j}{r_{ji}}\left[\frac{4\zeta_i(q_i)\zeta_j(q_j)\zeta_{q,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)e^{-\zeta_j(q_j)r_{ji}}}{r_{ji}}+r_{ji}d\Delta_{ji}\right] \qquad \text{(C.6)}
\end{aligned}$$

where $d\Delta_{ji}$ would represent $\left(\frac{\partial \Delta_{ji}}{\partial \rho_{nm}}\right)$. To find the derivation of this term please refer to Appendix I, subsection C.

Differentiation of $\alpha_{1,i}(q_i)$ with respect to ρ_{nm}

Recall[172]

$$\alpha_{1,i}(q_i) = \alpha_{1,i}(0)e^{-3B_iq_i}$$

Now, differentiating $\alpha_{1,i}(q_i)$ with respect to ρ_{nm}:

$$\begin{aligned}
d\alpha_{1,i} &= \frac{d\alpha_{1,i}(q_i)}{dq_i} \\
&= -3\alpha_{1,i}(0)B_i(dq_i)e^{-3B_iq_i} \\
&= -3B_i(dq_i)\alpha_{1,i}(q_i) \qquad \text{(C.7)}
\end{aligned}$$

Differentiation of $\eta_{1,i}(q_i)$ with respect to ρ_{nm}

Recall[172]

$$\eta_{1,i}(q_i) = \sqrt{\frac{N_{eff,i}(q_i)}{\alpha_{1,i}(q_i)}} \qquad \text{(C.8)}$$

where the $N_{eff,i}(q_i)$ term is a scaled function with the form

$$N_{eff,i}(q_i) = \begin{cases} N_{val,i}(q_i)N_{eff,i}(0)/N_{val,i}(0) & \text{if } N_{val,i}(q_i) > 0 \\ 0 & \text{if } N_{val,i}(q_i) \leq 0 \end{cases} \tag{C.9}$$

where $N_{val,i}(q_i)$ is defined as

$$N_{val,i}(q_i) = N_{val,i}(0) - q_i \tag{C.10}$$

where $N_{eff,i}(0)$ and $N_{val,i}(0)$ are constants.

Now, differentiating $\eta_{1,i}(q_i)$ with respect to ρ_{nm}, we define:

$$\begin{aligned} \frac{\partial \eta_{1,i}}{\partial \rho_{nm}} &= \frac{d\eta_{1,i}(q_i)}{dq_i} \\ &= \frac{dN_{eff,i}(q_i)}{2\sqrt{N_{eff}(q_i)\alpha_{1,i}(q_i)}} - \frac{d\alpha_{1,i}(q_i)\sqrt{N_{eff,i}(q_i)}}{2\sqrt{\alpha_{1,i}(q_i)^3}} \end{aligned} \tag{C.11}$$

where $dN_{eff,i}$ is the partial of $N_{eff,i}$ with respect to ρ_{nm}. Continuing with the derivation using Eqn. (C.9):

$$\frac{\partial N_{eff,i}(q_i)}{\partial \rho_{nm}} = \frac{N_{eff,i}(0)}{N_{val,i}(0)} dN_{val,i}(q_i) \tag{C.12}$$

where a similar notition is used for the partial of $N_{val,i}$ with respect to ρ_{nm}. Using Eqn. (C.10),

$$\frac{\partial N_{val,i}(q_i)}{\partial \rho_{nm}} = -\left(\frac{\partial q_i}{\partial \rho_{nm}}\right) \tag{C.13}$$

Combining Eqn. (C.12) and Eqn. (C.13)

$$\frac{\partial N_{eff,i}(q_i)}{\partial \rho_{nm}} = -\frac{N_{eff,i}(0)}{N_{val,i}(0)}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right) \tag{C.14}$$

Plugging Eqn. (C.7) and Eqn. (C.14) into Eqn. (C.11)

$$\begin{aligned}
\frac{\partial \eta_{1,i}(q_i)}{\partial \rho_{nm}} &= \frac{-\frac{N_{eff,i}(0)}{N_{val,i}(0)}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)}{2\sqrt{N_{eff,i}(q_i)\alpha_{1,i}(q_i)}} - \frac{-3\alpha_{1,i}(0)B_i\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)e^{-3B_i q_i}\sqrt{N_{eff,i}(q_i)}}{2\sqrt{\alpha_{1,i}(q_i)^3}} \\
&= -\frac{N_{eff,i}(0)\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)}{2N_{val,i}(0)\sqrt{N_{eff,i}(q_i)\alpha_{1,i}(q_i)}} - \frac{-3B_i(dq_i)\alpha_{1,i}(q_i)\sqrt{N_{eff,i}(q_i)}}{2\sqrt{\alpha_{1,i}(q_i)^3}} \\
&= -\frac{\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)}{2}\sqrt{\frac{N_{eff,i}(q_i)}{\alpha_{1,i}(q_i)}}\left(\frac{N_{eff,i}(0)}{N_{val,i}(0)N_{eff,i}(q_i)} - 3B_i\right) \\
&= \frac{\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)}{2}\sqrt{\frac{N_{eff,i}(q_i)}{\alpha_{1,i}(q_i)}}\left(3B_i - \frac{1}{N_{val,i}(q_i)}\right) \\
&= \frac{\eta_{1,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)}{2}\left(3B_i - \frac{1}{N_{val,i}(q_i)}\right) \qquad \text{(C.15)}
\end{aligned}$$

Differentiation of $\eta_{2,i}(q_i)$ with respect to ρ_{nm}

Recall[172]

$$\eta_{2,i} = \left(\frac{\sqrt{9N_{eff,i}(q_i)\alpha_{1,i}(q_i)}}{\alpha_{2,i}(q_i)}\right)^{1/2}$$

Now we must consider $\alpha_{2,i}(q_i)$. There will be two derivations; the first with $\alpha_{2,i}(q_i)$ demending on ρ_{nm} and the next with $\alpha_{2,i}(q_i)$ as a constant. During all of the following derivations "dX" will represent $\left(\frac{\partial X}{\partial \rho_{nm}}\right)$.

With $\alpha_{2,i}(q_i)$ demending on ρ_{nm}

$$\begin{aligned}
\frac{\partial \eta_{2,i}}{\partial \rho_{nm}} &= \left(\frac{1}{2}\right)\left(\frac{\sqrt{9N_{eff,i}(q_i)\alpha_{1,i}(q_i)}}{\alpha_{2,i}(q_i)}\right)^{-1/2} \\
&\quad * \left[\frac{d\left(\sqrt{9N_{eff,i}(q_i)\alpha_{1,i}(q_i)}\right)\alpha_{2,i}(q_i)}{\alpha_{2,i}(q_i)^2}\right. \\
&\quad \left. + \frac{(d\alpha_{2,i}(q_i))\sqrt{9N_{eff,i}(q_i)\alpha_{1,i}(q_i)}}{\alpha_{2,i}(q_i)^2}\right] \qquad \text{(C.16)}
\end{aligned}$$

Taking a closer look

$$\begin{aligned}\frac{\partial\left(\sqrt{9N_{eff,i}(q_i)\alpha_{1,i}(q_i)}\right)}{\partial\rho_{nm}} &= \left(\frac{9}{2}\right)(9N_{eff,i}(q_i)\alpha_{1,i}(q_i))^{-1/2} \\ &\quad * ((dN_{eff,i}(q_i))\alpha_{1,i}(q_i) + (d\alpha_{1,i}(q_i))N_{eff,i}(q_i))\end{aligned} \tag{C.17}$$

Now substituing in Eqn. (C.17) into Eqn. (C.16)

$$\begin{aligned}\frac{\partial\eta_{2,i}}{\partial\rho_{nm}} &= \left(\frac{9}{4}\right)\left(\frac{\alpha_{2,i}(q_i)^{1/2}}{(9N_{eff,i}(q_i)\alpha_{1,i}(q_i))^{1/4}}\right) \\ &\quad * \left(\alpha_{2,i}(q_i)^3\frac{dN_{eff,i}(q_i)\alpha_{1,i}(q_i) + d\alpha_{1,i}(q_i)N_{eff,i}(q_i)}{(9N_{eff,i}(q_i)\alpha_{1,i}(q_i))^{1/2}}\right. \\ &\quad \left.-d\alpha_{2,i}(q_i)\alpha_{2,i}(q_i)^2(9N_{eff,i}(q_i)\alpha_{1,i}(q_i))^{1/2}\right) \\ &= \left(\frac{9}{4}\right)\left(\frac{\alpha_{2,i}(q_i)^{1/2}}{(9N_{eff,i}(q_i)\alpha_{1,i}(q_i))^{1/4}}\right)\left(\alpha_{2,i}(q_i)^3(9N_{eff,i}(q_i)\alpha_{1,i}(q_i))^{1/2}\right) \\ &\quad * \left(\frac{dN_{eff,i}(q_i)\alpha_{1,i}(q_i) + d\alpha_{1,i}(q_i)N_{eff,i}(q_i)}{9N_{eff,i}(q_i)\alpha_{1,i}(q_i)} - \frac{d\alpha_{2,i}(q_i)}{\alpha_{2,i}(q_i)}\right) \\ &= \left(\frac{\alpha_{2,i}(q_i)^4}{4}\right)(\eta_{2,i}(q_i))\left(\frac{dN_{eff,i}(q_i)}{N_{eff,i}(q_i)} + \frac{d\alpha_{1,i}(q_i)}{\alpha_{1,i}(q_i)} - 9\frac{d\alpha_{2,i}(q_i)}{\alpha_{2,i}(q_i)}\right)\end{aligned} \tag{C.18}$$

Then using Eqn. (C.7) and Eqn. (C.12)

$$\frac{\partial\eta_{2,i}}{\partial\rho_{nm}} = -\left(\frac{\alpha_{2,i}(q_i)^4\eta_{2,i}(q_i)\left(\frac{\partial q_i}{\partial\rho_{nm}}\right)}{4}\right)\left(\frac{1}{N_{val,i}(q_i)} + 3B_i + 9\frac{\kappa_i}{\alpha_{2,i}(q_i)}\right) \tag{C.19}$$

If it is assumed that $\alpha_{2,i}$ is a *constant* then, using the same notation as before,

$$\begin{aligned}
\frac{\partial \eta_{2,i}}{\partial q_i} &= \left(\frac{3}{\alpha_{2,i}}\right)^{\frac{1}{2}} \frac{1}{2} \left(\sqrt{N_{eff,i}(q_i)\alpha_{1,i}(q_i)}\right)^{\frac{-1}{2}} \\
&\quad * \left[\frac{1}{2} N_{eff,i}(q_i)^{\frac{-1}{2}} \alpha_{1,i}(q_i)^{\frac{1}{2}} (dN_{eff,i}(q_i)) + \frac{1}{2} N_{eff,i}(q_i)^{\frac{1}{2}} \alpha_{1,i}(q_i)^{\frac{-1}{2}} (d\alpha_{1,i}(q_i))\right] \\
&= \left(\frac{3}{16\alpha_{2,i}\sqrt{N_{eff,i}(q_i)\alpha_{1,i}(q_i)}}\right)^{1/2} \left[\left(\frac{\alpha_{1,i}(q_i)(dN_{eff,i}(q_i))^2}{N_{eff,i}(q_i)}\right)^{\frac{1}{2}} + \left(\frac{N_{eff,i}(q_i)(d\alpha_{1,i}(q_i))^2}{\alpha_{1,i}(q_i)}\right)^{\frac{1}{2}}\right] \\
&= \left(\frac{3\sqrt{N_{eff,i}\alpha_{1,i}}}{16\alpha_{2,i}}\right)^{1/2} \left[\frac{(dN_{eff,i}(q_i))}{N_{eff,i}(q_i)} + \frac{(d\alpha_{1,i}(q_i))}{\alpha_{1,i}(q_i)}\right] \\
&= -\frac{(dq_i)}{4} \left(\frac{\sqrt{9N_{eff,i}(q_i)\alpha_{1,i}(q_i)}}{\alpha_{2,i}}\right)^{\frac{1}{2}} \left[\frac{1}{N_{val,i}(q_i)} + 3B_i\right] \\
&= -\frac{\eta_{2,i}\left(\frac{\partial q_i}{\partial \rho_{nm}}\right)}{4} \left[\frac{1}{N_{val,i}(q_i)} + 3B_i\right] \qquad \text{(C.20)}
\end{aligned}$$

Appendix D

Charge Table for Chloride/Methylchloride Reaction

	Tansition State		CH_3Cl	
	Avg. Charge	Std. Dev.	Avg. Charge	Std. Dev.
M06-2X gas	-0.6118	–	-0.1866	–
SRP gas	-0.7700	–	-0.1659	–
LJ_{Cl^-}	-0.8153	0.0451	-0.2089	0.0469
LJ_{CH_3Cl}	-0.8940	0.0447	-0.2465	0.0519
QXD	-0.7905	0.0694	-0.2104	0.0479

Table D.1: **Charges of the chlorine atoms in the methylchloride/chloride reaction pathway.** The value of the Mulliken charges of the chlorine atoms in the transition state and the methylchloride molecule are shown for gas phase M06-2X and SRP AM1 Hamiltonian methods, as well as in the solution phase for the SRP Hamiltonian with the LJ_{Cl^-} and LJ_{CH_3Cl} Lennard-Jones parameterizations and the QXD method.

Appendix E

Detailed Definition of the SCOSMO Switching Layer

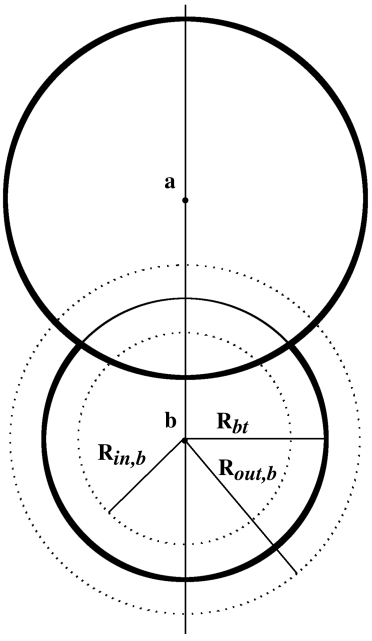


Figure E.1: **The intersection of two SCOSMO cavitation spheres.** Surface response elements within the switching layer are smoothly turned off as two cavities intersect.

As stated in Chapter 4 a switching function, p_t, smoothly transitions between 0 and 1, preventing surface elements from interacting with other surface elements which have been partially or fully removed by the intersection of two or more particles. The intersection of the cavitation spheres is shown in Figure E.1. The switching function takes the form,

$$p_{t\in a} = \prod_{b\neq a} S_{\text{on}}(R_{bt}^2, R_{\text{in},b}^2, R_{\text{out},b}^2). \tag{E.1}$$

where S_{on}, takes the form

$$S_{\text{on}}(x, x_{\text{lo}}, x_{\text{hi}}) = \begin{cases} 0 & \text{if } x < x_{\text{lo}} \\ 1 & \text{if } x > x_{\text{hi}} \\ 1 - 10u^3 - 15u^4 + 6u^5 & \text{otherwise} \\ u \equiv (x_{\text{hi}} - x)/(x_{\text{hi}} - x_{\text{lo}}) & \end{cases} \tag{E.2}$$

The bounds of the switching function are defined as

$$R_{\text{in},a} = (1 - oh)R_{\text{rad},a} \tag{E.3}$$

and

$$R_{\text{out},a} = (1 + (1 - o)h)R_{\text{rad},a} \tag{E.4}$$

where

$$o = \frac{1}{2} + \frac{1}{h} - \sqrt{\frac{1}{h^2} - \frac{1}{28}}. \tag{E.5}$$

For the present work the value of h is set to 0.15. In essence, h and o are scaling factors set by the user to modulate the 'thickness' of the switching layer around $R_{\text{rad},a}$

Appendix F

Comprehensive Table of VRSCOSMO Training Set Solvation Free Energies

Training Set					
Compound [Charge]	Ref Val	VRS	Diff	SCMO	Diff
1-Butene[0]	1.38	0.74	-0.63	0.80	-0.57
1-Butyne[0]	-0.16	-0.58	-0.42	-0.03	0.12
1-Hexene[0]	1.68	1.30	-0.37	1.26	-0.41
1-Hexyne[0]	0.29	0.04	-0.25	0.44	0.15
1-Pentyne[0]	0.01	-0.29	-0.30	0.18	0.17
2,2,4-Trimethylpentane[0]	2.85	2.65	-0.19	2.12	-0.72
2,4-Dimethylpentane[0]	2.88	2.43	-0.44	1.92	-0.95
2-Methylpentane[0]	2.52	2.20	-0.31	1.76	-0.75
2-Methylpropane[0]	2.32	1.59	-0.72	1.33	-0.98
Butenyne[0]	0.04	-0.75	-0.79	-0.11	-0.15
Cyclopentane[0]	1.20	1.64	0.44	1.34	0.14
Cyclopropane[0]	0.75	0.39	-0.35	0.62	-0.12
Ethane[0]	1.83	1.05	-0.77	0.88	-0.94
Ethyne[0]	-0.01	-1.42	-1.41	-0.58	-0.57

Training Set					
Compound [Charge]	Ref Val	VRS	Diff	SCMO	Diff
Methylcyclohexane[0]	1.71	2.32	0.61	1.75	0.04
Propene[0]	1.27	0.36	-0.90	0.55	-0.71
Toluene[0]	-0.89	-0.15	0.73	0.30	1.19
cis-1,2-Dimethylcyclohexane[0]	1.58	2.46	0.88	1.89	0.31
e-2-Pentene[0]	1.34	1.02	-0.31	1.06	-0.27
m-Xylene[0]	-0.84	0.10	0.94	0.49	1.33
n-Butane[0]	2.08	1.64	-0.43	1.33	-0.74
n-Hexane[0]	2.49	2.23	-0.25	1.80	-0.68
n-Octane[0]	2.89	2.78	-0.11	2.16	-0.72
n-Propane[0]	1.96	1.35	-0.60	1.12	-0.83
o-Xylene[0]	-0.90	0.03	0.93	0.41	1.31
s-trans-1,3-Butadiene[0]	0.61	0.03	-0.57	0.39	-0.21
1,2-Ethanediol[0]	-9.30	-9.21	0.08	-8.46	0.83
1,4-Dioxane[0]	-5.05	-6.17	-1.12	-4.77	0.28
1-Butanol[0]	-4.72	-4.40	0.31	-4.48	0.23
1-Heptanol[0]	-4.24	-3.54	0.69	-3.63	0.60
1-Hexanol[0]	-4.36	-3.68	0.67	-3.95	0.40
1-Octanol[0]	-4.09	-3.10	0.98	-3.54	0.54
1-Pentanol[0]	-4.47	-3.97	0.49	-4.21	0.25
2-Heptanone[0]	-3.04	-2.96	0.07	-2.29	0.74
2-Hexanone[0]	-3.29	-3.35	-0.06	-2.59	0.70
2-MethoxyEthanol[0]	-6.77	-7.24	-0.47	-6.80	-0.03
2-Octanone[0]	-2.88	-2.80	0.07	-2.27	0.60
2-Pentanone[0]	-3.53	-3.61	-0.08	-3.04	0.48
2-Propanol[0]	-4.76	-4.74	0.01	-4.92	-0.16
2-Propen-1-ol[0]	-5.08	-5.66	-0.58	-5.13	-0.05
4-Heptanone[0]	-2.93	-2.81	0.11	-2.30	0.62
ButylEthanoate[0]	-2.55	-4.21	-1.66	-3.38	-0.83
Cyclopentanol[0]	-5.49	-4.15	1.33	-4.48	1.00
DibutylKetone[0]	-2.67	-2.19	0.47	-1.86	0.80

Training Set					
Compound [Charge]	Ref Val	VRS	Diff	SCMO	Diff
DiethylEther[0]	-1.76	-2.16	-0.40	-1.71	0.04
DimethylKetone[0]	-3.85	-4.44	-0.59	-3.46	0.38
Ethanal[0]	-3.50	-3.79	-0.29	-2.62	0.87
EthylEthanoate[0]	-3.10	-4.98	-1.88	-3.83	-0.73
EthylMethanoate[0]	-2.65	-4.82	-2.17	-3.27	-0.62
EthylPhenylEther[0]	-2.22	-1.72	0.49	-1.00	1.21
HydrogenPeroxide[0]	-8.58	-8.32	0.25	-8.00	0.57
IsopropylEther[0]	-0.53	-1.66	-1.13	-1.41	-0.88
MethoxyBenzene[0]	-2.45	-2.15	0.29	-1.27	1.17
MethylBenzoate[0]	-3.91	-8.58	-4.67	-6.43	-2.52
MethylButanoate[0]	-2.83	-4.53	-1.70	-3.39	-0.56
MethylEthanoate[0]	-3.32	-5.41	-2.09	-4.13	-0.81
MethylHexanoate[0]	-2.49	-4.01	-1.52	-3.02	-0.53
MethylIsopropylEther[0]	-2.01	-2.15	-0.14	-1.93	0.07
MethylMethanoate[0]	-2.78	-5.34	-2.56	-3.45	-0.67
MethylOctanoate[0]	-2.04	-3.27	-1.23	-2.58	-0.54
MethylPentanoate[0]	-2.57	-4.25	-1.68	-3.15	-0.58
MethylPeroxide[0]	-5.28	-3.93	1.34	-2.71	2.56
MethylPhenylKetone[0]	-4.58	-4.02	0.55	-2.88	1.69
MethylPropanoate[0]	-2.93	-4.95	-2.02	-3.65	-0.72
MethylPropylEther[0]	-1.66	-2.01	-0.35	-1.67	-0.01
Octanal[0]	-2.29	-1.82	0.46	-1.19	1.09
PentanoicAcid[0]	-6.16	-6.56	-0.40	-6.16	-0.00
PentylEthanoate[0]	-2.45	-3.94	-1.49	-3.11	-0.66
Phenylmethanal[0]	-4.02	-3.55	0.46	-2.24	1.77
Propanal[0]	-3.44	-3.14	0.29	-2.20	1.23
PropanoicAcid[0]	-6.47	-7.36	-0.89	-6.67	-0.20
PropylEthanoate[0]	-2.86	-4.46	-1.60	-3.59	-0.73
Tetrahydrofuran[0]	-3.47	-3.03	0.43	-2.28	1.18
Tetrahydropyran[0]	-3.12	-2.22	0.89	-1.77	1.34

Training Set					
Compound [Charge]	Ref Val	VRS	Diff	SCMO	Diff
m-Cresol[0]	-5.49	-4.56	0.92	-4.20	1.28
m-Hydroxybenzaldehyde[0]	-9.51	-8.04	1.46	-6.82	2.68
o-Cresol[0]	-5.87	-4.68	1.18	-3.84	2.02
p-Hydroxybenzaldehyde[0]	-10.48	-8.37	2.10	-7.22	3.25
t-ButylMethylEther[0]	-2.21	-1.97	0.23	-1.78	0.42
t-ButylMethylKetone[0]	-2.89	-2.83	0.05	-2.33	0.55
trans-1-Propanol[0]	-4.83	-4.60	0.22	-4.62	0.20
1,1-Dimethyl-3-phenylurea[0]	-9.63	-11.26	-1.63	-9.36	0.26
1,1-Dimethylhydrazine[0]	-4.48	-5.64	-1.16	-8.17	-3.69
1,2-Ethanediamine[0]	-9.72	-7.65	2.06	-9.51	0.20
1,4-DimethylPiperazine[0]	-7.58	-1.77	5.80	-0.85	6.72
1-Methyl-2-Nitrobenzene[0]	-3.59	-7.29	-3.70	-4.59	-1.00
1-MethylPiperazine[0]	-7.77	-3.37	4.39	-2.83	4.93
1-Methylthymine[0]	-10.40	-14.13	-3.73	-12.25	-1.85
1-Nitrobutane[0]	-3.08	-7.85	-4.77	-3.53	-0.45
1-Nitropropane[0]	-3.34	-8.13	-4.79	-3.63	-0.29
2,4-Dimethylpyridine[0]	-4.86	-2.15	2.70	-2.87	1.98
2,5-Dimethylpyridine[0]	-4.72	-1.95	2.76	-2.66	2.05
2,6-Dimethylpyridine[0]	-4.60	-2.13	2.46	-2.19	2.40
2-Ethylpyrazine[0]	-5.51	-3.53	1.97	-4.18	1.33
2-Methoxyethanamine[0]	-6.55	-5.96	0.58	-6.81	-0.26
2-Methylaniline[0]	-5.56	-4.61	0.94	-4.81	0.74
2-Methylpyrazine[0]	-5.57	-3.84	1.72	-3.84	1.72
2-Methylpyridine[0]	-4.63	-2.28	2.34	-2.31	2.31
2-Nitropropane[0]	-3.14	-7.43	-4.29	-3.25	-0.11
3,4-Dimethylpyridine[0]	-5.22	-2.15	3.06	-2.83	2.38
3,5-Dimethylpyridine[0]	-4.84	-1.86	2.97	-2.59	2.24
3-Aminoaniline[0]	-9.92	-10.06	-0.14	-9.52	0.39
3-Methylaniline[0]	-5.67	-4.76	0.90	-4.83	0.83
3-Methylpyridine[0]	-4.77	-2.19	2.57	-1.80	2.96

Training Set					
Compound [Charge]	Ref Val	VRS	Diff	SCMO	Diff
4-Ethylpyridine[0]	-4.74	-2.34	2.39	-2.24	2.49
4-Methylaniline[0]	-5.55	-4.67	0.87	-4.61	0.93
4-Methylpyridine[0]	-4.94	-2.40	2.53	-1.95	2.98
9-methyladenine[0]	-13.60	-12.38	1.21	-13.08	0.51
Acetamide[0]	-9.71	-11.48	-1.77	-10.75	-1.04
Benzamide[0]	-10.90	-10.91	-0.01	-10.25	0.64
Benzonitrile[0]	-4.10	-4.24	-0.14	-5.42	-1.32
Butanonitrile[0]	-3.64	-4.50	-0.86	-6.00	-2.36
Butylamine[0]	-4.29	-2.96	1.32	-4.43	-0.14
Butylenimine[0]	-5.48	-2.10	3.37	-2.42	3.05
Diethyleneimine[0]	-7.40	-4.82	2.57	-4.76	2.63
Dimethylamine[0]	-4.29	-2.19	2.09	-2.47	1.81
Dimethylaniline[0]	-3.58	-2.24	1.33	-0.93	2.64
Dipropylamine[0]	-3.66	-0.81	2.84	-1.42	2.23
Ethanonitrile[0]	-3.89	-5.51	-1.62	-6.68	-2.79
Hydrazine[0]	-6.26	-8.39	-2.13	-14.05	-7.79
Methylhydrazine[0]	-5.31	-6.71	-1.40	-9.81	-4.50
Morpholine[0]	-7.17	-6.18	0.98	-6.29	0.87
1-Methyl-2-Nitrobenzene[0]	-3.59	-7.29	-3.70	-4.59	-1.00
Nitrobenzene[0]	-4.12	-7.62	-3.50	-4.34	-0.22
Nitroethane[0]	-3.71	-8.67	-4.96	-4.03	-0.32
Nitromethane[0]	-3.95	-9.33	-5.38	-4.35	-0.40
Pentylamine[0]	-4.10	-2.67	1.42	-4.25	-0.15
Propanonitrile[0]	-3.85	-4.95	-1.10	-6.35	-2.50
Propylamine[0]	-4.36	-3.18	1.17	-4.61	-0.25
Pyridine[0]	-4.70	-2.46	2.23	-2.53	2.16
Trimethylamine[0]	-3.23	-0.68	2.54	-0.46	2.77
Trimethylenimine[0]	-5.56	-2.69	2.86	-3.17	2.38
Urea[0]	-13.80	-17.77	-3.97	-16.93	-3.13
e-n-Methylacetamide[0]	-10.00	-10.25	-0.25	-9.06	0.93

Training Set					
Compound [Charge]	Ref Val	VRS	Diff	SCMO	Diff
n-Ethylaniline[0]	-4.62	-3.07	1.54	-2.07	2.54
n-Methylaniline[0]	-5.55	-3.61	1.93	-2.42	3.12
n-Methylmorpholine[0]	-6.34	-3.76	2.57	-3.23	3.10
z-n-Methylacetamide[0]	-10.0	-8.73	1.26	-8.79	1.20
2-Butanol[-1]	-86.0	-90.01	-4.01	-81.13	4.86
3-Hydroxyphenol[-1]	-76.0	-71.75	4.24	-66.98	9.01
EthylHydroperoxide[-1]	-91.0	-91.66	-0.66	-95.69	-4.69
Methoxyethanol[-1]	-91.0	-91.53	-0.53	-85.45	5.54
MethylHydroperoxide[-1]	-95.0	-92.82	2.17	-96.47	-1.47
PropanoicAcid[-1]	-78.0	-78.49	-0.49	-74.68	3.31
Propanol[-1]	-90.0	-91.73	-1.73	-82.99	7.00
m-Cresol[-1]	-73.0	-68.29	4.70	-63.41	9.58
o-Cresol[-1]	-72.0	-67.06	4.93	-61.71	10.28
p-Cresol[-1]	-74.0	-68.39	5.60	-63.24	10.75
2-Nitrophenol[-1]	-62.0	-64.77	-2.77	-63.41	-1.41
3-Nitrophenol[-1]	-60.0	-59.80	0.19	-56.20	3.79
4-Nitroaniline[-1]	-59.0	-61.88	-2.88	-64.01	-5.01
4-Nitrophenol[-1]	-60.0	-57.55	2.44	-56.55	3.44
Acetamide[-1]	-82.0	-82.39	-0.39	-83.69	-1.69
Diphenolamine[-1]	-56.0	-53.90	2.09	-47.48	8.51
DimethylEther[1]	-78.0	-68.66	9.33	-64.45	13.54
Ethanol[1]	-86.0	-74.55	11.44	-75.91	10.08
2-Methylaniline[1]	-68.0	-66.77	1.22	-70.22	-2.22
3-Methylaniline[1]	-68.0	-66.76	1.23	-70.73	-2.73
AllylAmine[1]	-70.0	-72.64	-2.64	-75.36	-5.36
Cyclohexamine[1]	-67.0	-67.03	-0.03	-70.21	-3.21
Diethylamine[1]	-62.0	-61.11	0.88	-59.91	2.08
Dimethylamine[1]	-67.0	-67.45	-0.45	-66.39	0.60
Isopropylamine[1]	-68.0	-70.03	-2.03	-72.93	-4.93
Methylamine[1]	-74.0	-76.66	-2.66	-79.98	-5.98

Training Set					
Compound [Charge]	Ref Val	VRS	Diff	SCMO	Diff
Triethylamine[1]	-53.0	-52.45	0.54	-49.15	3.84
n-Ethylaniline[1]	-60.0	-56.04	3.95	-56.69	3.30
n-Methylaniline[1]	-61.0	-60.54	0.45	-59.65	1.34
t-Butylamine[1]	-65.0	-66.89	-1.89	-69.70	-4.70
MSE	–	–	0.2	–	0.8
MUE	–	–	1.6	–	2.0
RMSE	–	–	2.4	–	3.0

Table F.1: All values reported in kcal/mol. VRS=VRSCOSMO, SCMO=SCOSMO.

Appendix G

Comprehensive Table of VRSCOSMO Test Set Solvation Free Energies

Test Set				
Compound [Charge]	Ref Val	VRS	SCMO	DFTBPR[269]
1-Pentene[0]	1.66	-0.69	-0.65	-1.9
2,2-Dimethylpropane[0]	2.50	-0.58	-0.95	-2.2
2-Methylpropene[0]	1.16	-0.74	-0.52	-1.6
Anthracene[0]	-4.23	3.30	4.12	2.6
Benzene[0]	-0.87	0.47	1.04	-0.1
Cyclohexane[0]	1.23	0.65	0.32	-0.9
Cyclopentene[0]	0.56	0.17	0.28	-1.0
Ethene[0]	1.27	-1.08	-0.87	-1.5
Ethylbenzene[0]	-0.80	0.99	1.38	-0.1
Methane[0]	2.00	-1.20	-1.31	-1.8
Naphthalene[0]	-2.39	1.70	2.43	1.1
Propyne[0]	-0.31	-0.75	-0.10	-1.7
n-Heptane[0]	2.62	-0.26	-0.62	-2.2
p-Xylene[0]	-0.81	0.94	1.32	-0.1

Test Set				
Compound [Charge]	Ref Val	VRS	SCMO	DFTBPR[269]
1,2-Dimethoxyethane[0]	-4.81	1.05	1.90	0.5
2-Butanone[0]	-3.64	-0.31	0.58	-1.5
2-Methyl-2-propanol[0]	-4.51	0.01	-0.13	0.8
3-Pentanone[0]	-3.41	-0.13	0.66	-1.1
AceticAcid[0]	-6.70	-0.94	-0.17	-1.4
Butanal[0]	-3.18	0.25	1.06	-1.0
ButanoicAcid[0]	-6.36	-0.53	0.24	-1.3
Cyclopentanone[0]	-4.68	0.98	1.71	0.0
DimethylEther[0]	-1.92	-0.91	-0.32	-0.8
Ethanol[0]	-5.01	-0.08	0.05	1.0
HexanoicAcid[0]	-6.21	0.08	0.41	-1.3
Methanol[0]	-5.11	-0.31	-0.15	1.2
Pentanal[0]	-3.03	0.28	1.18	-1.2
Phenol[0]	-6.62	1.77	2.28	2.7
p-Cresol[0]	-6.14	1.59	1.89	2.3
Aniline[0]	-5.49	0.36	0.81	2.1
Diethylamine[0]	-4.07	2.55	2.01	3.7
Ethylamine[0]	-4.50	0.94	-0.42	3.8
Methylamine[0]	-4.56	0.72	-0.67	3.8
Piperidine[0]	-5.11	3.55	2.88	4.8
2-Propanol[-1]	-88.00	-3.21	4.90	0
3-Pentanone[-1]	-76.00	5.27	10.20	2
4-Hydroxyphenol[-1]	-80.00	6.60	11.46	8
Acetaldehyde[-1]	-78.00	4.50	9.09	0
AceticAcid[-1]	-80.00	0.54	3.92	-2
Acetone[-1]	-78.00	4.66	10.90	0
AcrylicAcid[-1]	-76.00	-0.99	2.45	-2
AllylAlcohol[-1]	-88.00	-0.59	5.91	2
BenzylAlcohol[-1]	-87.00	-0.98	7.27	6
Ethanol[-1]	-93.00	-0.30	8.46	3

Test Set				
Compound [Charge]	Ref Val	VRS	SCMO	DFTBPR[269]
FormicAcid[-1]	-78.00	0.16	2.20	-3
HexanoicAcid[-1]	-76.00	-0.53	3.64	-4
Methanol[-1]	-97.00	1.71	10.36	5
Phenol[-1]	-74.00	5.52	10.30	5
PyruvicAcid[-1]	-70.00	-3.98	-0.38	-7
t-Butanol[-1]	-84.00	-5.14	2.54	-2
Aniline[-1]	-65.00	-9.50	-7.48	1
Methanol[1]	-91.00	12.12	10.08	9
Acetone[1]	-75.00	10.72	13.25	7
Acetophenone[1]	-63.00	10.26	12.75	6
DiethylEther[1]	-70.00	10.34	12.96	7
3-Aminoaniline[1]	-64.00	-5.26	-9.46	-2
4-Methyl-n,n-Dimethylaniline[1]	-54.00	2.36	5.90	0
4-Methylaniline[1]	-68.00	1.33	-2.36	1
Aniline[1]	-70.00	1.22	-2.51	1
Di-n-propylamine[1]	-59.00	0.23	0.97	-1
Diallylamine[1]	-60.00	-0.36	1.73	-2
Tri-n-propylamine[1]	-49.00	-0.32	2.56	-3
Trimethylamine[1]	-59.00	-0.87	2.67	-4
n,n-Diethylaniline[1]	-52.00	3.68	6.79	-1
n-Propylamine[1]	-70.00	-1.66	-5.02	-3
MSE	–	1.1	2.4	0.5
MUE	–	2.3	3.6	2.4

Table G.1: Table of model solvation energy errors. All values reported in kcal/mol. VRS=VRSCOSMO, SCMO=SCOSMO, DFTBPR=SCC-DFTBPR by Hou *et. al.*