A Model For the Calculation of Solvent Effects on Reaction Rates for Process Design Purposes

by

João Paulo Dantas Aumond

M.S. in Chemical Engineering Practice, MIT (1995)

Engenheiro Químico, Universidade de São Paulo (1992)

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

DOCTOR OF PHILOSOPHY
in Chemical Engineering
at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 1999

© Massachusetts Institute of Technology. All rights reserved.

Signature of Author

Department of Chemical Engineering
April 5th, 1999

Certified by

George Stephanopoulos
Arthur D. Little Professor of Chemical Engineering
Thesis Supervisor

Accepted by

Robert E. Cohen
St. Laurent Professor of Chemical Engineering
Chairman, Committee for Graduate Students
A Model For the Calculation of Solvent Effects on Reaction Rates for Process Design Purposes

by

João Paulo Dantas Aumond

Submitted to the Department of Chemical Engineering on March 3rd 1999
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Chemical Engineering

Abstract

Solvents are used in chemical processes as separation and cleaning agents as well as reaction media. Their selection has a significant impact on the environmental impact, efficiency and profitability of a process. For this reason it would be desirable that solvents be chosen with respect not only to their effectiveness in their respective process tasks but also for process-wide requirements such as their ease of recovery, low toxicity and environmental impact and possible applicability to other process tasks. Although there are models for the evaluation of solvents as separation agents, there is a need for models for the assessment of solvents as reaction media that are fast, and therefore able to screen large numbers of molecules in a reasonable time, as well as applicable to a wide range of solvents and reactions and able to provide results that are accurate for process design needs.

Solvent effects on reaction rates can be very marked. Reaction rates can vary by several orders of magnitude by simply substituting one solvent for another. This effect is due to differential solvation of reactants and transition state and can be summarized by the rule: if the activated complex has a charge distribution greater than the reactants, then that reaction's rate will be magnified by more polar solvents and vice-versa. Therefore, two elements are necessary to assess solvent effects on reaction rates: the charge distribution of the solvent and the evolution of the charge distribution from reactants to transition state.

To calculate the atomic charges of solvents, a novel scheme based on the concept of conjugation operators has been devised. Conjugation operators create charge separation along a bond or a set of contiguous bonds and can be defined in quantum mechanical terms as operators that act on the electronic population matrix. Operators are bond-specific and have weights associated to them, indicating how much a certain bond or chain of bonds contributes to the concentration (or dispersion) of electronic density around an atom. These weights are derived from standard quantum chemical calculations.
Quantifying the charge distribution evolution during the activation step of a chemical reaction is a difficult task because usually little or no information is known about the structure and charge distribution of the transition state. Since the task at hand is to determine the difference in charge distribution between the two species, this problem is circumvented by defining a symbolic molecular entity, named the reaction fingerprint. The reaction fingerprint represents the difference of charge distribution between reactants and transition state, and its Gibbs' free energy of solvation is equal to the difference between those of the activated complex and the reactants. It consists of adjustable parameters that describe its polarity as well as the geometric distribution of its charges.

To calculate solvent effects on a specific reaction, the reaction fingerprint parameters are regressed from experimental kinetic data in at least three solvents. With these parameters, the Gibbs' free energy of solvation of the reaction fingerprint in the solvent is calculated using a simplified statistical mechanical expression that uses the first term of a density expansion of the radial distribution function. The energy of interaction between solvent and reaction fingerprint is a function of both their dipole moment and geometric parameters that indicate their accessibility by the other species.

The simplicity, speed and wide applicability of the method make it a useful tool in the selection of solvents for reaction tasks during the conceptual design of chemical processes, enabling a process chemist to screen hundreds of potential candidate solvents and select compounds that are effective reaction media as well as environmentally sound and easily recoverable. The application of the method to a number of chemical reactions will be discussed.

Thesis Supervisor: George Stephanopoulos

Arthur D. Little Professor of Chemical Engineering
Acknowledgments

On my second ever day of skiing a good friend of mine took me to the top of the mountain and down one of its steepest trails. I wish to thank Professor George Stephanopoulos for replicating the sensation of the ride down over the course of my research. The way was hard, full of bumps and falls but with his guidance and support it was also an exciting experience and full of learning. Most important of all, I made it down alive! From Prof. Stephanopoulos not only did I learn about process engineering, but more importantly, his unique perspective on how to approach a problem, any problem, in a methodical way. For this I am forever indebted to him.

Prof. Michael Mavrovouniotis provided invaluable help with his will to spend long hours in discussions that resulted in many of the insights that generated this work. He also provided constant encouragement and meaningful criticism.

I am also thankful to the members of my thesis committee, Professors Daniel Blankschtein, T. Alan Hatton and Paul Barton. They have shown genuine interest in this work and provided valuable advice and criticism. They steered me through many difficult points in the work and were critical for its conclusion.

Prof. Osmar Antônio Ferraz and Prof. Liane Barata instilled in me the interest in physical chemistry and motivated me to follow a career in Chemical Engineering. They also showed me by example the value of a strong work ethic.

Gabrielle Joseph, Janet Fischer, Elaine Aufiero, Carol Phillips, Arline Benford and Glorianne Collver-Jacobson provided invaluable help (usually at last notice) but also created an environment of harmony and friendship that made my work much easier. They helped me out with my research and the MIT bureaucracy, provided constant encouragement and even gave me guitar lessons and engaged me as an overnight concert producer. I will miss them dearly.

Orhan Karsligil, Matthew Dyer, Enrique Salomone, Manuel Rodriguez, Jerry Biesczcad, Ajay Modi, Alex Koulouris, Shahin Ali, Christine Ng, Eleni Stephanopoulos, Prof. Andreas Linninger, Prof. Chonghun Han, John “Sunshine” Carrier, Prof. Bhavik Bakshi, Seiichi Sugawa and Atsushi Isotani managed to work in the same laboratory as me and spent many hours in meaningful and not so meaningful discussions. They provided a great work environment with their dedication and support.

Thanks to Coach Bill Paine and the crew at MIT Masters Swimming, which I had the pleasure to run for three years. You provided much needed sanity and companionship. I’ll see you at the pool.

During my stay at MIT, I had the fortune of meeting Steffen Ernst, Seth Rodgers, Ravi Srinivasan, Glen Bolton and Allison Begner. Whether camping in the high desert of
Utah, skating around Boston at night, showing me around in Copenhagen or just being the sincere friends that they are, they have greatly enriched my life.

Let us not forget the Brazilian Mafia. Frederico Gueiros, Patricia Vidal, Lúcio, Lívia and Tânia Soibelman Guilherme Xavier, Mônica Pontes, Luís and Chris Pedrosa, Paulo and Flávia Gonçalves and Marcos Campos are true friends who shared in both the frustration and the excitement of research. I will always remember our times running the MIT Brazilian Student Association, organizing the infamous Carnival parties and the “only one four times World Champion” BSA soccer team. I hope the legitimate Brazilian soccer players will be able to forgive us for that.

My brother Bernardo basically pushed me through the final months of my research and spent long hours discussing both mine and his research. He also introduced me to the joys of surfing in New England. In winter! My sister Mariana, a.k.a. Dra. Jabiraca was an inspiration with her brilliant passage through Medical School.

Renata Pereira provided loving support and encouragement during the most difficult months and endured me at my worst. Without her I am not sure I would have finished this work.

Finally, I want to thank my parents, Walter and Jória. My father interrupted his own doctoral thesis at Sorbonne to work with my mother for the re-establishment of democracy in Brazil. They have thought me the meaning of personal sacrifice for one’s vision and to appreciate of the value of education. This work is dedicated with great love and respect to them.
Ah, my friend, the human species struggles to impose to the throbbing world a little bit of routine and logic, but something or somebody makes a farce of all this to laugh at us.

We live, incorrigibly, distracted from the most important things.

Guimarães Rosa
# Table of Contents

1. Introduction

1.1. Electrostatic Basis of Solvent Effects on Reaction Rates

1.2. Ab Initio Models

1.3. Linear Free Energy Relationships

1.4. Thesis Objectives

2. Solvent Charge Distribution and Conjugation Operators

2.1. Conjugation Operators

2.2. Calculation of Atomic Charges with Conjugation Operators

2.2.1. Properties of Conjugation Charge Equations

2.2.2. Aromatic Conjugation Operators and Charge Equations

2.2.3. Substituent Effects on Aromatic Rings

2.2.4. Non-aromatic Carbons

2.3. Quantum Basis of Conjugation Operators

2.4. Calculation of Conjugation Operator Weights

3. The Reaction Fingerprint

3.1. Transition State Basis of the Reaction Fingerprint

3.2. Statistical Mechanical Calculation of Free Energy of Solvation

3.2.1. Radial Distribution Functions

3.2.2. Free Energy of Solvation

3.2.3. Statistical Mechanical Description of the Reaction Fingerprint

3.2.4. The Radial Distribution Function as a Function of the Reaction Fingerprint

3.3. Implementation of the Reaction Fingerprint Model

3.3.1. Potential Energy

3.3.2. Solvent Modeling

3.3.3. Spherical Asymmetry

3.3.4. Obtaining the Parameters of the Reaction Fingerprint
List of Figures

Figure 2-1: Application of the HC, CO and HOCH operators to the methanol molecule with corresponding conjugation structures. ........................................ 25
Figure 2-2: The operator C=C and its inverse. ........................................ 25
Figure 2-3: The CCCH operator decomposed into its constituent single-bond operators 26
Figure 2-4: The OdnC operator. ........................................................ 27
Figure 2-5: Decomposition of the OCC operator. .................................... 27
Figure 2-6: Conjugation operators applied to the ethanol molecule ............. 29
Figure 2-7: Atomic charges of ethane, ethylene and acetylene. ................. 34
Figure 2-8: Resonant forms in benzene, phenol and toluene. .................. 36
Figure 2-9: Conjugation modes for the carbon atom in Benzene. .............. 37
Figure 2-10: Effects of the CARC operator. ....................................... 39
Figure 2-11: The CorthoC and CparaC operators. .................................. 39
Figure 2-12: Decomposition of the CparaC operator. ............................. 40
Figure 2-13: Conjugate forms generated by the CorthoC and CparaC operators. ____ 40
Figure 2-14: Decomposition of conjugate operators HCoCH and H CpCH in their individual operators. ......................................................... 41
Figure 2-15: Decomposition of the CORTHO and CPARA operators. ....... 42
Figure 2-16: Decomposition of the CPARAC operator ................................ 43
Figure 2-17: Conjugation structures combining aliphatic and aromatic effects. ____ 44
Figure 2-18: Decomposition of the CarCarCH and HCarCarC operators. Part I. 44
Figure 2-19: Decomposition of HCarCarC and CarCarCH operators. Part II. 45
Figure 2-20: Two resonant conjugation structures resulting from the CarCarCH operator. .............................................................. 45
Figure 2-21: Decomposition of the HCarCarC operator. ......................... 46
Figure 2-22: Molecular structure of phenol ........................................... 47
Figure 2-23: Decomposition of OCorthoC and OCparaC operators. .......... 48
Figure 2-24: Comparison between atomic charges calculated by Natural Population Analysis and those calculated by conjugation charge equations. .... 64
Figure 3-1: The reaction fingerprint in its simplest form: a spherical molecular entity of radius R consisting of a centrally located dipole with charge parameter $\delta$ and length parameter L.

Figure 3-2: Possible forms of the reaction fingerprint including ellipsoidal molecular shapes and asymmetrically positioned dipole.

Figure 3-3: Representation of a typical radial distribution function.

Figure 3-4: Atomic charges, molecular structure and dipole moment of ethanol. The gray circles refer to the centers of positive and negative charge. The actual solvent geometry as calculated by quantum mechanical optimization is shown. Charge units in $10^{-19}$ C.

Figure 3-5: The six degrees of freedom for the orientation of the solvent with respect to the reaction fingerprint.

Figure 3-6: Lowest energy orientations for a pair of dipoles: (a) longitudinal orientation and (b) parallel orientation.

Figure 3-7: Ethanol molecule aligned in the most favorable orientations in the full atomic charge model.

Figure 3-8: The distances from the solvent dipole charge and the orientation of the solvent with respect to the reaction fingerprint in the solvent dipole model with most favorable orientation.

Figure 4-1: Schematic representation of the application of the reaction fingerprint model.

Figure 4-2: Methyl iodide-iodide ion exchange reaction.

Figure 4-3: Comparison of experimental and reaction fingerprint reaction rates for the Menschutkin reaction between triethylamine and ethyl iodide.

Figure 4-4: Comparison of experimental and reaction fingerprint reaction rates for the Menschutkin reaction between triethylamine and ethyl iodide.

Figure 4-5: Results of reaction fingerprint calculation using solvents with anomalous behavior.

Figure 4-6: Response surface for Case Study 1 - Least Squares error as a function of charge and dipole length parameters.

Figure 4-7: Response surface for Case Study 1 - Least Squares error as a function of charge and sphere radius parameters.
Figure 4-8: Response surface for Case Study 1 - Least Squares error as a function of dipole length and sphere radius parameters. 120
Figure 4-9: Menschutkin reaction rate as a function of the charge parameter. 121
Figure 4-10: Menschutkin reaction rate as a function of the sphere radius parameter. 122
Figure 4-11: Menschutkin reaction rate as a function of the dipole length parameter. 123
Figure 4-12: Menschutkin reaction rate as a function of the dipole length parameter with potential energy calculated using only the parallel dipole configuration. 124
Figure 4-13: Menschutkin reaction rate as a function of the dipole length parameter with potential energy calculated using only the longitudinal dipole configuration. 126
Figure 4-14: Schematic representation of the interactions of the dipolar representation of a solvent (left) and its full atomic charge representation (right) with the reaction fingerprint. 127
Figure 4-15: Results of reaction fingerprint calculation for the Menschutkin reaction using full atomic charge description. Apolar solvents are included. 128
Figure 4-16: Correlation between solvent dielectric constant and reaction rates for the reaction between triethylamine and methyl iodide. 130
Figure 4-17: Correlation between solvent Kirkwood parameter and reaction rates for the reaction between triethylamine and methyl iodide. 131
Figure 4-18: Average squared log(k) error as a function of charge and dipole length using the full atomic charge model. Radius = 1.5x10^-10 m. 132
Figure 4-19: Average squared log(k) error as a function of radius and charge using the full atomic charge model. Dipole length = 2.8x10^-10 m. 133
Figure 4-20: Average squared log(k) error as a function of radius and dipole length using the full atomic charge model. Charge = 0.05x10^-19 C. 133
Figure 4-21: Sensitivity analysis of reaction rate with respect to the reaction fingerprint’s charge parameter - Menschutkin reaction between ethyl iodide and triethylamine using full atomic charges. 134
Figure 4-22: Sensitivity analysis of reaction rate with respect to the reaction fingerprint’s dipole length parameter - Menschutkin reaction between ethyl iodide and triethylamine using full atomic charges. 135
Figure 4-23: Sensitivity analysis of reaction rate with respect to the reaction fingerprint’s radius parameter - Menschutkin reaction between ethyl iodide and triethylamine using full atomic charges.

Figure 4-24: Application of the reaction fingerprint parameters from the reaction between Et$_3$I + Et$_3$N to the reaction between ICH$_2$COOC$_2$H$_5$ + Et$_3$N using the solvent dipole model.

Figure 4-25: Application of the reaction fingerprint parameters from the reaction between Et$_3$I + Et$_3$N to the reaction between ICH$_2$COOC$_2$H$_5$ + Et$_3$N using the full atomic charge model.

Figure 4-26: Correlation between solvent dielectric constant and reaction rates for the reaction ICH$_2$COOC$_2$H$_5$ + Et$_3$N.

Figure 4-27: Correlation between solvent Kirkwood parameter and reaction rates for the reaction ICH$_2$COOC$_2$H$_5$ + Et$_3$N.

Figure 4-28: Formation of charge transfer complex between bromine and 1-pentene.

Figure 4-29: First-order mechanism for bromination of 1-pentene: the limiting step is the formation of bromonium ion.

Figure 4-30: Second-order mechanism for bromination of 1-pentene: the limiting step is a bimolecular nucleophilic attack in concert with bromine bond cleavage.

Figure 4-31: Comparison of experimental [65] and reaction fingerprint reaction rates for the bromination of 1-pentene.

Figure 4-32: Solvent electrophilic assistance to the transition state during limiting step in the bromination of 1-pentene.

Figure 4-33: Experimental rates [65] and reaction fingerprint calculation results for the bromination of 1-pentene using full atomic charges.

Figure 4-34: Correlation between solvent dielectric constant and reaction rates for the bromination of 1-pentene.

Figure 4-35: Correlation between solvent Kirkwood parameter and reaction rates for the bromination of 1-pentene.

Figure 4-36: Average squared log(k) error as a function of charge and dipole length using the full atomic charge model for the bromination of 1-pentene. Radius = 0.6 Å.
Figure 4-37: Average squared log(k) error as a function of radius and dipole length using the full atomic charge model for the bromination of 1-pentene. Charge = $0.5 \times 10^{-19}$ C. 152

Figure 4-38: Average squared log(k) error as a function of radius and charge using the full atomic charge model for the bromination of 1-pentene. Dipole length = $0.4 \, \text{Å}$. 152

Figure 4-39: Reaction rate sensitivities of t-butanol, acetic acid and water with respect to the reaction fingerprint charge parameter using the full atomic charge model. $R = 0.6 \, \text{Å}$, $l = 0.4 \, \text{Å}$. 153

Figure 4-40: Reaction rate sensitivities of t-butanol, acetic acid and water with respect to the reaction fingerprint dipole length parameter using the full atomic charge model. $q = 0.5 \times 10^{-19}$ C, $l = 0.4 \, \text{Å}$. 154

Figure 4-41: Reaction rate sensitivities of t-butanol, acetic acid and water with respect to the reaction fingerprint dipole length parameter using the full atomic charge model. $q = 0.5 \times 10^{-19}$ C, $R = 0.6 \, \text{Å}$. 155

Figure 4-42: Schematic representation for orientation sampling (right) as opposed to using only the most favorable orientations (left). 156

Figure 4-43: Experimental rates and reaction fingerprint calculation results for the bromination of 1-pentene using full atomic charges. 157

Figure 4-44: Average squared log(k) error as a function of radius and dipole length using the full atomic charge and orientation sampling model for the bromination of 1-pentene. Charge = $0.3 \times 10^{-19}$ C. 160

Figure 4-45: Average squared log(k) error as a function of radius and charge using the full atomic charge and orientation sampling model for the bromination of 1-pentene. Dipole length = $0.9 \, \text{Å}$. 160

Figure 4-46: Average squared log(k) error as a function of charge and dipole length using the full atomic charge and orientation sampling model for the bromination of 1-pentene. Dipole length = $0.9 \, \text{Å}$. 161

Figure 4-47: Solvolysis of t-butyl chloride. 162

Figure 4-48: Comparison of reaction fingerprint results and experimental results for solvolysis of t-butyl chloride in mixtures of water and ethanol. 164
List of Tables

Table 2-1: Effects of the PARA conjugation on carbon-carbon bonds in both resonance forms as shown in Figure 2-9. 38
Table 2-2: Compounds used in the calculation of conjugation operator weights. 63
Table 2-3: Weights of the conjugation operators. 65
Table 4-1: Methyl iodide-iodide ion exchange reaction rates in test solvents. 105
Table 4-2: kinetic rate and free energy of transfer (with formamide as reference value) data for the reaction between triethylamine and ethyl iodide [56]. 109
Table 4-3: Reaction fingerprint calculation results for the Menschutkin reaction between triethylamine and ethyl iodide. 113
Table 4-4: Comparison of experimental and calculated rates for the Menschutkin reaction between ethyl iodide and triethylamine using full atomic charges. 129
Table 4-5: Absolute average error and correlation coefficient for linear correlation between reaction fingerprint calculated rates, dielectric constant, Kirkwood parameter and experimental reaction rates. 131
Table 4-6: application of the reaction fingerprint parameters from the reaction between Et$_3$I + Et$_3$N to the reaction between ICH$_2$COOC$_2$H$_5$ + Et$_3$N using the solvent dipole model. 138
Table 4-7: application of the reaction fingerprint parameters from the reaction between Et$_3$I + Et$_3$N to the reaction between ICH$_2$COOC$_2$H$_5$ + Et$_3$N using the full atomic charge model. 140
Table 4-8: Absolute average error and correlation coefficient for linear correlation between reaction fingerprint calculated rates, dielectric constant, Kirkwood parameter and experimental reaction rates. 142
Table 4-9: Experimental rates [65] and reaction fingerprint calculation results for the bromination of 1-pentene. 145
Table 4-10: Experimental rates [65] and reaction fingerprint calculation results for the bromination of 1-pentene using full atomic charges. 148
Table 4-11: Absolute average error and correlation coefficient for linear correlation between reaction fingerprint calculated rates, dielectric constant, Kirkwood parameter and experimental reaction rates. 150
Table 4-12: Experimental rates and reaction fingerprint calculation results for the bromination of 1-pentene using full atomic charges and orientation sampling. 158
1. Introduction

Solvents are widely used in chemical processes and have a direct impact on the environmental impact, efficiency and profitability of a manufacturing scheme. High toxicity can lead to the need of extensive downstream processing and recycling, which impacts the cost of a process. On the other hand, a solvent that is easily recovered and affords high reaction rates can reduce operating costs and batch time. Moreover, by choosing a suitable set of solvents, the number of distinct solvent as well as the complexity of the process can be reduced.

Solvents are employed in process tasks such as separation, reaction and cleaning [1]. As a separation agent, a solvent can be an entrainer in azeotropic distillation, an extraction solvent or a crystallization medium among other functions. In a reaction task, the solvent serves the dual purpose of dissolving the reactants and reducing the activation energy of the reaction so as to accelerate its kinetics.

There is a wealth of literature on the selection [2] [3] [4] [5] and the design of solvents for separation applications. This is due to the fact that there are models for the calculations of activity coefficients such as UNIFAC [6] that enable not only the selection of solvents from a database, but also the design of solvent molecules through the assembly of molecular groups. Design methods [7, 8] [9] follow two distinct paradigms. The first is the generation of molecular structures and subsequent pruning based on physical property constraints for distillation [10], liquid extraction[11], mixtures with property constraints [12] [13] and general separation processes [14]. The other approach is based on discrete optimization based on an objective function embodying the desired physical properties of the solvents where the optimization variables are the chemical groups and their connections [15] [16] [17].

Although there are phase equilibrium models in place for the selection of solvents for separation tasks, the problem remains of evaluating solvents as reaction media. A model
for the calculation of solvent effects on reaction rates that is accurate for process design purposes is the last building block needed to enable rational selection of solvents that satisfy plant-wide requirements [18] such as good performance in its specific task but also applicability to other process tasks, low environmental impact and toxicity and ease of recovery. The main difficulty in assessing solvents as reaction media is the dependence of the reaction activation energy on not only the reactants, but also the transition state [19]. Since it is very difficult to obtain transition state geometries and charge distributions except for the simplest cases, this hinders the development of solution kinetic models that are fast and have wide applicability.

Ab initio quantum mechanical calculations have been extensively used to determine the structure of the transition state and to model solution kinetics. Although in some cases they allow accurate predictions of the structure of the activated complex and its “activity coefficient”, enabling the calculation of kinetic rates from first principles, they are not suitable for process design applications due to their excessive detail level and corresponding data demand, computational expense and limited range of applicability (quantum mechanical models are tailored to specific systems). The challenge, as in many other process design endeavors, is to develop a model that correctly balances data needs, complexity and accuracy.

1.1. Electrostatic Basis of Solvent Effects on Reaction Rates

Solvents greatly influence reaction rates through differential solvation of both reactants and activated complex. Differences on solvation energy are mainly due to changes in charge distribution in the reacting species that occur during the transition from reactants to transition state. Therefore, one would assume that differences in the Gibbs’ free energy of solvation for the reacting species would be mainly due to enthalpy contributions. Indeed, it has been shown by abundant experimental evidence that enthalpy effects are largely dominant [20] [21] in most reaction systems, and therefore by modeling kinetic rates through electrostatic interactions, one should have enough
information to calculate rates with reasonable accuracy. This led to the development of a formal classification of reactions based on charge distribution as well as a set of rules that allowed qualitative analysis of solvent effects [22]:

1) Reactions in which there is an increase in charge magnitude during activation. 
2) Reactions in which there is destruction of charge during activation. 
3) Reactions in which there is dispersion of charge during activation. 
4) Reactions in which there is no appreciable change in charge density during activation.

a) Reactions of type 1 will be accelerated by an increase in solvent polarity and slowed by a decrease in solvent polarity.
b) Reactions of type 2 will be accelerated by a decrease in solvent polarity and slowed by an increase in solvent polarity.
c) Reactions of type 3 will be accelerated by a decrease in solvent polarity and slowed by an increase in solvent polarity, but to a lesser extent than reactions of types 1 and 2.
d) Reactions of type 4 will not be significantly affected by changes in solvent polarity.

To understand these rules it is useful to consider the case of a solvolysis reaction, whereby a neutral molecule is broken into ions. In a low polarity solvent, the reactants have a lower energy of solvation while the transition state, which is more polar, will not be solvated as well and will have a higher energy of solvation. To increase the polarity of the solvent will make the reactants’ energy of solvation increase while decreasing the energy of solvation of the transition state, which will be better solvated in a higher polarity solvent. The net result when solvent polarity is increased in such a reaction is to reduce the activation energy of the reaction and, as result, to accelerate its rate. This change in the activation energy is proportional to the extent that charge is being created or concentrated. Reactions where charge is created in the activation process, such as solvolysis, will have much more pronounced solvent effects than reactions where charge is concentrated or there is a change in geometry that increases the dipole moment during the activation step.
Solvent effects have been modeled quantitatively by *ab initio* studies and correlations between the free energy of activation in different media.

### 1.2. *Ab Initio* Models

*Ab initio* models [23] require the three-dimensional structures and charge distributions of both reactants and transition state. In the Reaction Field Theory framework, the solvent is modeled as a continuum with a uniform dielectric constant $\varepsilon$. The reactants are modeled as a cavity in the continuum where atomic charges are located. The reaction field $\xi$ is a function of solvent parameters

$$
\xi = 2(\mu / a^3) \left( \frac{d - 1}{2d + 1} \right)
$$

(1-1)

where $a$ is the diameter of the solute cavity, $d$ is the dielectric constant of the solvent and $\mu$ the dipole moment of the solute. The term above represents the screening energy of a dipole in a cavity immersed in a dielectric continuum. Once the field has been calculated, the free energy of solvation $\Delta G^{sol}$ of the solute is given by

$$
\Delta G^{sol} = -c \xi \mu
$$

(1-2)

where $c$ is a constant that depends only on the unit system being used. The main difficulty in reaction field calculations is to estimate the geometry and the dipole moment of the transition state. It is also necessary to estimate the radius of the cavity, $a$. 

19
1.3. Linear Free Energy Relationships

Linear Free Energy Relationships are models that correlate the free energy of activation of reactions in different media. The free energy of activation $\Delta G^\ddagger$ is postulated to be a linear function of some polarity measure $Y$ of the solvent

$$\Delta G^\ddagger = aY + b \quad \text{(1-3)}$$

The coefficients $a$ and $b$ are determined from kinetic data in at least two distinct solvents. One suggested measure of solvent polarity is the logarithm of the ratio of the rate of solvolysis of t-butyl chloride in the solvent $k$ and the rate in a standard solvent $k_0$.

$$Y = \log \left( \frac{k}{k_0} \right) \quad \text{(1-4)}$$

Linear free energy relationships are not applicable to all kinds of reactions and generally incur large errors.

1.4. Thesis Objectives

Given that *ab initio* models are highly customized for each case and that statistical techniques such as Linear Free Energy Relationships and other statistical mappings from physical properties to reaction rates do not allow the selection of solvents with acceptable accuracy for process design purposes, the objective of this work is to develop a model for the calculation of solvent effects on reaction rates that is fast, applicable to a wide variety of reactions and solvents and that yields results that are accurate enough to allow the selection of solvents. These requirements pose the problem of modeling solution kinetics in enough detail to obtain reasonable accuracy but at the same time make simplifications in order to keep the model general and fast. Rather than a statistical aritifice, the model
should also be rigorous and have a molecular basis.

The electrostatic basis of solvent effects on reaction rates indicates that two types of information are needed: the charge distribution of the solvent and the charge distribution change that takes place when the reaction moves from reactants to transition state.

In order to model the charge distribution of the solvent, the concept of conjugation operators [24] has been applied to the calculation of atomic charges. Conjugation operators act on a molecule localizing the electrons in one bond in one of the respective atoms, thereby creating conjugation structures. Atomic charges in the molecule are modeled as the result of the interplay of different modes of conjugation in the form of conjugation charge equations which have assigned weights to each conjugation mode. These weights are obtained from \textit{ab initio} atomic charge calculations. It is demonstrated that charge conjugation equations can be obtained directly from the population matrix obtained in quantum mechanical calculations. Although in this work conjugation operators are used to calculate solvent atomic charges, they can be applied to other problems, such as the determination of activity coefficients in nonideal mixtures.

In modeling the reaction, the main problem to be tackled is the difficulty of obtaining the transition state’s geometry and charge distribution. Since reaction rates are a function of the difference between the energies of the transition state and the reactants, a molecular entity termed the reaction fingerprint is introduced that has the property that its energy of formation in solution is equal to the difference of energies of formation in solution between transition state and reactants, i.e. the activation energy in solution. Since the reaction fingerprint is a molecular entity, its energy of solvation is calculated in the realm of statistical mechanics. The reaction fingerprint has adjustable parameters which describe its polarity and the geometric availability of its charge. These parameters are adjusted so as to minimize the error between actual experimental values of activation energies and those calculated by the reaction fingerprint model. Once the parameters for the reaction fingerprint corresponding to a reaction are obtained, reaction rates for solvents whose rates are not known can be obtained. This is done by calculating the
statistical mechanical energy of solvation of the reaction fingerprint in the new solvent.

The reaction fingerprint model has been applied to several reactions with very different charge transfer characteristics and involving a wide variety of solvents. In general it is able to calculate reaction rates in solution with an error smaller than one order of magnitude.
2. Solvent Charge Distribution and Conjugation Operators

The first building block in a model for the calculation of solvent effects on reaction rates is a description of the electrostatic charge distribution of the solvent. This is routinely done by assigning electric charges to the atomic centers of a molecule. This discretization of the continuous electronic distribution in a molecule obtained by quantum mechanical calculation is artificial and for these reasons there are numerous schemes for the determination of atomic charges. They can be obtained by partitioning electrons in overlapping molecular orbitals between their respective atoms, such as in the Hartree-Fock model [25, 26] or by partitioning a single electronic density function among the atoms, as is done in Density Functional Theory [27, 28]. Within these two themes, many variations are possible. Charges can be calculated so as to fit the electrostatic potential of the molecule [29], [30-33]. The charge partition can also be done so that the electronegativity of both atoms in the overlap is equaled in the distribution [34, 35]. Other models take into account the three-dimensional structure of the molecule [36] and can make charges dependent on the molecular geometry [37]. However, there is not a database for most commonly employed solvents, and charges obtained through a certain method are not easily transferred to another model in many cases [38].

To model the solvent charges, a model that encodes the atomic charges obtained from \textit{ab initio} calculations into conjugation operators has been developed. This model assigns weights to conjugation structures that are obtained by localizing the electrons in a bond in one of its corresponding atoms. Atomic charges are obtained by the summation of the weights of all operators that apply to an atom.
2.1. Conjugation Operators

Conjugation operators were originally developed for the calculation of macroscopic properties such as heat of formation or critical parameters [24, 39-43]. Conjugation operators act on molecules by dislocating electrons in molecular bonds and thus creating charge separations along one or more bonds. Conjugation operators are also specific to the kind of bond. For instance, the CH conjugation operator is applied to carbon-hydrogen bonds, allocating the pair of electrons in the bond to the hydrogen atom and creating the C\(^+\)...H\(^-\) conjugate form. Its symmetric operator, the HC operator, allocates the bond electron pair to the carbon atom creating the H\(^+\)...C\(^-\) conjugate form. Figure 2-1 exemplifies the application of some conjugation operators to the methanol molecule.

The theoretical foundation of conjugation operators is the concept that a molecule is a hybrid of its multiple conjugation structures and as a result its properties are a result of the contribution of each conjugation structure. The contributions of each conjugation structure can be statistically determined and used to calculate molecular properties. This work has extended the concept to the calculation of atomic-level properties (in this case, atomic charge) in both aliphatic and aromatic compounds.

Each conjugation operator has a weight associated to it. The weights of operators that span more than one bond are derived from single-bond operators. The CCCH operator, for instance, is derived by first applying the CC and CH operators to the end bonds of a CCCH chain. At this point, two bonds have been broken with the delocalization of two pairs of electrons. For the next step, the C=C operator must be considered. This operator acts on a double bond, delocalizing one of the two pairs of bonding electrons and creating a structure where the atoms previously involved in the double bond are charged and have a single bond. Its inverse operator, C=C\(^-\), creates the opposite effect on a single bond with two charged atoms, resulting in a single bond, as shown in Figure 2-2. It can be seen that the operator C=C\(^-\) can be applied to the central carbon atoms in the conjugation structure, creating the final conjugation structure as shown in Figure 2-3.
Figure 2-1: Application of the HC, CO and HOCH operators to the methanol molecule with corresponding conjugation structures.

\[
\text{C} = \text{C} \quad \xleftrightarrow{\text{C} = \text{C}^{-1}} \quad \text{C}^{+1} - \text{C}^{-1}
\]

Figure 2-2: The operator C=C and its inverse.
Figure 2-3: The CCCH operator decomposed into its constituent single-bond operators

The weight of the CCCH operator is calculated as the product of the weights of the single-bond operators. The reason for this is the fact that, if a single bond conjugation mode has a large energy impediment to it, i.e. its weight is small, then all multiple-bond conjugation modes derived from it should also be impeded.

\[ w_{CCCH} = w_{CC} \cdot w_{CH} \cdot \frac{1}{w_C} = C \]  

(2-1)

Besides conjugation operators that delocalize electrons that are in bonding orbitals, there exists another type of operators, whereby an atom with a free electron pair in a high energy orbital, such as oxygen or sulfur, can donate this electron pair, sharing it with an adjacent atom. An example of this kind of operator is the OdnC operator (the \( dn \) between the atoms denotes an electron donor operator), which acts on a single OC bond, creating a double bond in its place, as shown in Figure 2-4. The oxygen atom has a positive charge in the resulting conjugation structure as it donates the electron pair. An important characteristic of the OdnC operator is that its symmetric operator, CdnO, does not exist since the carbon atom does not have a free electron pair to donate. However, operators consisting of two atoms with available free electron pairs have their symmetrics, such as OdnN and NdnO.
Figure 2-4: The OdnC operator.

With the free-electron donor operators, it is possible to construct longer operators such as the OCC operator as shown in Figure 2-5. The OdnC operator creates a double bond and charged oxygen and carbon atoms. The CC operator then transfer the charge accumulated in the form of the donated electron pair to the terminal carbon, thereby transferring the oxygen free electron pair form the oxygen atom to the carbon atom.

Figure 2-5: Decomposition of the OCC operator.

The weight of the OCC operator is then given by

\[ w_{OCC} = w_{OdnC} w_{CC} \]  \hspace{1cm} (2-2)\n
The OCC operator does not have a symmetric operator since there is not a CdnC operator.

The defined conjugation operators can now be employed in the calculation of atomic charges.
2.2. Calculation of Atomic Charges with Conjugation Operators

The weights associated to conjugation operators describe the likelihood, or the energy barrier, for a valence electron to dislocate itself from a bonding orbital and become localized at an atomic center. The larger the weight of the conjugation operator associated to a bond, the easier it is for an electron in that bond to become localized along the direction specified by the operator. It follows that if one considers a single bond, the difference between the weights of operators is a measure of the asymmetry of charge distribution along that bond or, since orbitals are functions that describe the distribution of electrons over space, the dislocation of the overlapping region between the bonding orbitals from the bond’s center. Conjugation operators discretize this asymmetry in charge distribution by concentrating the charge on the atomic center. As a result, the atomic charge of an atom in a molecule is a result of the asymmetry of the distribution of charge in all the bonds, and chains of bonds, it is part of, and conjugation charge equations can be written for each atom in a molecule describing its charge as a function of conjugation operators that apply to that atom. Figure 2-6 shows the conjugation operators that act on the ethanol molecule represented as arrows connecting the atoms where the charge is localized. Double-ended arrows represent sets of symmetric operators whereas the OCH and OCC operators, which have no symmetrics, are represented by a single-ended arrow.

Following the principle that the difference in symmetric operator conjugation weights accounts for the asymmetry in charge distribution, conjugation charge equations can be written for each atom in the molecule.
The carbon atom connected to the carboxyl group is also connected to two hydrogen atoms and another carbon atom. Therefore, the OC, CO, OH and HO single bond operators apply to that atom, as follows in its conjugation charge equation:

$$q_{C1} = w_{CO} - w_{OC} + 2(w_{CH} - w_{HC})$$  \hspace{1cm} (2-3)$$

where $q_{C1}$ is the charge of the carbon atom (the subscript $C_1$ is used to differentiate this atom from the other carbon atom). The CO operator dislocates the bond electrons to the hydrogen atom creating a positive charge in the carbon atom and thus its weight has a positive sign; the OC operator, which creates a negative charge, has a negative weight sign. The same applies for the CH and HC operators, but as they apply to two carbon-hydrogen bonds the carbon atom is involved in, their weight is multiplied by two. The CC operator, which is its own symmetric, is canceled out in the equation, as there is no asymmetry in charge distribution in conjugations involving bonds with identical atoms.

The methyl group carbon atom is directly connected to three hydrogen atoms and a carbon atom; this atom is also affected by multiple bond conjugations.
The weights of the CH and HC operators are multiplied by three to express the contribution of three distinct CH/HC conjugation modes to the carbon atom’s charge. The free electron donor OCC conjugation operator expresses the influence of the carboxyl oxygen atom on the charge of the carbon atom. Finally, the CCOH (positive charge-inducing), and its symmetric, HOCC (negative charge-inducing) operators account for the conjugation mode involving the carboxyl hydrogen.

The oxygen atom’s charge is affected by single and multiple bond conjugation modes involving all the atoms in the molecule.

\[ q_{C_2} = 3(w_{CH} - w_{HC}) - w_{OCC} + w_{CCOH} - w_{HOCC} \]  \hspace{1cm} (2-4)

Besides the single bond conjugations, the oxygen atom is involved in three free electron donor (one OCC and two OCH modes) and one three bond conjugation modes (OCCH/HCCO). The OCC and OCH operators have no symmetrics as the carbon and hydrogen atoms have no free electrons.

The charge of the hydroxyl hydrogen (denoted by the subscript \( H_1 \) to differentiate it from the other hydrogen atoms) is calculated by

\[ q_{H_1} = w_{HO} - w_{OH} + w_{OCC} - w_{CCH} + 2(w_{OCH} - w_{HCOH}) \]  \hspace{1cm} (2-5)

where the two HOCH/HCOH conjugations refer to the two distinct hydrogen atoms connected the carbon atom closest to the hydroxyl group (denoted by the subscript \( H_2 \)), whose charge is calculated by
Because both hydrogen atoms are connected to the same atoms and in the same way, the conjugation charge equation is the same for both of them. Since the HCCH operator is its own symmetric operator, it cancels itself out and therefore does not appear in the charge conjugation equation.

The charges of the hydrogen atoms in the methyl group (denoted by the subscript \( H_3 \)) are given by

\[
q_{H_3} = w_{HC} - w_{CH} + w_{HOCC} - w_{CCH} + 2(w_{HCO} - w_{CO})
\]  

(2-8)

Again, because the hydrogen atoms have the same connectivity, the conjugation charge equation is the same for all three atoms.

In summary, atomic charges can be determined by expressions involving the weights of conjugation operators, which represent the asymmetry in charge distribution along a bond or set of bonds. The charge of an atom will then be determined by its connections and their number. This implies some properties of the conjugation charge equations that will be discussed below.

### 2.2.1. Properties of Conjugation Charge Equations

Conjugation charge equations have certain characteristics which, in order to be validated, need to examined with respect to atomic charge data obtained by \textit{ab initio} calculations.

The first property is that, by using the conjugation charge equations, \textit{the sum of all atomic charges in a neutral molecule is zero}. For instance, if the charge conjugations for each atom in the ethanol molecule, Eqs. 2-1 to 2-6, are summed (including repeated equations
for atoms with the same connectivity, such as the methyl group hydrogens), the conjugation operator weights cancel each other out and the result of the summation is zero, i.e., the sum of atomic charges calculated by conjugation charge equation is zero for a neutral molecule. This is due to the fact that a conjugation operator always acts on two atoms simultaneously creating a positive charge on one and a negative charge on the other. Therefore the weight of an operator will always appear with a positive sign on one charge conjugation operator and a negative sign in the other. Summing all the equations will make the weights of opposing signs cancel each other. This is an important characteristic that contributes to the self-consistency of the conjugation charge equations.

Conjugation charge equations also rely on the important assumption that atomic charges can be determined exclusively from the connectivity of molecules, i.e. which atoms are bonded and whether these bonds are single, double, triple or aromatic, and without any information on the geometry of the molecule. In \textit{ab initio} calculations, atomic charges are obtained by first defining the geometry of the molecule by minimizing the potential energy of interaction among the electronic orbitals. Atomic charges are then obtaining by allocating the electrons in non-bonding orbitals to their respective atom and dividing the population in the overlapping regions between bonding orbitals in some way between the bonding atoms. Since the bonding electrons define atomic charge and the overlapping regions of bonding orbitals depend on their arrangement in space, it follows that atomic charges depend on the geometry of the molecule. By creating conjugation charge equations that map connectivity to atomic charge without any information on molecular geometry, the assumption is implied that there is a strong correlation between connectivity and geometry. For instance, the connectivity of the atoms that make up a methyl group will determine a certain basic structure which on its turn will define the atomic charges of the atoms in the group; connections between this methyl group with other groups will affect its structure and as a consequence the atomic charges. The bond with the other group will be expressed in the conjugation charge equation in the form of conjugation operators that affect the atomic charge.
If the connectivity of the atoms in a molecule provides enough information to determine the atomic charges, then the charges of atoms whose conjugation charge equations are identical should be equal when calculated by *ab initio* methods. For instance, in the ethanol molecule, the charges of the three hydrogen atoms in the methyl group, which have the same conjugation charge equations, and those of the two hydrogen atoms connected to the carbon atom closest to the hydroxyl group, should be the same. Indeed, they are; atomic charges calculated by Hartree-Fock calculations using the 6-31G** basis set show that their charges are identical: The three methyl hydrogen atoms have the same charge of $0.222 \times 10^{-19}$ C while the two other hydrogens have a charge of $0.177 \times 10^{-19}$ C. Additionally, the same atomic charge calculations were conducted for other 80 compounds and the largest observed difference in atomic charges for atoms with the same conjugation charge equations was $0.002 \times 10^{-19}$ C, or less than 2%, which is a strong indication that the connectivity hypothesis is correct. Details of these charge calculations will be provided in a later section.

Not only the types of conjugations that act on an atom, but also their number, influence its charge. In other words, the charge of an atom will be a function of the kinds of atoms in its neighborhood and their number. This will be further explained when conjugation operators are examined in the context of quantum mechanics. As an illustration, the homologue series of hydrocarbons ethane, ethylene and acetylene with their atomic charges are shown in Figure 2-7.
The conjugation charge equations for the three molecules are:

Ethane:

\[ q_C = 3(w_{CH} - w_{HC}) \]  
\[ q_H = w_{CH} - w_{HC} \]  
(2-9)

Ethylene

\[ q_C = 2(w_{CH} - w_{HC}) \]  
\[ q_H = w_{CH} - w_{HC} \]  
(2-10)

Acetylene

\[ q_C = (w_{CH} - w_{HC}) \]  
\[ q_H = w_{CH} - w_{HC} \]  
(2-11)
Symmetrical conjugation modes such as HCCH are canceled out and do not appear in the above equations.

According to Equations 2-9 to 2-11, the charges of the carbon atoms in the acetylene molecule are two times smaller than in ethylene and three times smaller than in ethane, while the charges of the hydrogen atoms are all identical and remain the same for the three molecules. The atomic charges show in Figure 2-7 were obtained by a Hartree-Fock calculation using the 6-31G** basis set and Natural Population Analysis. They show that the charges of the carbon atoms closely follow the proportionality suggested by the charge conjugation equations and that the charges of the hydrogen atoms are roughly the same, virtually identical for ethane and ethylene with a small deviation for the acetylene molecule. Within the molecules, the conjugation equations suggest that the carbon charges are three times as big as the hydrogen charges in the ethane molecule, twice as big in ethylene and the same in acetylene. The calculated atomic charges show that this is rigorously true. This example confirms the additivity property of conjugation operators, i.e. the number of conjugation modes affects atomic charges. It also provides confirmation of the hypothesis that there exists a strong correlation between connectivity and atomic charges, as evidenced by the almost identical charges for hydrogen atoms with the same connectivity in three different molecules.

2.2.2. Aromatic Conjugation Operators and Charge Equations

Aromatic compounds contain bonds that cannot be classified as single, double or triple. Aromatic bonds form conjugated systems with highly mobile electrons and low energy barriers for the formation of conjugate structures. Substituent atoms connected to the aromatic ring also introduce aliphatic conjugation modes in the molecule and therefore atomic charges in an aromatic structure arise from the interplay of aromatic and, if applicable, aliphatic conjugation effects.
The resonant forms of benzene, phenol and toluene are shown in Figure 2-8:

![Diagram of benzene, phenol, and toluene](image)

Figure 2-8: Resonant forms in benzene, phenol and toluene.

If one examines the possible conjugation modes of benzene with the number 1 carbon atom in the top position, one obtains the conjugation forms shown in Figure 2-9:
Figure 2-9: Conjugation modes for the carbon atom in Benzene.
In accordance with the notation used to describe substituents in aromatic rings, conjugations that affect carbons one, two and three bonds removed from the original carbon atom are called ORTHO, META and PARA conjugations. The conjugation modes are shown applied to both resonant structures. Clearly, the META conjugation is not allowed as it creates a carbon atom (number 3) with 5 bonds. As a result, only ORTHO and PARA conjugations need to be considered.

The reason why aliphatic operators cannot be applied to aromatic structures is the fact that carbon-carbon bonds in an aromatic ring are a resonance hybrid between single and double. Therefore, aliphatic operators that act on single and double bonds are not applicable.

If one looks closer, however, it is possible to identify common characteristics in the two conjugate structures of benzene. For instance, in both resonant forms of the ORTHO conjugate structure, the effect of the operator is to decrease the order of the bond between carbon atoms 1 and 2, from double to single in the left form and from single to broken in the right form. Similarly, in the PARA conjugate set the effects are:

<table>
<thead>
<tr>
<th></th>
<th>Left Form</th>
<th>Right Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>atoms 1 and 2</td>
<td>decrease bond order</td>
<td>double to single</td>
</tr>
<tr>
<td>atoms 2 and 3</td>
<td>increase bond order</td>
<td>single to double</td>
</tr>
<tr>
<td>atoms 3 and 4</td>
<td>decrease bond order</td>
<td>double to single</td>
</tr>
</tbody>
</table>

Table 2-1: Effects of the PARA conjugation on carbon-carbon bonds in both resonance forms as shown in Figure 2-9.

Based on the common features between both resonant conjugation forms, one can then define a single bond aromatic operator CarC. This operator decreases the order of an aromatic bond, creating a broken aromatic bond with charge separation. The operator CARC is also defined. This operator acts on a hypothetical bond with higher order than an aromatic bond creating an aromatic bond with charge separation.
Figure 2-10: Effects of the CARC operator.

The operators CorthoC and CparaC are also defined in Figure 2-11:

Figure 2-11: The CorthoC and CparaC operators.
The CorthoC and CparaC operators can be related to the CarC and CARC operators. Obviously, the CorthoC operator is identical to the CarC operator. The CparaC operator is decomposed as shown below:

\[
\text{CarC} \rightarrow \text{CarC CARC}^{-1} \rightarrow \text{CarC CARC}^{-1}
\]

Figure 2-12: Decomposition of the CparaC operator.

And its weight is given by

\[
w_{CparaC} = w_{CarC}w_{CARC}^{-1}w_{CarC}
\]  
(2-12)

With the CorthoC and CparaC operators in place, one can now account for conjugation effects due to atoms connected to the aromatic carbons. In the specific case of benzene, only hydrogen atoms are attached to the aromatic ring and, therefore, conjugate forms can be generated such as:

Figure 2-13: Conjugate forms generated by the CorthoC and CparaC operators.
These operators can be decomposed in terms of both aliphatic and aromatic operators, as shown below:

\[ w_{HCoCH} = w_{HC}w_{CORTHOC}^{-1}w_{CH} = w_{HC}w_{CARC}^{-1}w_{CH} \]

\[ w_{HCpCH} = w_{HC}w_{CPARAC}^{-1}w_{CH} = w_{HC}[w_{CarC}w_{CARC}^{-1}]^{-1}w_{CH} \]  

(2-13)

Figure 2-14: Decomposition of conjugate operators HCoCH and HCpCH in their individual operators.

It should be noted that the CORTHOC and CPARAC operators are not the same as the CorthoC and CparaC operators. The CorthoC and CparaC operators create a charge separation in an aromatic ring, resulting in a charged broken ring, whereas the CORTHOC and CPARAC operators create a charge separation in a hypothetical ring of higher order than an aromatic ring resulting in a charged aromatic ring.

The CORTHOC and CPARAC operators can be decomposed as follows:
\[ w_{\text{CORTHO}} = w_{\text{CARC}} \]
\[ w_{\text{CPARA}}^{-1} = w_{\text{CarC}} w_{\text{CARC}}^{-2} \] (2-14)

Figure 2-15: Decomposition of the CORTHO and CPARA operators.
To the aromatic conjugation effects, aliphatic conjugation effects can be superimposed with the difference that these "aliphatic" operators refer to aromatic bonds rather than single and double bonds.
Figure 2-17: Conjugation structures combining aliphatic and aromatic effects.

The longer operators CarCarCH and HCarCarC can be decomposed as follows:

Figure 2-18: Decomposition of the CarCarCH and HCarCarC operators. Part I.

It can be seen from examination of the bonds between carbons 1 and 2 and between carbons 2 and 3 that, while the bond between carbons 2 and 3 has been broken by the operator CarC, the bond between atoms 1 and 2 is still intact, although there is charge
separation. It can be concluded that the charge separation between carbons 1 and 2 is the result of the application of an operator to a bond that has a higher order than an aromatic bond.

By applying the operator CARC, the following conjugation structure is obtained:

![Conjugation Structure with CARC](image)

**Figure 2-19:** Decomposition of HCarCarC and CarCarCH operators. Part II.

Examining the two possible resonance structures for this conjugate form:

![Resonance Structures](image)

**Figure 2-20:** Two resonant conjugation structures resulting from the CarCarCH operator.

In both structures, the order of the bond between carbons 1 and 2 has been increased by the CARC operator.

Therefore, the CarCarCH operator can be decomposed as follows:
Similarly, for the HCarCarC operator:

![Diagram showing the decomposition of HCarCarC](image)

\[ \text{HCarCarC} = \text{HC} (\text{CARC}^{-1}) \text{CarC} \]

Figure 2-21: Decomposition of the HCarCarC operator.

With these operators, it is now possible to calculate the charges in the benzene molecule:

**Carbon:**

\[ q_C = w_{CH} - w_{HC} + 2(w_{\text{CarCarCH}} - w_{\text{HC} \text{CarCarC}}) \tag{2-16} \]

**Hydrogen:**

\[ q_H = w_{HC} - w_{CH} + 2(w_{\text{HC} \text{CarCarC}} - w_{\text{CarCarCH}}) \tag{2-17} \]

### 2.2.3. Substituent Effects on Aromatic Rings

The ORTHO and PARA conjugation operators do not appear in the equations for the charges in benzene because they cancel each other due to the symmetry of the molecule. However, when substituents are introduced, the ORTHO and PARA operators account for the asymmetry in charge distribution.
Consider phenol, for instance:

![Molecular structure of phenol](image)

Figure 2-22: Molecular structure of phenol

The equations are:

**Carbon attached to oxygen:**

\[
q_C = w_{CO} - w_{OC} + 2(w_{CarCarCH} - w_{HCarCarC}) \quad (2-18)
\]

**ORTHO carbon**

\[
q_C = w_{CH} - w_{HC} + 2(w_{CarCarCH} - w_{HCarCarC}) + w_{CarCOH} - w_{HOCarC} + w_{OCoC} \quad (2-19)
\]

**META carbon**

\[
q_C = w_{CH} - w_{HC} + w_{CarCarCO} - w_{OCarCarC} + w_{CarCarCH} - w_{HCarCaC} \quad (2-20)
\]

**PARA Carbon**

\[
q_C = w_{CH} - w_{HC} + 2(w_{CarCarCH} - w_{HCarCarC}) - w_{OCparaC} \quad (2-21)
\]

**ORTHO hydrogen**

\[
q_H = w_{HC} - w_{CH} + w_{HCarCO} - w_{OCarCH} + 2(w_{HCarCarC} - w_{CarCarCH}) \quad (2-22)
\]
META hydrogen

\[ q_H = w_{HC} - w_{CH} + 2(w_{HCarCarC} - w_{CarCarCH}) \]  

(2-23)

PARA hydrogen

\[ q_H = w_{HC} - w_{CH} + w_{OCparaCH} + 2(w_{HCarCarC} - w_{CarCarCH}) \]  

(2-24)

Note that because oxygen has a free electron pair, it can donate that pair and therefore the operators OCparaC and OCorthoC can be defined as:

\[ w_{OCorthoC} = w_{OdnC}w_{CarC} \]

\[ w_{OCparaC} = w_{OdnC}w_{CparaC} = w_{OdnC}w_{CarC}w_{CarC}w_{CARC}^{-1} \]

Figure 2-23: Decomposition of OCorthoC and OCparaC operators.

The following operators are also defined:

\[ w_{OCarCarCH} = w_{OdnC}w_{CarC}w_{CarC}w_{CarC}w_{CH} \]  

(2-25)
\[ w_{OC_{para}CH} = w_{OC}w_{CPARAC}^{-1}w_{CH} \]  \hspace{1cm} (2-26)

where \[ w_{CPARAC} = w_{CarC}^{-1}w_{CARC}^2 \]  \hspace{1cm} (2-27)

Finally, the charge equation for the oxygen and hydrogen atoms in the carboxyl group can be written:

**Oxygen**
\[ qO = w_{OH} - w_{HO} + w_{OC} - w_{CO} + 2w_{OC_{O}C} + 2w_{O_{Car}CH} - 2w_{H_{Car}CO} + 2w_{OC_{Car}C} - 2w_{Car_{Car}C} + w_{OC_{para}CH} - w_{HC_{para}CO} + w_{OC_{pC}} \]  \hspace{1cm} (2-28)

**Hydroxyl hydrogen**
\[ qO = w_{HO} - w_{OH} + 2w_{HOC_{ar}C} - 2w_{Car_{COH}} \]  \hspace{1cm} (2-29)

### 2.2.4. Non-aromatic Carbons

In the toluene molecule, the carbon atom in the methyl group does not form an aromatic bond. Furthermore, it cannot donate an electron pair, unlike the oxygen atom in the phenol molecule. With these facts in mind, the charge equations for toluene are written down as follows:

**Top aromatic carbon:**
\[ q_{C} = 2(w_{Car_{Car}CH} - w_{H_{Car}CarC}) \]  \hspace{1cm} (2-30)

49
ORTHO carbon

\[ q_C = -CH - HC + 2(w_{CarCarCH} - w_{HCarCarC}) + 3(w_{CarCCCH} - w_{HCCarC}) \]  \hspace{1cm} (2-31)

META carbon

\[ q_C = w_{CH} - w_{HC} + w_{CarCarCC} - w_{CCarCarC} + 2(w_{CarCarCH} - w_{HCarCarC}) \]  \hspace{1cm} (2-32)

PARA carbon

\[ q_C = w_{CH} - w_{HC} + 2(w_{CarCarCH} - w_{HCarCarC}) \]  \hspace{1cm} (2-33)

Methyl carbon

\[ q_C = 3(w_{CH} - w_{HC}) + 2(w_{CCarCH} - w_{HCarCC}) + 2(w_{CCarCarC} - w_{CarCarCC}) \]  \hspace{1cm} (2-34)

ORTHO hydrogen

\[ q_H = w_{HC} - w_{CH} + w_{HCarCC} - w_{CCarCH} + 2(w_{HCarCarC} - w_{CarCarCH}) \]  \hspace{1cm} (2-35)

META hydrogen

\[ q_H = w_{HC} - w_{CH} + 2(w_{HCarCarC} - w_{CarCarCH}) \]  \hspace{1cm} (2-36)

PARA hydrogen

\[ q_H = w_{HC} - w_{CH} + 2(w_{HCarCarC} - w_{CarCarCH}) + w_{HCparaCC} - w_{CCparaCH} \]  \hspace{1cm} (2-37)

It is now necessary to examine the quantum mechanical meaning of aliphatic and aromatic conjugation operators.
2.3. Quantum Basis of Conjugation Operators

The weights of conjugation operators can be described as the coefficients of a decomposition of the density matrix that can be obtained from quantum calculations [44] [45]. To understand this decomposition one needs to know how atomic charges are calculated using molecular orbitals. The calculation will be demonstrated for the case of closed shell systems, i.e. molecules where all occupied orbitals contain one electron pair.

The state of any system is given in quantum terms by Schrödinger’s Equation

\[ H \psi = E \psi \]  \hspace{1cm} (2-38)

where \( H \) is the Hamiltonian operator, which applied to the wavefunction \( \psi \) yields the total energy of the system. According to Schrödinger’s Equation, the total energy of the system is equal to the product of its energy \( E \) by its wavefunction. In other words, the wavefunction of the system is the eigenvector of the Hamiltonian operator and \( E \) its eigenvalue.

In a molecule, Schrödinger’s Equation applies to a system of electrons and nuclei. However, since electron speeds are much higher than those of nuclei, electrons are assumed to change their configuration instantaneously given a change in the positions of the nuclei. For this reason, the electronic wavefunction can be calculated independently of that of the nuclei, once the nuclei coordinates are defined. This is known as the Born-Oppenheimer Approximation.

Since Schrödinger’s Equation cannot be solved exactly but for simple systems, wavefunctions for pairs of electrons, also known as molecular orbitals, are defined as a linear combination of basis functions.

\[ \psi_i = \sum_j c_{ij} \phi_j \]  \hspace{1cm} (2-39)
where $c_{ij}$ are the coefficients of the basis functions $\phi_i$. These basis functions are known in advance. They can be three-dimensional Gaussians, for instance or more complicated forms. The basis functions define the shape of the molecular orbital and consequently the shape of the three-dimensional distribution of electrons; they are centered on the nuclei coordinates. Therefore the basis function set is defined so as to obtain a faithful representation of the electronic distribution. The choice of basis set has great impact on the accuracy of the results. In general, the more accurate a basis set is, the higher the computational cost of the calculation. Once a basis set has been chosen, the problem of calculating the quantum state of the electrons in the molecule becomes the determination of the coefficients $c_{ij}$ of the basis functions. These are determined based on the variational principle that, if the wavefunction $\psi$ is the exact wavefunction of the system, its energy $E$ is minimized

$$E = \int \psi^* \mathbf{H} \psi \, dr$$  

Therefore, once a basis set has been picked, the above equation can be minimized with respect to the coefficients of the basis functions, with the constraint that the basis functions be orthonormal.

$$\int \phi_i^* \phi_j^* \, dr = \delta_{ij}$$  

The constraints are incorporated in the minimization by using the Lagrange multiplier framework, enabling the calculation of the coefficients of the basis functions.

The number density of electrons in space of a closed system is given by multiplication of the orbital wavefunctions $\psi(r)$ by their complex conjugate $\psi^*(r)$. Since it is a closed system, each orbital contains exactly two electrons and therefore the electronic density $n(r)$ at coordinate $r$ is given by the summation of all orbital products multiplied by two.
\[ n(r) = 2 \sum_i \psi_i^* (r) \psi_i (r) \]  

(2-42)

The results are easily extended to open systems. Since the orbital wavefunctions \( \psi_i \) are a linear combination of basis functions \( \phi_j \), the above equation can be written as

\[ n(r) = 2 \sum_i \left[ \sum_j c_{ij} \phi_j^* (r) \sum_j c_{ij} \phi_j (r) \right] = 2 \sum_i \sum_j c_{ij}^2 \phi_j^* (r) \phi_j (r) = \sum_i \sum_j P_{ij} \phi_i^* (r) \phi_j (r) \]  

(2-43)

To obtain the total number of electrons \( N \) in the system is obtained by integrating the electronic density \( n(r) \) over space.

\[ N = \int n(r) dr = \sum_i \sum_j P_{ij} \int \phi_i^* (r) \phi_j (r) dr = \sum_i \sum_j P_{ij} S_{ij} \]  

(2-44)

where \( S_{ij} \) is called the overlap integral between basis functions \( \phi_i \) and \( \phi_j \). The matrix \( S \) containing the overlap integrals \( S_{ij} \) is called the overlap matrix. Since the basis function set is orthonormal, the diagonal elements of the overlap matrix are equal to unity and the term \( P_{ij} S_{ij} \) in the population matrix becomes \( P_{ii} \) in the diagonal elements. \( P_{ii} \) is referred to as the net population of the basis function \( \phi_i \). Therefore, the total number of electrons associated to a basis function is given by:

\[ p_i = P_{ii} + \sum_{j \neq i} P_{ij} S_{ij} \]  

(2-45)

Since each basis function is assigned to an orbital and orbitals are assigned to atoms by being centered about the atom’s coordinates, the electrons associated to a certain basis
functions can also be assigned to an atom. The term $P_{ij}$ represents the population of electrons that is associated to single basis function and, by extension, to a single orbital and a single atom; these are electrons in low energy orbitals close to the atomic nuclei. The summation of $P_{ij}S_{ij}$ terms refers to those electrons in overlapping orbitals, i.e. electrons in high energy orbitals, or valence electrons involved in bonding composed by the common regions of the orbitals.

Using the population matrix, it is possible to discretize the continuous distribution of electrons represented by the orbitals, assigning a discrete atomic charge to the nuclei coordinates. Atomic charges are fundamental to molecular simulation and modeling as calculations of interactions between orbitals are very computationally expensive. In order to calculate atomic charges, it is necessary to partition the electrons in overlapping orbitals among their related atoms. Since this discretization is artificial, there are multiple ways by which the overlap electronic populations can be partitioned. The simplest of them, Mulliken Population Analysis, simply divides the overlap populations equally between the atoms corresponding to the overlapping orbitals. The charge $q_i$ of an atom can be calculated by subtracting the total electronic population $T_i$ assigned to that atom in the molecule, which is the sum of all the electrons assigned to basis functions which comprise orbitals associated to that atom, $p_i$, from the number of electrons in the isolated atom, $N_i$. Using Mulliken Population Analysis, atomic charges are given by

\[ q_i = N_i - T_i = N_i - \sum_i p_i = N_i - \sum_i P_{ii} - \frac{1}{2} \sum_{ij} \sum_{j \neq i} P_{ij}S_{ij} \]  

(2-46)

Since the partition of electrons in overlapping orbitals is strictly ad hoc and atomic charges are not a measurable property, the validity of different schemes for the calculation of atomic charges is determined by how well the charges model the behavior of the molecules in simulations or how well they can be used to estimate molecular properties such as the dipole moment [44]. It has been shown [46] that charges calculated by the Natural Population Analysis model [47] [48] using the 6-31G** basis
set best model molecular properties. The Natural Population Analysis uses only information from the population matrix based on the concept of obtaining orbitals termed Natural Bond Orbitals (NBOs) which provide description of the delocalized basis functions in terms of localized Lewis structures. This is done by maximizing the contributions of Lewis-type orbitals to the electronic density.

It is now possible to express conjugation operators within the framework of molecular orbitals. If one considers a system consisting of 2 basis functions, one for each of two atoms, the population $Q$ of the system is of the form

$$Q = \begin{bmatrix} a & c \\ c & b \end{bmatrix}$$

(2-47)

$Q$ is a symmetric matrix as the off-diagonal elements correspond to the overlap between the same basis functions. There are two conjugation structures $Q_1$ and $Q_2$ that can be obtained from $Q$, involving the assignment of overlap electrons to a single atom.

$$Q_1 = \begin{bmatrix} a + c & 0 \\ 0 & b \end{bmatrix} \quad Q_2 = \begin{bmatrix} a & 0 \\ 0 & b + c \end{bmatrix}$$

(2-48)

The actual electronic population of the first atom is somewhere between $a$ and $a+c$ and, for the second atom, between $b$ and $b+c$.

$$P_A = a + fc$$
$$P_B = b + (1 - f)c$$

(2-49)

$$0 \leq f \leq 1$$

The variation in electronic population is given by the difference in the assigned populations between the distribution of the conjugation structure and that of the actual configuration.
\[ \Delta P_1^A = (a + c) - (a + fc) = (1 - f)c \]
\[ \Delta P_2^A = a - (a + fc) = -fc \]
\[ \Delta P_1^B = b - [b + (1 - f)c] = -(1 - f)c \]
\[ \Delta P_2^B = (b + c) - [b + (1 - f)c] = fc \]

where \( \Delta P_i^A \) refers to the change in population in atom A from the actual population to the conjugate structure \( Q_i \).

The actual population of both atoms can be written as the sum of the number of electrons in the isolated atoms A and B, \( N_A \) and \( N_B \) and a linear combination of the population changes in all the conjugate forms.

\[
P_A = N_A + e_1 \Delta P_1^A + e_2 \Delta P_2^A = N_A + e_1 (1 - f)c - e_2 fc
\]
\[
P_B = N_B + e_1 \Delta P_1^B + e_2 \Delta P_2^B = N_B - e_1 (1 - f)c + e_2 fc
\]

where \( e_1 \) and \( e_2 \) are the coefficients of the conjugation modes. The atomic charge is equal to the difference between the number of electrons in the isolated atom and the actual atomic population.

\[
q_A = N_A - P_A = -e_1 (1 - f)c + e_2 fc
\]
\[
q_B = N_B - P_B = e_1 (1 - f)c - e_2 fc
\]

where the terms with negative signs refer to conjugations that increase the electron population of an atom and those with positive charge decrease the population. The two equations cancel each other and a conjugation mode which adds electrons to one atom and has a negative sign will have the opposite sign in the other atom. The conjugation operator weights can then be defined as
\[ w_1 = e_1 (1 - f) c \]
\[ w_2 = e_2 f c \]  
(2-53)

And the charge equation can be cast in conjugation operator form

\[ q_A = -w_1 + w_2 \]
\[ q_B = w_1 - w_2 \]  
(2-54)

To prove the additivity of conjugation modes, the same analysis will be conducted for the water molecule. The population matrix of water as calculated using the Hartree-Fock Method using the 6-31G** basis set is

\[
P = H \begin{bmatrix}
0.314 & 0.376 & -0.025 \\
0.314 & -0.025 & 0.376
\end{bmatrix}
\]

In this matrix, the population of every atomic orbital pertaining to an atom has been condensed in the diagonal matrix element corresponding to that atom; each off-diagonal element contains half of the overlap population. For instance, the overlap population between the hydrogen and oxygen atoms is 0.314 + 0.314 = 0.628. Since the hydrogen atoms have the same connectivity, their overlap populations with the oxygen atom is the same. This is a fact that has been observed even in asymmetric molecules. The conjugation structures are:
The population changes for the atoms in each conjugation mode are

\[
\begin{align*}
\Delta P_1^0 &= (8.043 + 2 \times 0.314) - (8.043 + 2 \times f \times 0.314) = 0.628(1 - f) \\
\Delta P_2^0 &= 8.043 - (8.043 + 2f \times 0.314) = -0.628f \\
\Delta P_3^0 &= (8.043 + 2 \times 0.314) - (8.043 + 2f \times 0.314) = 0.628(1 - f) \\
\Delta P_4^0 &= 8.043 - (8.043 + 2f \times 0.314) = -0.628f
\end{align*}
\]
Two important points about the population change equations must be noted: the first is that since both hydrogen atoms have the same connectivity, their overlap populations with the oxygen atom are the same. Therefore the partition of the overlap population is the same for both atoms and as a consequence the factor \( f \) is the same. The second point is that since conjugation modes are independent, the population of the hydrogen atom \( H_1 \) is not affected by the conjugation modes 3 and 4 and therefore there is no change in its population. The same happens for the \( H_2 \) in conjugation modes 1 and 2.

The equations for the atomic electronic populations are

\[
P_O = N_O + e_{HO} \Delta P_1^O + e_{OH} \Delta P_2^O + e_{HO} \Delta P_3^O + e_{OH} \Delta P_4^O = \\
8 + 2e_{HO} 0.628(1 - f) - 2e_{OH} 0.628f
\]

\[
P_{H_1} = N_H + e_{HO} \Delta P_1^{H_1} + e_{OH} \Delta P_2^{H_1} + e_{HO} \Delta P_3^{H_1} + e_{OH} \Delta P_4^{H_1} = \\
2 - e_{HO} 0.628(1 - f) + e_{OH} 0.628f + e_{HO} 0 + e_{OH} 0
\]
And, finally, casting the equations in conjugation operator form:

\[
\begin{align*}
ph_2 &= N_h + e_{HO} \Delta p_{1 H_2} + e_{OH} \Delta p_{2 H_2} + e_{HO} \Delta p_{3 H_2} + e_{OH} \Delta p_{4 H_2} = \\
          & \quad 2 - e_{HO} 0 + e_{OH} 0 - e_{HO} 0.628(1 - f) + e_{OH} 0.628 f
\end{align*}
\]  

(2-58)

As a result, conjugation charge equations have been derived directly from the population matrix, proving that atoms with the same connectivity have identical charge conjugation equations and that not only the types but also the number of conjugation modes that apply to an atom determine its atomic charge. Using Natural Population Analysis, the atomic charges of water are:

Oxygen charge = -0.972
Hydrogen charges = 0.486

The partition coefficient for these charge is \( f = 0.739 \). Since there is only one linearly independent charge conjugation equation, the weights of the operators cannot be calculated. In reality, the operator weights are calculated using data from several molecules so as to minimize the error between the charges calculated by the Hartree-Fock/6-31G**/Natural Population Analysis method and those calculated using conjugation charge equations.
2.4. Calculation of Conjugation Operator Weights

The weights of the conjugation operators were calculated so as to minimize the square error between \textit{ab initio} charges $q_j^{ai}$ and those calculated by charge conjugation equations, $q_j^{ccc}$ referring to the same atom in the same molecule.

$$\{w_i^*\} = \min_{w_i} \sum_j \left( q_j^{ai} - q_j^{ccc}(w_i) \right)^2$$ (2-60)

Atomic charges were calculated for common solvents as well as alkanes molecules using the Hartree-Fock method with the 6-31G** basis function using the Gaussian 94 software [49]. Atomic charges were calculated using the Natural Population analysis method. Prior to the atomic charge calculation, the molecular geometry was optimized in the Hartree-Fock calculation. The initial guess for the molecular geometry was obtained from the molecular geometries in the Insight II software. The molecules used in the regression of the weights of conjugation operators and their extrapolation are shown in Table 2-2.

<table>
<thead>
<tr>
<th></th>
<th>Regression</th>
<th>Extrapolation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inorganic</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Alcohols</strong></td>
<td>methanol</td>
<td>1-propanol</td>
</tr>
<tr>
<td></td>
<td>ethanol</td>
<td>2-butanol</td>
</tr>
<tr>
<td></td>
<td>1,2-ethanediol</td>
<td>2-pentanol</td>
</tr>
<tr>
<td></td>
<td>2-propanol</td>
<td>3-pentanol</td>
</tr>
<tr>
<td></td>
<td>2-methyl-2-propanol</td>
<td>2-methyl-2-butanol</td>
</tr>
<tr>
<td></td>
<td>1-butanol</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3-methyl-1-butanol</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-pentanol</td>
<td></td>
</tr>
</tbody>
</table>

61
<table>
<thead>
<tr>
<th><strong>Acids and Esthers</strong></th>
<th><strong>Nitriles</strong></th>
<th><strong>Aromatic</strong></th>
<th><strong>Halogenated</strong></th>
<th><strong>Ketones</strong></th>
<th><strong>Aldehydes</strong></th>
<th><strong>Amines</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>formic acid</td>
<td></td>
<td></td>
<td></td>
<td>acetone</td>
<td>formaldehyde</td>
<td>vinylamine</td>
</tr>
<tr>
<td>acetic acid</td>
<td></td>
<td></td>
<td></td>
<td>2-butanone</td>
<td>acetaldehyde</td>
<td>1,2-diaminoethane</td>
</tr>
<tr>
<td>diethyl carbonate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ethyl benzoate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>methyl acetate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>acetic anhydride</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ethyl acetate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>acetonitrile</td>
<td>propionitrile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>acetonitrile</td>
<td></td>
<td>benzene</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>propionitrile</td>
<td></td>
<td>toluene</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nitriles</td>
<td></td>
<td>1,2-dimethylbenzene</td>
<td>1,2-dichloroethane</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aromatic</td>
<td></td>
<td>1,4-dimethylbenzene</td>
<td>chloroform</td>
<td>acetone</td>
<td>formaldehyde</td>
<td></td>
</tr>
<tr>
<td>benzene</td>
<td></td>
<td>1,4-dimethylbenzene</td>
<td>dichloromethane</td>
<td></td>
<td>acetaldehyde</td>
<td></td>
</tr>
<tr>
<td>toluene</td>
<td></td>
<td>tetrachloromethane</td>
<td>1,1-dichloroethane</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,2-dimethylbenzene</td>
<td></td>
<td>1,1,1-trichloroethane</td>
<td>1,1,2,2-tetrachloroethane</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,4-dimethylbenzene</td>
<td></td>
<td>1,1,1,1-tetrachloroethane</td>
<td>1,1,1,1-tetrachloroethane</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Halogenated</td>
<td></td>
<td>1,1-dichloroethene</td>
<td>fluoromethane</td>
<td>acetone</td>
<td>formaldehyde</td>
<td></td>
</tr>
<tr>
<td>Halogenated</td>
<td></td>
<td>chloroform</td>
<td>difluoromethane</td>
<td></td>
<td>acetaldehyde</td>
<td></td>
</tr>
<tr>
<td>Halogenated</td>
<td></td>
<td>dichloromethane</td>
<td>tetrafluoromethane</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Halogenated</td>
<td></td>
<td>tetrachloromethane</td>
<td>chlorocyclohexane</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Halogenated</td>
<td></td>
<td>1,1-dichloroethene</td>
<td></td>
<td>acetone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ketones</td>
<td></td>
<td>chloroform</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>acetone</td>
<td></td>
<td>dichloromethane</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-butanone</td>
<td></td>
<td>tetrachloromethane</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aldehydes</td>
<td></td>
<td>1,1-dichloroethene</td>
<td></td>
<td>acetone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>formaldehyde</td>
<td></td>
<td>fluoromethane</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>acetaldehyde</td>
<td></td>
<td>difluoromethane</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,2-diaminoethane</td>
<td></td>
<td>tetrafluoromethane</td>
<td></td>
<td>acetone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amines</td>
<td></td>
<td>chlorocyclohexane</td>
<td></td>
<td>2-butanone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vinylamine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>formaldehyde</td>
<td></td>
</tr>
<tr>
<td>1,2-diaminoethane</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>acetaldehyde</td>
<td></td>
</tr>
<tr>
<td>Nitrogenated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nitromethane</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Amides</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Formamide</td>
<td>Dimethyl formamide</td>
</tr>
<tr>
<td>Methyl formamide</td>
<td>Dimethyl acetamide</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alkanes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethane</td>
<td>Pentane</td>
</tr>
<tr>
<td>Ethene</td>
<td>Hexane</td>
</tr>
<tr>
<td>Acetylene</td>
<td></td>
</tr>
<tr>
<td>Cyclohexane</td>
<td></td>
</tr>
<tr>
<td>Heptane</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ethers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Diethyl ether</td>
<td>1,4-dioxane</td>
</tr>
<tr>
<td>1,2-dimethoxyethane</td>
<td></td>
</tr>
<tr>
<td>Tetrahydrofuran</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Multifunctional</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1,1,1-trifluor-3-propanol</td>
<td>Chlorobenzene</td>
</tr>
<tr>
<td>Chlorobenzene</td>
<td>1,1,1-trifluor-3-propanol</td>
</tr>
<tr>
<td>Benzonitrile</td>
<td></td>
</tr>
<tr>
<td>Acetophenone</td>
<td></td>
</tr>
<tr>
<td>Nitrobenzene</td>
<td></td>
</tr>
<tr>
<td>Phenol</td>
<td></td>
</tr>
</tbody>
</table>

Table 2-2: Compounds used in the calculation of conjugation operator weights.

Figure 2-24 shows the fit of the atomic charges obtained by charge conjugation equations against those calculated by Natural Population Analysis.
Figure 2-24: Comparison between atomic charges calculated by Natural Population Analysis and those calculated by conjugation charge equations.

From Figure 2-24 it can be seen that the charges obtained by conjugation charge equations have an excellent fit against those calculated by \textit{ab initio} calculations. The average error is 0.0253 and the maximum error is 0.1320. The calculated weights of the conjugation operators are given in Table 2-3.
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>0.0311</td>
</tr>
<tr>
<td>CdC</td>
<td>0.2816</td>
</tr>
<tr>
<td>CtC</td>
<td>8.8354</td>
</tr>
<tr>
<td>CarC</td>
<td>-0.3777</td>
</tr>
<tr>
<td>CARC</td>
<td>69.8282</td>
</tr>
<tr>
<td>CH</td>
<td>-1.6347</td>
</tr>
<tr>
<td>HC</td>
<td>-1.4253</td>
</tr>
<tr>
<td>CO</td>
<td>0.0291</td>
</tr>
<tr>
<td>OC</td>
<td>-0.29531</td>
</tr>
<tr>
<td>CdO</td>
<td>5.0641</td>
</tr>
<tr>
<td>OdC</td>
<td>4.4124</td>
</tr>
<tr>
<td>CF</td>
<td>1.2158</td>
</tr>
<tr>
<td>FC</td>
<td>0.7842</td>
</tr>
<tr>
<td>CCl=</td>
<td>1.1324</td>
</tr>
<tr>
<td>CIC</td>
<td>0.6441</td>
</tr>
<tr>
<td>CN</td>
<td>2.1997</td>
</tr>
<tr>
<td>NC</td>
<td>2.0184</td>
</tr>
<tr>
<td>CtN</td>
<td>31.2954</td>
</tr>
<tr>
<td>NtC</td>
<td>27.2949</td>
</tr>
<tr>
<td>NH</td>
<td>-1.9928</td>
</tr>
<tr>
<td>HN</td>
<td>-1.6407</td>
</tr>
<tr>
<td>NdO</td>
<td>2.1215</td>
</tr>
<tr>
<td>OdN</td>
<td>1.7940</td>
</tr>
<tr>
<td>OH</td>
<td>-0.7942</td>
</tr>
<tr>
<td>HO</td>
<td>-0.3754</td>
</tr>
<tr>
<td>OdnC</td>
<td>0.0479</td>
</tr>
<tr>
<td>OddnC</td>
<td>-0.0463</td>
</tr>
<tr>
<td>OddnN</td>
<td>0.0077</td>
</tr>
<tr>
<td>NddnO</td>
<td>1.0000</td>
</tr>
<tr>
<td>NdnC</td>
<td>-0.0362</td>
</tr>
<tr>
<td>NtdnC</td>
<td>3.6787</td>
</tr>
</tbody>
</table>

Table 2-3: Weights of the conjugation operators.

The charges calculated through conjugation charge equations were used to model the solvent in the reaction fingerprint model. The next chapter describes the model and how the solvent atomic charges were incorporated in it.
3. The Reaction Fingerprint

Solvent effects on reaction rates are mostly due to differences in electrostatic interaction between the solvent/reactants and solvent/transition state pairs. This differential solvation, which affects the energy of activation of the reaction and consequently its rate, depends on the electrostatic characteristics of the solvent and the change in charge distribution that happens to the reactive species when they evolve from reactant to transition state. In Chapter 2, conjugation operators were used to calculate the atomic charges of the solvent providing a picture of its electrostatic makeup. It is now necessary to describe the charge distribution evolution during the activation step of the reaction. Ideally, this could be done by calculating the atomic charges for both the reactants and the transition state. However, it is very difficult to obtain transition state geometries and electronic distributions for the transition state except for the simplest reactions. On the other hand, one can exploit the fact that, for the purpose of calculating reaction rates, one is interested only in the difference between the solvation energies between reactants and transition state rather than their individual values. This leads to a model where the reactants and the transition state are aggregated in a single hypothetical molecular entity, called the reaction fingerprint, whose energy of solvation is equal to the difference in solvation energies between transition state and reactants. Since the reaction fingerprint is a molecular entity, its energy of solvation is calculated within the context of statistical thermodynamics.

3.1. Transition State Basis of the Reaction Fingerprint

In the context of Transition State theory the rate of a reaction is given by

\[ k = \frac{RT}{N_A h} e^{\frac{-\Delta G^\ddagger}{RT}} \]  

(3-1)
where \( N_A \) is the Avogadro number, \( h \) is the Planck constant, \( R \) is the ideal gas constant, \( T \) is the temperature and \( \Delta G^\# \) is the free energy of activation of the reaction. The free energy term appears in the expression for the rate constant as a consequence of the assumption of equilibrium between transition state and reactants.

In a reaction happening in a solution, the free energy of activation is given by

\[
\Delta G_{\text{sol}}^\# = (\Delta G_0^{TS} + \Delta G_{sol}^{TS}) - (\Delta G_0^R + \Delta G_{sol}^R) \tag{3-2}
\]

where the subscripts \( TS \) and \( R \) refer to transition state and reactants respectively and the superscripts \( 0 \) and \( \text{sol} \) refer to the free energy of formation in the reference state (gas phase) and in solution, respectively. The two first terms in Eq. 3-2 refer to intrinsic characteristics of the reaction which depend only on the molecular structures of transition state and reactants whereas the two last terms account for the effect of the interaction of the environment with the reacting species.

The ratio of the rates of the same reaction in a solvent \( A \) and in gas phase is, according to Eq. 3-1.

\[
\frac{k_A}{k_0} = e^{-\frac{(-\Delta G_A^\# + \Delta G_0^\#)}{RT}} \tag{3-3}
\]

Substituting Eq. 3.2 into Eq. 3.3:

\[
\frac{k_A}{k_0} = e^{-\left[ \frac{(\Delta G_0^{TS} + \Delta G_A^{TS}) - (\Delta G_0^R + \Delta G_A^R) - (\Delta G_0^{TS} - \Delta G_0^R)}{RT} \right]} \quad \left[ \frac{\Delta G_A^{TS} - \Delta G_A^R}{RT} \right] = e^{-\frac{(\Delta G_A^{TS} - \Delta G_A^R)}{RT}} \tag{3-4}
\]

or in other terms:
\[ \ln k_A = \ln k_0 - \left( \frac{\Delta G_A^{TS} - \Delta G_A^R}{RT} \right) = \ln k_0 - \frac{\Delta G_A^\neq}{RT} \] (3-5)

According to Eq. 3-5, the reaction rate in solution can be expressed as a contribution of a reference term \((k_0)\), which depends only on the molecular structures of the transition state and the reactants, and a solvation term \((\Delta G_A^\neq)\), which describes the interactions of these molecular structures with the solvent as the difference in free energies of solvation between transition state and reactants.

The main difficulty related to calculating the reaction rate expressed in Eq. 3-5 is the fact that obtaining the molecular structure of the transition state is impractical for but the simplest cases [44]. This is because to incorporate the effect of partially formed (or broken) bonds it is necessary to use large electronic orbital basis sets and to take into account electronic correlation effects. Moreover, the transition state is located at a saddle point in the molecular structure energy surface and current saddle point location methods are not yet effectively applied to the problem of determining transition state structures.

However, to calculate the rate in solution of a reaction it is not necessary to know the free energies of solvation of the individual components, but the *difference* between the solvation energies of the transition state and reactants. The question then is how to calculate this difference and how to apply it to the calculation of rates in multiple solvents.

The problem of determining the transition state structure is avoided by defining a hypothetical, but physically meaningful molecular entity that will be henceforth called the *reaction fingerprint*. Its main property is that its free energy of solvation in a solvent is equal to the free energy of the reaction it corresponds to in the same solvent.

\[ \Delta G_A^{RF} = \Delta G_A^{TS} - \Delta G_A^R \] (3-6)
where $\Delta G^{RF}_A$ is the free energy of solvation of the reaction fingerprint in a solvent $A$.

The reaction fingerprint is a molecular entity, consisting of physical parameters. In its simplest form, the reaction fingerprint is a sphere containing an electric dipole in its center. It is defined by three parameters: dipole charge, dipole length and molecular radius, as shown in Figure 3-1.

Since the reaction fingerprint is a molecular entity, its free energy of solvation is calculated at the molecular level using a statistical mechanical expression. This is done by computing the distribution of the solvent molecules around the reaction fingerprint and then calculating its interaction with the solvation shells. This has the practical effect of decoupling the problem of calculating solution reaction rates in two distinct elements: the reaction fingerprint, which represents the reaction itself and the solvent, which represents the environment where the reaction occurs. Because the reaction fingerprint incorporates characteristics of both reactants and transition state into a single molecular entity and does not include any description of the environment, it provides a parametrized description of the intrinsic aspects of the reaction where the charge evolution during the activation step is encoded in the reaction fingerprint’s parameters. The interaction between reaction fingerprint and solvent allows the computation of the free energy of activation in solution.

The molecular parameters in the reaction fingerprint not only provide a description of charge and polarization (or lack thereof) in the reaction, but also basic geometric information. While the dipole length and charge parameters express the evolution of charge distribution in the reaction, the radius parameter determines how available those charges are available to interaction with a solvent. The representation of reactions by a set of molecular parameters comprising energetic and geometric characteristics of the species involved is the main feature of the reaction fingerprint model.
Figure 3-1: The reaction fingerprint in its simplest form: a spherical molecular entity of radius $R$ consisting of a centrally located dipole with charge parameter $\delta$ and length parameter $L$.

More complicated forms for the reaction fingerprint are possible including asymmetric dipoles (not centered in the sphere) or an ellipsoidal reaction fingerprint. It is also possible to include other interaction terms such as van der Waals parameters or quadrupoles. However, more sophisticated representations of the reaction fingerprint also require more parameters and experimental data.

### 3.2. Statistical Mechanical Calculation of Free Energy of Solvation

Since the reaction fingerprint is a molecular entity, its free energy of solvation must be calculated at the molecular level. To do this one must use the tools of statistical mechanics and thermodynamics [50-52].
In the liquid phase, the relative positions of molecules are correlated by the energy of interaction among them. The structure of the molecules in a liquid can be modeled by fixing an arbitrary coordinate center on one molecule and counting the number of molecules around it as a function of the radial distance from the center. The function that maps the number, or rather the probability of the distribution of molecules around the coordinate center is called the radial distribution function. Its calculation is described in the next section.

3.2.1. Radial Distribution Functions

In a liquid, the probability of a certain configuration among the molecules is a function of the total potential energy of interaction among them. The lower the potential energy of a molecular arrangement, the higher its statistical weight. According to statistical thermodynamics, the probabilistic weight of an ensemble of $N$ molecules is given by the Boltzmann distribution:

$$w(r_1, r_2, ..., r_N) = e^{-U_N(r_1, r_2, ..., r_N)} kT$$

(3-7)
where \( r_1, r_2, \ldots, r_N \) are the coordinates of molecules 1, 2, \ldots, \( N \), \( k \) is the Boltzmann constant, \( T \) is the temperature and \( U_N \) is the total potential energy of interaction among the \( N \) molecules. \( w \) is the statistical weight of the ensemble when molecule 1 is in \( r_1 \), molecule 2 in \( r_2 \) and so on. In molecular simulations it is useful to express the total potential energy as a sum of pair-wise potentials.

\[
U_N(r_1, r_2, \ldots, r_N) = \sum_{i=1, j > i}^{N} u_{ij}(r_i, r_j) \tag{3-8}
\]

where \( u_{ij} \) is the potential energy of interaction between two molecules \( i \) and \( j \), which depends only on the distance between the two molecules and their relative orientations if they are not spherical.

Using the Boltzmann distribution, the probability of a fixed configuration where molecule 1 is at position \( dr_1 \), molecule 2 at \( dr_2 \) and so on in an ensemble containing \( N \) molecules in a volume \( V \) at temperature \( T \) is given by:

\[
P(r_1, r_2, \ldots, r_N) = \frac{e^{-U_N(r_1, r_2, \ldots, r_N) / kT}}{\sum_{r_1, r_2, \ldots, r_N} e^{-U_N(r_1, r_2, \ldots, r_N) / kT}} \tag{3-9}
\]

The denominator of the expression is a normalization factor, the configuration integral, which is the summation of the statistical weights for every possible configuration of the system, i.e. by varying the coordinates of each molecule to span the whole volume of the ensemble.

If \( k \) molecules have their coordinates fixed irrespective of the coordinates of the other \((N-k)\) molecules, the probability of that configuration is
\[ p^{(k)}(r_1, r_2, ..., r_N) = \frac{\int \int ... \int e^{-U_N(r_1, r_2, ..., r_N)} kT^{dr_{k+1}...dr_N}}{\int \int ... \int e^{-U_N(r_1, r_2, ..., r_N)} kT^{dr_1dr_2...dr_N}} \] (3-10)

where the positions of the \( N-k \) molecules which are not fixed are integrated over space.

Eq. 3-10 only considers one specific set of molecules in fixed positions. Since molecules are indistinct, all possible combinations and permutations of all \( N \) molecules must be considered. To take that into account, the number of possible ways to make groups of \( k \) molecules from a set of \( N \) molecules is

\[ C(N,k) = \frac{N!}{(N-k)!k!} \] (3-11)

The \( k! \) multiple permutations of the fixed molecules over the fixed coordinates also have to be taken into account and so the total number of arrangements is \( C(N,k).k! \). Therefore, the probability that a set of any \( k \) molecules is at positions \( r_1, r_2...r_k \) is:

\[ \rho^{(k)}(r_1, r_2, ..., r_N) = \frac{N!}{(N-k)!k!} \frac{\int \int ... \int e^{-U_N(r_1, r_2, ..., r_N)} kT^{dr_{k+1}...dr_N}}{\int \int ... \int e^{-U_N(r_1, r_2, ..., r_N)} kT^{dr_1dr_2...dr_N}} \] (3-12)

or \( \rho^{(k)}(r_1, r_2, ..., r_N) = \frac{N!}{(N-k)!} P^{(k)}(r_1, r_2, ..., r_k) \)

When \( k=1 \), \( \rho^{(1)} \) is simply the probability that any one molecule be found at position \( dr_1 \).

In an ideal gas where there are no interactions between particles (i.e. \( U_N = 0 \)), \( \rho^{(1)} \) becomes
\[ \rho^{(1)}(r_1, r_2, \ldots, r_N) = \frac{N!}{(N-1)!} \frac{\int \cdots \int dr_2 \cdots dr_N}{V^N} = N \frac{V^{N-1}}{V^N} = \frac{N}{V} \]  

which is an intuitive result, as in the absence of correlation between molecules, any region of the volume has the same probability that a molecule can be found there and the probability is \( N/V \), the number density. By extension, the probability of finding \( k \) molecules of an ideal gas in \( k \) distinct positions is \( (N/V)^k \).

Since the uncorrelated probability for finding a molecule is only a function of a power of the number density, it is useful to model this probability for a liquid, where correlation is present, as a product of the uncorrelated probability and a factor to introduce correlation. This factor is the radial distribution function, denoted by \( g^{(k)} \).

\[ \rho^{(k)}(r_1, r_2, \ldots, r_k) = \rho^k g^{(k)}(r_1, r_2, \ldots, r_k) \]  

From Eq. 3-13, the expression for the radial distribution function is:

\[ g^{(k)}(r_1, r_2, \ldots, r_k) = \frac{V^k N!}{N^k (N-k)!} \frac{\int \cdots \int e^{-U_N} dr_{k+1} \cdots dr_N}{\int \cdots \int e^{-U_N} dr_1 \cdots dr_N} \]  

Since the intermolecular potential energy is usually calculated using pair-wise potentials, it is useful to calculate the distributions of pairs of molecules. When \( k = 2 \) the distribution function describes the correlation between pairs of molecules in an ensemble. For spherical molecules and pair-wise potentials, the potential energy, and by extension the
radial distribution function are functions of the distance between the pairs of molecules, not their absolute coordinates or

\[ g^{(2)}(r_1, r_2) = g^{(2)}(r_{12}) = g(r) \] (3-16)

For simplicity, the superscript of \( g \) will be dropped from now on and \( r \) will be the distance between two molecules. Integrating \( \rho g(r) \) over space:

\[
\int \rho g(r) dr_2 = \int \frac{N}{V} \frac{V^2 N!}{V^2 (N-2)!} \int \frac{-U_N}{kT} dr_3 ... dr_N \\
= \frac{N}{V} \frac{V^2 N!}{V^2 (N-2)!} \frac{1}{V} = N - 1
\] (3-17)

The quotient between the two integrals reduces to \( 1/V \) because \( dr_I \) can be seen as the position of the “coordinate center molecule”, which is immaterial to the potential function, and therefore can be brought out of the \( N \)-dimensional integral in the denominator. Therefore, given that a molecule is at an arbitrary location, \( \rho g(r)d(r) \) is the number of molecules at a distance \( dr \) from it.
Figure 3-3: Representation of a typical radial distribution function

Figure 3-3 shows a representation of the shape of a typical radial distribution function. Since molecules cannot be superposed, $g(r)$ is equal to zero in the region within the solute’s radius, achieving a sharp peak in its immediate vicinity indicating the first layer, or solvation shell, of solvent molecules. As the distance from the central molecule increases, smaller peaks appear, indicating the second and third solvation shells, and at sufficiently large distance, $g(r)$ tends to one and the number of molecules at that distance becomes $\rho dr$, meaning that the coordinates of the solute and those of molecules far removed from it are uncorrelated.

Once the radial distribution function has been defined, the statistical mechanical expression for the free energy of solvation can be determined.

### 3.2.2. Free Energy of Solvation

It is a well known result in statistical mechanics [50-52] that the Helmholtz energy $A$ of an ensemble of molecules is given by:
\[ A = -kT \ln Z_N - kT \ln N! - 3NkT \ln \Lambda \]

where \( Z_N = \int \ldots \int e^{-\frac{U_N}{kT}} \, dr_1 \ldots dr_N \), the configuration integral

and \( \Lambda = \sqrt{\frac{h^2}{2\pi mkT}} \)

where \( h \) is Planck’s constant and \( \Lambda \) is the de Broglie wavelength of the molecule, which depends on the molecule’s mass \( m \).

The free energy of solvation, or more generally, the chemical potential is thermodynamically related to \( A \) by:

\[ \mu = \left( \frac{\partial A}{\partial N} \right)_{V,T} \]  

(3-19)

and since \( N \) is very large, one can write:

\[ \left( \frac{\partial A}{\partial N} \right)_{V,T} = A(N, V, T) - A(N - 1, V, T) \]  

(3-20)

Substituting Eqs. 3-18 and 3-19 into Eq. 3-20:

\[ \mu = -kT \ln \left( \frac{Z_N}{Z_{N-1}} \right) + kT \ln N + kT \ln \Lambda^3 \]  

(3-21)

At this point, a mathematical artifice, called the coupling parameter \( \xi \) is introduced. \( \xi \) describes the existence of the solute particle in solution. When \( \xi = 0 \), the solute is not
present in the solution and when \( \xi = 1 \), the solute is present and fully interacting with the solution. Changing \( \xi \) from 0 to 1 means to gradually bring the solute molecule into existence and allows the integration from one state to another (from no solute to solute in solution) to be performed as a line integral to obtain the chemical potential.

The coupling parameter is applied as a factor to the potential energy of interaction between solvent and solute. If the index 1 is assigned to the solute molecule, the potential energy of interaction \( U_N \) can be expressed as a function of the coupling parameter \( \xi \):

\[
U_N(\xi) = \xi \sum_{j=2}^{N} u_{1j} + \sum_{i=2}^{N} \sum_{j=i+1}^{N} u_{ij}
\]

(3-22)

where the first summation term refers to the interaction between the solute (subscript 1) and the solvent molecule pairs \((u_{ij})\) and the second term the interactions between solvent molecule pairs \((u_{ij})\). When \( \xi = 0 \), the first term vanishes, and the solute molecule is not present. When \( \xi = 1 \), the solute is interacting fully with the solvent molecules.

The configuration integrals \( Z_N \) (solute present) and \( Z_{N-1} \) (solute not present) can also be expressed as functions of \( \xi \) and their ratio in Eq. 3-21 calculated.

\[
Z_N(\xi = 0) = \int \cdots \int e^{\frac{kT}{V}} dr_1 dr_2 \cdots dr_N = \\
\int \cdots \int e^{\frac{kT}{V}} dr_1 dr_2 \cdots dr_N = \int dr_1 Z_{N-1} = VZ_{N-1}
\]

(3-23)
As a result

\[ \frac{Z_N}{Z_{N-1}} = \frac{Z_N(\xi = 1)}{Z_N(\xi = 0)} \]  \hspace{1cm} (3-24)

and the first term in Eq. 3-21 can be rewritten as:

\[ \ln \frac{Z_N}{Z_{N-1}} = \ln Z_N(\xi = 1) - \ln Z_N(\xi = 0) + \ln V = \]

\[ = \int_0^1 \frac{1}{\partial \xi} \ln Z_N d\xi + \ln V \]  \hspace{1cm} (3-25)

It is now necessary to evaluate the partial derivative of \( Z_N \) with respect to \( \xi \) as it appears in Eq. 3-25.

\[ \frac{\partial Z_N}{\partial \xi} = \int \int \int \cdots \int \frac{1}{kT} \left( \sum_{j=2}^{N} \sum_{i=2}^{N} u_{1j} \right) e^{-\frac{-\xi \sum_{j=2}^{N} u_{1j} - \sum_{i=2}^{N} \sum_{j=i+1}^{N} u_{ij}}{kT}} dr_1 dr_2 \cdots dr_N \]  \hspace{1cm} (3-26)

The above integral consists of \( N-1 \) identical terms since the molecules are indistinguishable.
\[ \frac{\partial Z_N}{\partial \xi} = - \frac{(N-1)}{kT} \int \int \ldots \int u_{1j} e^{ \frac{-U_N(\xi)}{kT} } \, \frac{d\mathbf{r}_1}{kT} d\mathbf{r}_2 \ldots d\mathbf{r}_N = \]

\[ = - \frac{(N-1)}{kT} \int \int \ldots \int u_{1j} e^{ \frac{-U_N(\xi)}{kT} } d\mathbf{r}_1 d\mathbf{r}_2 \ldots d\mathbf{r}_N. \quad (3-27) \]

Dividing Eq. 3-27 by \( Z_N \):

\[ \frac{1}{Z_N} \frac{\partial Z_N}{\partial \xi} = \frac{\partial \ln Z_N}{\partial \xi} = \]

\[ = - \frac{(N-1)}{kT} \int \int \ldots \int u_{1j} e^{ \frac{-U_N(\xi)}{kT} } \, \frac{d\mathbf{r}_1}{kT} d\mathbf{r}_2 \ldots d\mathbf{r}_j d\mathbf{r}_{j-1} d\mathbf{r}_{j+1} \ldots d\mathbf{r}_N \quad (3-28) \]

and substituting Eq. 3-15 to bring the radial distribution function into Eq. 3-28:

\[ \frac{\partial \ln Z_N}{\partial \xi} = - \frac{(N-1)}{kT} \int \int \frac{\rho^2}{N(N-1)} g(\mathbf{r}_{ij}) d\mathbf{r}_1 d\mathbf{r}_j. \quad (3-29) \]

Since both \( u_{ij} \) and \( g(r_{ij}) \) depend only on the distance between molecules \( i \) and \( j \) irrespective of their absolute coordinates, Eq. 3-28 can be simplified by fixing the coordinate center on molecule \( i \) and defining the variable

\[ \mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i \]
and, since \( r_1 \) is fixed,

\[
dr_{1j} = dr_j
\]  

(3-30)

Substituting into Eq. 3-29:

\[
\frac{\partial \ln Z_N}{\partial \xi} = -\frac{(N-1)}{kT} \frac{\rho^2}{V} \int dr_1 \int u_{1j} \frac{\rho^2}{N(N-1)} g(\eta_{1j})d\eta_{1j} =
\]

\[
= -\frac{(N-1)}{kT} V \int u_{1j} \frac{\rho^2}{V} g(\eta_{1j})d\eta_{1j} = -\frac{1}{kT} \frac{V}{N} \rho^2 \int u_{1j} g(\eta_{1j})d\eta_{1j} = (3-31)
\]

\[
= -\frac{1}{kT} \rho \frac{V}{V} u_{1j} g(\eta_{1j})d\eta_{1j} = -\frac{1}{kT} \rho \frac{V}{V} u(\eta_{1j}) g(\eta_{1j})d\eta_{1j}
\]

Since the potential energy functions are assumed to depend only on the intermolecular distance and not on the angles between them, the volume element \( dr \) can be written in spherical coordinates.

\[
\frac{\partial \ln Z_N}{\partial \xi} = -\frac{1}{kT} \rho \frac{V}{V} \int u(\eta_{1j}) g(\eta_{1j}) d\eta_{1j} =
\]

(3-32)

\[
= -\frac{1}{kT} \rho \frac{\pi^2}{2} \frac{R}{\rho} \int u(\eta_{1j}) g(\eta_{1j}) r^2 \sin \theta dr d\theta d\phi
\]

Integrating over the symmetrical angular coordinates:

\[
\frac{\partial \ln Z_N}{\partial \xi} = -\frac{\rho}{kT} \frac{R}{\rho} \int u(r) g(r) 4\pi r^2 dr
\]  

(3-33)
And finally, substituting into Eq. 3-21 and 3-25, the expression for the chemical potential is obtained:

$$\mu = -kT \ln \left( \frac{Z_N}{Z_{N-1}} \right) + kT \ln N + kT \ln \Lambda^3 =$$

$$= -kT \left[ \int_0^1 \frac{\partial \ln Z_N}{\partial \xi} d\xi + \ln V \right] + kT \ln N + kT \ln \Lambda^3 =$$

$$= -kT \left[ \int_0^R \frac{\rho}{kT} \int_0^1 u(r) g(r) 4\pi r^2 dr d\xi - \ln V + \ln N + \ln \Lambda^3 \right]$$

$$\mu = \rho \int_0^R \int_0^1 u(r) g(r, \xi) 4\pi r^2 dr d\xi + kT \ln \rho \Lambda^3$$

The calculation of the chemical potential through Eq. 3-34 includes two terms: the last is the ideal gas contribution to the chemical potential, when there are no intermolecular interactions. The first term includes the interaction between the new molecule and its surroundings (embodied in the potential function) and the energy necessary to create a cavity in the ensemble to accommodate the new molecule (included in \(g(r, \xi)\) which describes how the ensemble molecules distribute themselves around the new molecule). The portion of the chemical potential that is due to intermolecular interactions is then calculated by multiplying the energy of interaction \(u(r)\) by the number of molecules at a distance \(r\) from the new molecule (given by \(\rho g(r) 4\pi r^2 dr\)) over the whole space as the new molecule gradually comes into existence (tantamount to integrating over the coupling parameter \(\xi\) from 0 to 1).

With an expression for the calculation of the chemical potential at the molecular level, the reaction fingerprint can be described from a statistical mechanical point of view.
3.2.3. Statistical Mechanical Description of the Reaction Fingerprint

In the problem of determining reaction rates in solution, the chemical potential is the free energy of solvation, the molecular ensemble is composed by solvent molecules and the reactants and transition state are the solute. The rate of a reaction in solution can be expressed in statistical mechanical terms by

$$\ln k_A = \ln k_0 - \frac{\Delta G^\#}{RT} = \ln k_0 - \frac{\mu_A^{TS} - \mu_A^R}{RT} = \ln k_0 - \frac{RT}{k_B T} \left[ \rho \int_0^1 \int_0^1 \left[ u_A^{TS}(r) g_A^{TS}(r, \xi) - u_A^R(r) g_A^R(r, \xi) \right] 4 \pi r^2 dr d\xi \right]$$

(3-35)

where the superscripts $TS$ and $R$ refer respectively to the transition state and the reactants and the subscript $A$ denotes that the chemical potential of species in a solvent $A$ is being calculated. Since each species interacts in its own way with the solvent, and the distribution of solvent molecules changes with each different solute, the radial distribution function (as well as the potentials) is different for each species.

The transition state is a unique species. The reactants, however, can be either one or two molecules (reactions involving collisions among three molecules are impractical in solution). The evolution of charge distribution depends on both reactants and transition state and the reaction fingerprint is a single molecular species which substitutes them in order to provide a compact description of the reaction. In statistical mechanical terms, the reaction fingerprint (denoted by the superscript $RF$) substitutes all the species involved in the reaction providing a parametrized description of the reaction.
\[ \int_{0}^{R} \int_{0}^{R} u_A^{RF}(r) g_A^{RF}(r, \xi) 4\pi r^2 dr d\xi = \]
\[ = \int_{0}^{R} \int_{0}^{R} [u_A^{TS}(r) g_A^{TS}(r, \xi) - u_A^{R}(r) g_A^{R}(r, \xi)] 4\pi r^2 dr d\xi \]

(3-36)

and as a result, the expression for the calculation of reaction rates using the reaction fingerprint is

\[ \ln k_A = \ln k_0 - \frac{1}{RT} \int_{0}^{R} u_A^{RF}(r) g_A^{RF}(r, \xi) 4\pi r^2 dr d\xi \]

(3-37)

It has been previously shown that the reaction fingerprint is a molecular entity with parameters that describe its polarity and the geometrical availability of its charges. These parameters affect Eq. 3-37 as both the potential energy and the radial distribution function terms depend on them. The dipole length and charge parameters determine the polarity of the reaction fingerprint and thus the potential electrostatic energy of interaction between the reaction fingerprint and the solvent, \( U_A^{RF} \). The radius parameter controls how close solvent molecules can be to the reaction fingerprint and so it affects the way the solvent molecules distribute themselves around the reaction fingerprint, i.e. the radial distribution function, \( g_A^{RF} \).

The reaction rate can now be calculated by a single mechanical statistical expression involving a single molecular species, the reaction fingerprint, whose characteristics are adjustable parameters. However, it still not clear how the radial distribution will depend on the parameters of the reaction fingerprint. The next section addresses this issue.
3.2.4. The Radial Distribution Function as a Function of the Reaction Fingerprint

It has been show previously that the pair-wise radial distribution function describing the distribution of solvent molecules around a solute where \( r \) is the distance between a solvent molecule and the solute is:

\[
g(r) = \frac{N(N-1)}{\rho^2} \left( \frac{-U_N}{\int V e^{\frac{-U_N}{kT}} dr_3...dr_N} - \frac{U_N}{\int V e^{\frac{-U_N}{kT}} dr_1...dr_N} \right)
\]

(3-38)

However, Eq. 3-38 is not practical for the calculation of the pair correlation function because it involves \( N \)-dimensional integrals, where \( N \) is very large and the integrand \( U_N \) depends on all the integrating variables except \( r_i \). Since \( \rho g(r)dr \) is the number of solvent molecules in a spherical crust of thickness \( dr \) at a distance \( r \) of the solute, \( g(r) \) can be computed through molecular dynamic simulations [53], which generally consist of the following steps:

1. Initialize an ensemble, with the solute fixed at the coordinate center, of molecules at random positions, velocities and orientations.
2. Using molecular forcefield parameters (charges, van der Waals potentials, etc.) compute the potential energy of the ensemble molecules.
3. Compute the direction, velocity and intensity of the forces acting on the ensemble molecules by taking the gradient of the potential energy.
4. Update incrementally the coordinates and orientation of the ensemble molecules according to the forces acting on each molecule.
5. Compute the macroscopic thermodynamic properties of the ensemble by statistical mechanical expressions.
6. Repeat steps 2-5. Continue until the thermodynamic properties of the ensemble reach a constant value (within a specified margin), i.e., the system has attained equilibrium.

Once the equilibrium has been reached, the radial distribution $g(r)$ can be computed simply by creating a histogram of the number of molecules (divided by the number density $\rho = N/V$) at various distances from the solute molecule.

Since the aim of the reaction fingerprint model is to enable the screening of a large number of solvents, it is impractical to determine the radial distribution function through molecular dynamics simulations as they are quite computationally expensive. It is necessary then to obtain an approximation of the radial distribution function.

The radial distribution function can be expanded as a power series of the number density $\rho$.

$$g(r) = a_0 + a_1 \rho + a_2 \rho^2 + \ldots$$  \hspace{1cm} (3-39)

By expanding $g(r)$ and truncating the series, it is possible to obtain a simple expression for the radial distribution function. The details of the calculation of the coefficients of the expansion are very involved and a well known result \cite{50} and thus will not be presented. It suffices to say that the reaction fingerprint model uses only the first term for simplicity. The approximate radial distribution function is then

$$g(r) = e^{\frac{-u(r)}{kT}}$$  \hspace{1cm} (3-40)

where $u(r)$ is the potential energy of interaction between the solute and one solvent molecule. This is much simpler than considering the whole molecular ensemble although it incurs a significant error, which will be justified later in this chapter. This approximate radial distribution function depends on the potential energy of interaction between the reaction fingerprint and the solvent, and consequently on the reaction fingerprint's
parameters. Now that a simple, closed-form expression has been obtained for the radial distribution function, it can be expressed as a function of the coupling parameter

\[
g(r, \xi) = e^{\frac{-\xi u(r)}{kT}}
\]

(3-41)

and applied to Eq. 3-38 so that an expression for reaction rates based on the reaction fingerprint can be obtained.

\[
\ln k_A = \ln k_0 - \frac{1}{RT} \rho \int_0^1 \int_0^R u_A^{RF}(r) e^{\frac{-\xi u_A^{RF}(r)}{kT}} 4\pi r^2 dr d\xi
\]

(3-42)

Integrating over the coupling parameter \( \xi \)

\[
\ln k_A = \ln k_0 - \frac{1}{RT} \rho \int_0^1 u_A^{RF}(r) \left( \frac{-kT}{u_A^{RF}(r)} - 1 \right) 4\pi r^2 dr
\]

(3-43)

which is the final expression for the reaction rate in the reaction fingerprint model.

**3.3. Implementation of the Reaction Fingerprint Model**

In using Eq. 3-44 for the calculation of the free energy of solvation of the reaction fingerprint and the corresponding reaction rate, one must take into account that both the solvent and the reaction fingerprint are in most cases not spherically symmetrical and
have significant dimensions. Moreover, considerations such as speed have to be taken into account so that the model is effective for screening large numbers of candidate molecules. For these reasons, the reaction fingerprint model was implemented in a number of different ways in order to assess their impact on its robustness and accuracy. This section describes the different implementations and Chapter 4 discusses their performance in a number of cases studies.

### 3.3.1. Potential Energy

The first implementation issue is the calculation of the potential energy of interaction. Since electrostatic interactions are the most important solvent effect on reaction rates, the potential energy of interaction includes only charge-charge interactions for simplicity. It is possible to include other terms such as van der Waals potentials and quadrupoles, but they require more adjustable parameters and, in the interest of simplicity and a compact description of the reaction fingerprint, were not included in this work. Chapter 5 includes a discussion on the implementation of van der Waals potentials in the reaction fingerprint model. The electrostatic potential energy between two point charges $q_i$ and $q_j$ separated by a distance $r_{ij}$ is given by

$$U(r) = \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ij}}$$  \hspace{1cm} (3-44)

where $\varepsilon_0$ is the vacuum electrical permittivity ($\varepsilon_0 = 8.854 \times 10^{-12}$ C²N⁻¹m⁻²). Since the reaction fingerprint in its simplest form has two electric charges of same magnitude and opposite signs, the total potential electrostatic energy of interaction between reaction fingerprint and solvent can be written as
where $q$ is the reaction fingerprint's charge magnitude parameter, $q_i$ are the solvent atomic charges and $r_{i+}$ and $r_{i-}$ are the distances between a solvent atomic charge and the reaction fingerprint’s positive and negative charges respectively.

In the calculation of the potential energy, solvent charges come into play. Solvent charges are calculated using conjugation operators as described in Chapter 2. However, one might be interested in using alternative representations of the solvent charges in order to keep the computational cost down. The next section discusses two different implementations of solvent charges and geometry.

### 3.3.2. Solvent Modeling

It is necessary to describe the solvent’s geometry and charges in order to obtain the charges and distances. The fullest solvent description for electrostatic energy calculation purposes is the set of molecular orbitals describing the electronic density over space; each molecular orbital is centered on the atomic center of its corresponding atom, so it is also necessary to obtain the coordinates of the atomic centers in the molecule. This approach involves the computation of spatial integrals for the calculation of the potential energy of interaction between each pair of orbitals and as such it has a very high computational cost. Solvents can also be described by their atomic charges, which are a discretization of the charge density distribution of the molecular orbitals; atomic charges are generally located at their corresponding atomic center coordinates and so their coordinates are also necessary. An even simpler representation is to use only the dipole moment of the solvent, by calculating the resulting positive and negative charges of the dipole moment and their coordinates. The dipole moment representation of the solvent will be henceforth called the *solvent dipole model* and the atomic charge representation will be called the *full atomic charge model*. 
The dipole moment $d$ of a solvent with atomic charges $q_i$ and atomic center coordinates $r_i$ is calculated by

$$d = \sum_i q_i r_i$$

(3-46)

and the positive and negative charge magnitudes $q^+$ and $q^-$ and respective coordinates $r^+$ and $r^-$ of the solvent’s dipole are given by

$$r^+ = \frac{\sum_{i,q_i>0} q_i r_i}{\sum_{i,q_i>0} q_i} \quad \text{and} \quad r^- = \frac{\sum_{i,q_i<0} q_i r_i}{\sum_{i,q_i<0} q_i}$$

(3-47)

and

$$q^+ = -q^- = \frac{|d|}{|r^+ - r^-|}$$

(3-48)

Figure 3-4 shows the molecular structure and atomic charges of ethanol as well as the resulting electric dipole and its coordinates.
Both the dipolar and the full charge solvent representations can be used when calculating the energy of interaction between the solvent and the reaction fingerprint. The solvent dipole model involves interactions between a smaller number of charges and it is therefore less computationally expensive, but it is also significantly less accurate than the full atomic charge molecule, as will be seen in Chapter 5.

### 3.3.3. Spherical Asymmetry

The expression for calculating a reaction rate in solution as a function of the reaction fingerprint’s parameters

\[
\ln k_A = \ln k_0 + \frac{1}{RT} \rho \int_0^1 kT \left[ -\frac{u_A^{RF}(r)}{kT} e^{-\frac{u_A^{RF}(r)}{kT}} - 1 \right] 4\pi r^2 dr
\]  

(3-49)
assumes that both the reaction fingerprint and the solvent molecule are spherically symmetrical. They are not and for this reason the potential energy of interaction $U_{A}^{RF}$ is not only a function of the distance between the species but also of their relative orientation. Therefore, at every step of the integration over the radial distance $r$, a value of $U_{A}^{RF}(r)$ has to be obtained by averaging over a number of possible relative orientations between the solvent and the reaction fingerprint separated by a distance $r$.

For simplicity, in the calculation of the integral in Eq. 3-49, the reaction fingerprint is positioned with its center over the coordinate center and its dipole aligned with the $x$ axis. The solvent molecule has then three translational degrees of freedom and three rotational (internal) degrees of freedom with respect to its orientation towards the reaction fingerprint. Figure 3-5 shows the arrangement of the reaction fingerprint and the six degrees for orientation of the solvent.

Since the calculation of the potential energy involves only two molecules at a time, for a radial position $r$ the potential energy $U_{A}^{RF}(r, \alpha, \beta, \gamma, \phi, \theta, \varphi)$ can obtained by Boltzmann averaging over a sample of relative orientations.

$$U_{A}^{RF}(r) = \frac{-U_{A}^{RF}(r, \alpha_{i}, \beta_{j}, \gamma_{k}, \phi_{m}, \theta_{n}, \varphi_{p})}{kT} \sum_{i} \sum_{j} \sum_{k} \sum_{m} \sum_{n} \sum_{p} e$$

$$U_{A}^{RF}(r, \alpha_{i}, \beta_{j}, \gamma_{k}, \phi_{m}, \theta_{n}, \varphi_{p})$$

$$= \frac{U_{A}^{RF}(r, \alpha_{i}, \beta_{j}, \gamma_{k}, \phi_{m}, \theta_{n}, \varphi_{p})}{kT} \sum_{i} \sum_{j} \sum_{k} \sum_{m} \sum_{n} \sum_{p} e$$

$$\sum_{i} \sum_{j} \sum_{k} \sum_{m} \sum_{n} \sum_{p} e$$

(3-50)

This approach was implemented in the reaction fingerprint model and henceforth it will be referred to as the orientation sampling approach.
Figure 3-5: The six degrees of freedom for the orientation of the solvent with respect to the reaction fingerprint.

The six orientation degrees of freedom require that a very large number of orientations be sampled so that a statistically meaningful value of the potential energy may be obtained. For instance, if only 3 values for each degree of freedom are sampled, 729 orientations will have to be calculated at every step in the integration, which is very computationally expensive. For this reason, an alternative approach was also applied.

Since the Boltzmann distribution's statistical weight is an exponential function of the potential energy, it is expected that low energy (i.e. large negative values of $U_{ARF}(r)$) will have very large statistical weights and determine the average value. For this reason, instead of sampling a large number of orientations, it is possible to use only the orientations with the lowest potential energy in the calculation. This approach henceforth will be referred to as the most favorable orientation.

There are two ways to align two dipoles so that their potential energies are at a minimum. They can be aligned along a line so that the negative charge of one dipole is closest to the positive charge of the other or vice-versa (longitudinal alignment) or they can be placed in parallel with their centers on a line perpendicular to the dipoles so that opposite sign
charges are closer than charges of the same sign (parallel alignment). These two alignments are shown in Figure 3-6.

![Figure 3-6: Lowest energy orientations for a pair of dipoles: (a) longitudinal orientation and (b) parallel orientation.](image)

The most favorable configuration approach has been applied to both the solvent dipole and the full atomic charge models. When the solvent is modeled as a single dipole, it is simply aligned according to Figure 3-6. When the full atomic charge is in use, the atomic charges are aligned so that the dipole moment of the molecule is aligned with the reaction fingerprint’s dipole, as shown for the structure of ethanol in Figure 3-4. Due to the asymmetry of the solvent, there are two different longitudinal orientations and infinite different parallel orientations. Only two parallel orientations are used for potential energy calculations in the model. The resulting potential energy at distance \( r \), \( U_{RF}^R(r) \) is obtained by Boltzmann averaging the energies of the four most favorable orientations as shown in Eq. (3-50).

There remains one detail to be considered when calculating the potential energy of interaction between the reaction fingerprint and the solvent: both species have actual dimensions and thus it is necessary to determine how close they can be. That is the reason why the reaction fingerprint has a radius parameter; it specifies how close the solvating charges can be to the reaction fingerprint, i.e. it provides a simple measure of the steric impediment to solvation.
Figure 3-7: Ethanol molecule aligned in the most favorable orientations in the full atomic charge model.

The solvent also needs a measure of the exposure of its charges. Therefore it is necessary to define a solvent boundary. In the reaction fingerprint model this boundary is the radius of the smallest sphere containing the whole molecule. For a molecule containing $A$ atoms with coordinates $r_i$ and radii $R$, the sphere is computed according to the following algorithm:

1. Start with the coordinate center $c$ of the molecule as an estimate of the center of

$$c = \frac{1}{A} \sum_{i=1}^{A} r_i$$

the sphere:
2. Compute the largest distance \( r_{\text{max}} \) from the coordinate center to the periphery of the molecule (atomic coordinate + radius): 
\[
|r_{\text{max}}| = \max(|r_i + R_i - c|)
\]

3. Update the center of the sphere \( c \) by moving it a fraction of the module of \( r_{\text{max}} \) along its direction: 
\[
c_{\text{new}} = c + \rho r_{\text{max}}, \quad 0 < \rho < 1
\]

4. Compute the new largest distance \( r_{\text{max,\ new}} \) using the updated center \( c_{\text{new}} \):
\[
|r_{\text{max,\ new}}| = \max(|r_i + R_i - c_{\text{new}}|)
\]

5. If \( |r_{\text{max,\ new}} - r_{\text{max}}| < 0.01 \text{ Å} \) stop: \( c_{\text{new}} \) is the center of the sphere and \( r_{\text{max,\ new}} \) its radius. Otherwise repeat steps 2-5.

The radial distance between the reaction fingerprint and the solvent refers to the distance between the centers of the two species. Therefore the minimum distance between them is \( R + R_s \), where \( R \) is the radius of the reaction fingerprint and \( R_s \) is the radius of the smallest sphere encompassing the solvent molecule. Therefore the reaction fingerprint radius parameter determines the starting point of the integration, since value of the radial distribution function will be zero for \( r < R \) due to the “hardness” of both the reaction fingerprint and the solvent.

The distances between the charges of the solvent and the reaction fingerprint in the various orientations are computed previously and then used in the calculation of the potential energy. However, in the case of the solvent dipole model using only the most favorable orientations, it is easier to obtain the distances to the boundary of the solvent to speed up the calculation. Figure 3-8 shows the longitudinal distances of the solvent dipole charges, \( d_p \) for the positive charge and \( d_n \) for the negative, and the parallel distance of the dipole center \( y \), from its boundary, as well as the alignment between solvent and reaction fingerprint for the most favorable orientations. The distances \( d_p, d_n \) and \( y \) can be calculated according to Eq 3-52.
Figure 3-8: The distances from the solvent dipole charge and the orientation of the solvent with respect to the reaction fingerprint in the solvent dipole model with most favorable orientation.

\[
\begin{align*}
y &= r - |c - d|\sqrt{1 - \cos^2 \theta} \\
\cos \theta &= \frac{(p - d) \cdot (c - d)}{|p - d||c - d|} = -\frac{(n - d) \cdot (c - d)}{|n - d||c - d|} \\
d_p &= \frac{(p - d) \cdot (c - d)}{|p - d|} - \frac{l_s}{2} + \sqrt{R^2 - (R - y)^2} \\
d_n &= \frac{(n - d) \cdot (c - d)}{|n - d|} - \frac{l_s}{2} + \sqrt{R^2 - (R - y)^2}
\end{align*}
\]

(3-51)

With \(d_p, d_n\) and \(y\), closed expressions can be obtained for the electrostatic potential energies of the four most favorable orientations, two longitudinal and two parallel, as shown in Eq. 3-52.
\[U^1_{\text{longitudinal}}(r) = \frac{q_s q}{4\pi \varepsilon_0} \left( \frac{1}{r - \frac{l}{2} - (d_n - R_s + l_s)} - \frac{1}{r + \frac{l}{2} - (d_n - R_s - l_s)} \right)\]

\[U^2_{\text{longitudinal}}(r) = \frac{q_s q}{4\pi \varepsilon_0} \left( \frac{1}{r - \frac{l}{2} - (d_p - R_s + l_s)} - \frac{1}{r + \frac{l}{2} - (d_p - R_s - l_s)} \right)\]

\[U^1_{\text{parallel}}(r) = \frac{q_s q}{4\pi \varepsilon_0} \left( \frac{2}{\sqrt{\left( r - (R - y) \right)^2 + \left( \frac{l_s - l}{2} \right)^2}} \right) \geq \frac{2}{\sqrt{\left( r - (R - y) \right)^2 + \left( \frac{l_s + l}{2} \right)^2}} \geq \frac{2}{\sqrt{\left( r - (2R - y) \right)^2 + \left( \frac{l_s - l}{2} \right)^2}} \geq \frac{2}{\sqrt{\left( r - (2R - y) \right)^2 + \left( \frac{l_s + l}{2} \right)^2}} \geq (3-52)\]

\[U^2_{\text{parallel}}(r) = \frac{q_s q}{4\pi \varepsilon_0} \left( \frac{2}{\sqrt{\left( r - (2R - y) \right)^2 + \left( \frac{l_s - l}{2} \right)^2}} \right) \geq \frac{2}{\sqrt{\left( r - (2R - y) \right)^2 + \left( \frac{l_s + l}{2} \right)^2}} \geq \frac{2}{\sqrt{\left( r - (2R - y) \right)^2 + \left( \frac{l_s - l}{2} \right)^2}} \geq \frac{2}{\sqrt{\left( r - (2R - y) \right)^2 + \left( \frac{l_s + l}{2} \right)^2}} \]

where \( p \) and \( n \) are the coordinates of the positive and negative charges of the solvent dipole, \( d \) is the coordinate of the dipole center, \( c \) is the coordinate of the sphere center, \( l_s \) is the length of the solvent dipole, and \( \theta \) is the angle between the dipole axis and the line connecting \( c \) and \( d \).
3.3.4. Obtaining the Parameters of the Reaction Fingerprint

As seen in the previous sections, the reaction fingerprint parameters affect the calculation of the potential energy of interaction between the reaction fingerprint and the solvent and, consequently, the chemical potential of the reaction fingerprint in solution and the corresponding reaction rate. Effectively, the rate of a reaction in a certain solvent, as calculated by the reaction fingerprint model, $k_{rf}$, is a function of the reaction fingerprint’s parameters

$$k_{rf} = k_{rf}(q,l,R) \quad (3-53)$$

Given the experimental values of the rates of a reaction in a number of solvents, $k_{exp,i}$, the reaction fingerprint parameters can be computed so as to minimize the error between the experimental values and those calculated by the reaction fingerprint model.

$$q^*, l^*, R^* = \min \sum_{i} (k_{exp,i} - k_{rf,i}(q,l,R))^2 \quad \text{w.r.t. } q,l,R \quad (3-54)$$

The optimal values of the parameters, $q^*, l^*, R^*$ are effectively a parametrized description of the reaction. With these parameters, the rate of the reaction in a solvent whose rate has not been experimentally determined can be calculated using the reaction fingerprint’s parameters and the solvent molecular structure and charges obtained by conjugation operators (full atomic charges or only the solvent’s dipole, depending on the desired level of detail). These data are used in Eq. 3-49 to determine the reaction rate for the new solvent. The simplest representation of the reaction fingerprint has three parameters and therefore, so as to obtain a statistically meaningful set of parameters, experimental rates of the same reaction in at least three solvents are necessary.
3.3.5. Reference Rate

The reaction fingerprint model computes the effects of the solvent environment on the energy barrier for the transition from reactants to transition state. However, it does not account for the component in the energy barrier that is due to the intrinsic energy of each species, which depends only on their molecular structures. This portion of the rate coefficient is equivalent to the rate of the reaction when there is no potential energy interaction between the reacting species and will be called the reference rate \( k_0 \), which appears in Eq. 3-49.

If the reaction fingerprint computed the reaction rate in solution exactly, the reference rate could be obtained by taking the difference between the experimental rate and the calculated solvent effect.

\[
\ln k_{\text{exp}} - \frac{1}{RT} \rho \int_0^1 kT e^{-u_A^{RF}(r)/kT} - 1 \, 4\pi r^2 \, dr = \ln k_{\text{exp}} - \ln k_{rf} + \ln k_0 \quad (3-55)
\]

However that is not the case and the reference rate \( \ln k_0 \) also includes an error term which compensates for errors due to short range interactions, polarizability, specific interactions, etc. To reduce the biasing of the reference rate due to model errors, \( \ln k_0 \) is computed as the average difference between the logarithm of the experimental rate and the solvent effect computed by the reaction fingerprint model..

\[
k_0 = \frac{1}{n_s} \sum_{i=1}^{n_s} \left[ \ln k_{rf,i} - \frac{1}{RT} \rho \int_0^1 kT e^{-u_i^{RF}(r)/kT} - 1 \, 4\pi r^2 \, dr \right] \quad (3-56)
\]
where $n_s$ is the number of solvents whose solution reaction rates were used in the calculation of the reaction fingerprint's parameters.

### 3.4. Implicit and Explicit Assumptions of the Reaction Fingerprint Model.

In order to provide balance between accuracy and simplicity, the reaction fingerprint model makes a number of assumptions and simplifications whose validity will be examined in Chapter 5. The main premise of the model is that, by considering only electrostatic interactions between a solvent and an abstract entity which aggregates all reacting species using a simplified statistical mechanical framework, one can obtain variations in behavior from one solvent to another that are rich and descriptive enough to allow the computation of the effects of those solvents on reaction rates. In other words, the interaction of a solvent with respect to a simple dipole includes enough information to characterize its interaction with more complex structures. From this main premise, the following assumptions were incorporated in the reaction fingerprint model:

1. The interactions of a solvent with the reacting species in a chemical reaction can be adequately described by its interaction with a single, hypothetical molecular entity.

2. Electrostatic interactions are the principal basis of solvent effects on reaction rates and therefore contain enough information for the determination of solvent effects, even when no other kind of interaction is considered.

3. In the calculation of the free energy of solvation of the reaction fingerprint a simplified statistical mechanical expression is used whereby the radial distribution function of the solvent around the reaction fingerprint is computed using only a pair-wise electrostatic potential term. This expression neglects the many-to-many interaction among the solvent molecules as they rearrange themselves to
accommodate the reaction fingerprint; it also neglects the one-to-many interaction between the reaction fingerprint and the surrounding ensemble of solvent molecules. The underlying assumption is that even though these interactions are not considered, the pair-wise interaction is enough to allow the differentiation between solvents of different electrostatic characteristics and the calculation of solution reaction rates at an accuracy suitable for the selection of solvents.

4. When only the lowest potential energy orientations between the solvent and the reaction fingerprint are considered in the calculation of the potential energy, its value is consistently underestimated. It is assumed that the discrepancy between the actual and calculated values will be small, considering that low potential energy orientations have much higher statistical weights, even if the solvent has low polarity.
4. Case Studies

The reaction fingerprint model was applied to five different chemical reactions for which kinetic data is available for several solvents. These reactions have widely different charge transfer (or lack thereof) mechanisms during the activation step.

The reaction fingerprint model is based on the concept of a hypothetical molecular entity which has three parameters in its most simple incarnation: charge, dipole length and radius parameters. While the charge and dipole length parameters describe a measure of the polarity of the solvent, the radius parameter provides geometrical information, describing the availability of the dipole to solvation.

The reaction fingerprint describes the charge evolution during the activation step of a reaction in that its free energy of solvation is equal to the difference in energies of solvation between transition state and reactants, or in other words, the solvent effect on the free energy of activation and thus on the reaction rate. The free energy of activation is calculated by a statistical mechanical expression which depends on the electrostatic potential energy of interaction between solvent and reaction fingerprint, which on its turn is a function of the reaction fingerprint’s parameters, as follows:

All other factors remaining the same, the potential energy of interaction increases with the charge parameter, as the reaction fingerprint becomes more polarized.

All other factors remaining the same, the potential energy of interaction increases with the dipole length parameter, due to the fact that the charges become more available to solvation.

All other factors remaining the same, the potential energy of interaction decreases with the radius parameter, as the reaction fingerprint’s dipole becomes more shielded from solvation.
By adjusting these parameters, one can calculate the rate of a reaction in a certain solvent. Since the reaction fingerprint parameters are intrinsic to the reaction in question and do not depend on the solvent environment, they are a parametrized description of the reaction that can be applied to any solvent.

The set of reaction fingerprint parameters that describes a reaction is obtained by minimizing the error between the rates calculated by the model and the experimental values. Solvents for which kinetic data is known for a certain reaction are used to determine the reaction fingerprint parameters. These solvents are termed test solvents.

Once a set of reaction fingerprint parameters has been obtained for a reaction, its rate in a new solvent for which experimental data is not available can be calculated provided the solvent’s atomic charges and molecular geometry is known. By using the solvent data in conjunction with the reaction fingerprint that has been obtained from experimental data in the statistical mechanical expression for the calculation of the free energy of activation in solution, a reaction rate for the new solvent is calculated. This process is represented schematically in Figure 4-1.

To exemplify the application of the reaction fingerprint model, the Finkelstein halide exchange reaction between methyl iodide and radioactive labeled iodide ion [54] [55], shown in Figure 4-2, will be considered. This is an ion-dipole type of reaction where the charge present in the iodide ion is dispersed over the transition state. Since there is dispersion of charge during the activation process, the rate of this reaction decreases significantly with increasing solvent polarity. There is a reduction of four orders of magnitude when water is substituted for acetone, as shown in Table 4-1, which lists the test solvents and their respective kinetic rates.
Figure 4-1: Schematic representation of the application of the reaction fingerprint model.

\[
\Delta G^\circ = k T \rho \int_0^\infty \int U(r) g(r, \xi) 4 \pi r^2 dr d\xi,
\]

\[
\ln k = \ln k_0 - \left( \frac{\Delta G^\circ - \Delta G_o^\circ}{RT} \right)
\]

Reaction fingerprint parameters
- Charge
- Dipole
- Radius

New Solvent
- Atomic charges
- Molecular geometry

Figure 4-2: Methyl iodide-iodide ion exchange reaction.

\[ I^- + CH_3I \rightleftharpoons [I\ldots CH_3\ldots I] \rightarrow I^-CH_3 + I^- \]

Table 4-1: Methyl iodide-iodide ion exchange reaction rates in test solvents.

<table>
<thead>
<tr>
<th>Solvent</th>
<th>Rate (1 mol(^{-1})s(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>2.94x10(^{-5})</td>
</tr>
<tr>
<td>Methanol</td>
<td>6.54x10(^{-4})</td>
</tr>
<tr>
<td>Acetone</td>
<td>0.79</td>
</tr>
</tbody>
</table>

The reaction fingerprint parameters are defined so as to minimize the error between calculated and experimental values of the reaction rates in different solvents. For the methyl iodide-iodide ion exchange reaction and the test solvents above, the reaction fingerprint parameters are

\[
\begin{align*}
q &= 1 \times 10^{-19} \text{ C} \\
\ell &= 0.3 \text{ Å} \\
R &= 5 \text{ Å}
\end{align*}
\]

and the reference rate is \(10^{-9} \text{ mol}^{-1}\text{s}^{-1}\). These parameters can now be used to calculate the rate of the reaction in a new solvent. Applying the parameters to ethanol, the calculated rate is \(1.6 \times 10^{-3} \text{ mol}^{-1}\text{s}^{-1}\). The experimental rate is \(2 \times 10^{-2} \text{ mol}^{-1}\text{s}^{-1}\), or an error of 1.1 orders of magnitude.

The reaction fingerprint model was applied to four other reactions with different charge evolution characteristics during the activation process, as follows.

### 4.1. Case Study 1 - Menschutkin reaction between triethylamine and ethyl iodide

The reaction between tertiary amines and alkyl halides, also known as Menschutkin reaction, has been extensively studied and is considered a standard system for the study of solvent effects on reactions. Its mechanism involves the reaction of two uncharged species to form charged products and as such there is creation of charge during the activation step, which points to marked acceleration of the reaction as solvent polarity is increased. This reaction was selected as the first test of the reaction fingerprint model not only due to the availability of experimental kinetic data for a large number of solvents but more importantly because extensive chemical insights on the system allow for a thorough test of the assumptions of the model.
Abraham and Grellier [56] studied the effects of 32 aprotic and 7 hydroxylic solvents on the Menschutkin reaction between triethylamine and ethyl iodide:

\[ \text{Et}_3\text{N} + \text{EtI} \rightarrow \text{Et}_4\text{N}^+ + \text{I}^- \] (4-1)

In their analysis, solvent effects were decomposed into reactant and transition state contributions. This was done by calculating the free energy of solvation of the reactants using their limiting values of Raoult’s Law coefficient, obtained experimentally for the solvents in question (since both reactants are volatile solutes), according to Eq. 4.2:

\[
\Delta G_t^0(s) = \Delta G_2^0(s) - \Delta G_1^0(s)
\]

\[
\Delta G_i^0(s) = RT \ln \left( \frac{\gamma_2^\infty(s)}{\gamma_1^\infty(s)} \right) \quad (4-2)
\]

where \( \Delta G_t^0 \) is the free energy of transfer, defined as the difference between the free energies of formation in solution of solute \( s \) in solvents 2 and 1 and \( \gamma_2^\infty \) and \( \gamma_1^\infty \) are the Raoult’s Law coefficients at infinite dilution in solvent \( i \). Using Eq. 4.2 to determine the free energy of solvation of the reactants, the transition state contribution was obtained by:

\[
\Delta G_t^0(\text{TS}) = \Delta G_t^0(\text{Et}_3\text{N}) + \Delta G_t^0(\text{EtI}) + \delta \Delta G^\neq
\]

where \( \delta \Delta G^\neq = \Delta G_2^\neq - \Delta G_1^\neq \)

TS refers to the transition state and \( \Delta G_i^\neq \) is the free energy of activation in solvent \( i \), obtained from kinetic data. Kinetic and thermodynamic data are summarized in Table 4-1.

In the reaction fingerprint calculations, a number of solvents were omitted. As the reaction fingerprint cannot currently handle apolar solvents, hexane, cyclohexane, carbon tetrachloride, benzene and symmetric molecules 1,4-dioxane, 1,2-dichloroethane and 1,1,2,2-tetrachloroethane were not considered. Moreover, as conjugation operators
involving sulfur and iodine have not been calculated, dimethyl sulfoxide and iodobenzene were also omitted.

Four solvents were used as test solvents to derive the reaction fingerprint parameters: diethyl ether, ethyl acetate, acetonitrile and nitromethane.

The reaction fingerprint parameters were obtained by searching a space of parameters in the ranges

\[ q: \text{from } 0.01 \times 10^{-19} \text{ C to } 1 \times 10^{-19} \text{ C} \]
\[ l: \text{from } 0.1 \text{ to } 10 \text{ Å} \]
\[ R: \text{from } 0.2 \text{ to } 15 \text{ Å} \]

For every parameter coordinate in the search space, the least squares error between experimental and calculated rates for the test solvents were computed. The set of parameters which minimized the average squared error between experimental and calculated values, i.e. the reaction fingerprint is:

\[ q = 0.04 \times 10^{-19} \text{ C} \]
\[ l = 2.3 \text{ Å} \]
\[ R = 3 \text{ Å} \]

The reference rate \( k_0 \) is \( 0.257 \times 10^{-6} \text{ mol}^{-1}\text{s}^{-1} \). Results of its application to the full set of solvents are detailed in Table 4-2 and Figure 4-3.
<table>
<thead>
<tr>
<th>Solvent</th>
<th>$k \times 10^6$ l mol$^{-1}$ s$^{-1}$</th>
<th>$\Delta G^\circ$ Kcal/mol</th>
<th>$\delta \Delta G^\circ$ Kcal/mol</th>
<th>$\Delta G^\circ_{(Et_3N)}$ Kcal/mol</th>
<th>$\Delta G^\circ_{(EtI)}$ Kcal/mol</th>
<th>$\Delta G^\circ_{(TS)}$ Kcal/mol</th>
</tr>
</thead>
<tbody>
<tr>
<td>hexane</td>
<td>0.0135</td>
<td>26.99</td>
<td>6.05</td>
<td>-1.28</td>
<td>-0.01</td>
<td>4.76</td>
</tr>
<tr>
<td>cyclohexane</td>
<td>0.0216</td>
<td>26.60</td>
<td>5.66</td>
<td>-1.26</td>
<td>-0.21</td>
<td>4.19</td>
</tr>
<tr>
<td>diethyl ether</td>
<td>0.359</td>
<td>24.91</td>
<td>3.97</td>
<td>-1.06</td>
<td>0.10</td>
<td>3.01</td>
</tr>
<tr>
<td>carbon tetrachloride</td>
<td>0.422</td>
<td>24.77</td>
<td>3.83</td>
<td>-1.39</td>
<td>-0.26</td>
<td>2.18</td>
</tr>
<tr>
<td>1,1,1-trichloroethane</td>
<td>3.14</td>
<td>23.60</td>
<td>2.66</td>
<td>-1.22</td>
<td>-0.26</td>
<td>1.18</td>
</tr>
<tr>
<td>toluene</td>
<td>3.37</td>
<td>23.60</td>
<td>2.66</td>
<td>-1.29</td>
<td>-0.33</td>
<td>1.04</td>
</tr>
<tr>
<td>chlorocyclohexane</td>
<td>5.21</td>
<td>23.41</td>
<td>2.47</td>
<td>-0.99</td>
<td>-0.39</td>
<td>1.09</td>
</tr>
<tr>
<td>bromocyclohexane</td>
<td>6.15</td>
<td>23.32</td>
<td>2.38</td>
<td>-1.02</td>
<td>-0.37</td>
<td>0.99</td>
</tr>
<tr>
<td>benzene</td>
<td>5.37</td>
<td>23.21</td>
<td>2.27</td>
<td>-1.12</td>
<td>-0.35</td>
<td>0.80</td>
</tr>
<tr>
<td>ethyl acetate</td>
<td>7.78</td>
<td>23.05</td>
<td>2.11</td>
<td>-0.74</td>
<td>-0.18</td>
<td>1.19</td>
</tr>
<tr>
<td>dioxan</td>
<td>11.8</td>
<td>22.72</td>
<td>1.78</td>
<td>-0.56</td>
<td>-0.21</td>
<td>1.01</td>
</tr>
<tr>
<td>tetrahydrofuran</td>
<td>11.7</td>
<td>22.70</td>
<td>1.76</td>
<td>-1.06</td>
<td>-0.43</td>
<td>0.27</td>
</tr>
<tr>
<td>ethyl benzoate</td>
<td>23.8</td>
<td>22.61</td>
<td>1.67</td>
<td>-0.86</td>
<td>-0.30</td>
<td>0.51</td>
</tr>
<tr>
<td>chlorobenzene</td>
<td>19.5</td>
<td>22.53</td>
<td>1.59</td>
<td>-1.16</td>
<td>-0.38</td>
<td>0.05</td>
</tr>
<tr>
<td>bromobenzene</td>
<td>34.4</td>
<td>22.37</td>
<td>1.43</td>
<td>-1.09</td>
<td>-0.36</td>
<td>-0.02</td>
</tr>
<tr>
<td>1,1-dichloroethane</td>
<td>23.7</td>
<td>22.30</td>
<td>1.36</td>
<td>-1.35</td>
<td>-0.17</td>
<td>-0.16</td>
</tr>
<tr>
<td>chloroform</td>
<td>30.1</td>
<td>22.13</td>
<td>1.19</td>
<td>-1.92</td>
<td>-0.41</td>
<td>-1.14</td>
</tr>
<tr>
<td>2-butanone</td>
<td>39.6</td>
<td>22.03</td>
<td>1.10</td>
<td>-0.68</td>
<td>-0.18</td>
<td>0.24</td>
</tr>
<tr>
<td>iodobenzene</td>
<td>50.2</td>
<td>22.02</td>
<td>1.08</td>
<td>-0.99</td>
<td>-0.46</td>
<td>-0.37</td>
</tr>
<tr>
<td>acetone</td>
<td>65.4</td>
<td>21.62</td>
<td>0.68</td>
<td>-0.30</td>
<td>0.05</td>
<td>0.43</td>
</tr>
<tr>
<td>1,2-dichloroethane</td>
<td>94.1</td>
<td>21.47</td>
<td>0.51</td>
<td>-0.83</td>
<td>-0.17</td>
<td>-0.49</td>
</tr>
<tr>
<td>dichloromethane</td>
<td>79.9</td>
<td>21.42</td>
<td>0.48</td>
<td>-1.31</td>
<td>0.04</td>
<td>-0.79</td>
</tr>
<tr>
<td>aceto phenone</td>
<td>164.0</td>
<td>21.35</td>
<td>0.41</td>
<td>-0.61</td>
<td>-0.24</td>
<td>-0.44</td>
</tr>
<tr>
<td>benzonitrile</td>
<td>152.0</td>
<td>21.32</td>
<td>0.38</td>
<td>-0.66</td>
<td>-0.10</td>
<td>-0.38</td>
</tr>
<tr>
<td>propionitrile</td>
<td>118.0</td>
<td>21.25</td>
<td>0.31</td>
<td>-0.42</td>
<td>0.17</td>
<td>0.06</td>
</tr>
<tr>
<td>nitrobenzene</td>
<td>184.0</td>
<td>21.20</td>
<td>0.26</td>
<td>-0.52</td>
<td>-0.11</td>
<td>-0.37</td>
</tr>
<tr>
<td>dimethylformamide</td>
<td>216.0</td>
<td>20.94</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1,1,2,2-tetrachloroethane</td>
<td>312.0</td>
<td>20.90</td>
<td>-0.03</td>
<td>-2.51</td>
<td>-0.53</td>
<td>3.07</td>
</tr>
<tr>
<td>acetone</td>
<td>227.0</td>
<td>20.64</td>
<td>-0.25</td>
<td>0.13</td>
<td>0.47</td>
<td>0.35</td>
</tr>
<tr>
<td>nitromethane</td>
<td>333.0</td>
<td>20.47</td>
<td>-0.47</td>
<td>0.21</td>
<td>0.54</td>
<td>0.28</td>
</tr>
<tr>
<td>propylene carbonate</td>
<td>684.0</td>
<td>20.31</td>
<td>-0.63</td>
<td>0.62</td>
<td>0.29</td>
<td>0.28</td>
</tr>
<tr>
<td>dimethyl sulfoxide</td>
<td>873.0</td>
<td>20.06</td>
<td>-0.88</td>
<td>0.74</td>
<td>0.29</td>
<td>0.15</td>
</tr>
<tr>
<td>2-butanone</td>
<td>3.79</td>
<td>23.35</td>
<td>2.41</td>
<td>-0.93</td>
<td>0.44</td>
<td>1.92</td>
</tr>
<tr>
<td>2-propanol</td>
<td>3.31</td>
<td>23.43</td>
<td>2.49</td>
<td>-1.06</td>
<td>0.41</td>
<td>1.84</td>
</tr>
<tr>
<td>n-butanol</td>
<td>1.66</td>
<td>23.84</td>
<td>2.90</td>
<td>-1.38</td>
<td>0.29</td>
<td>1.81</td>
</tr>
<tr>
<td>n-propanol</td>
<td>2.53</td>
<td>23.59</td>
<td>2.65</td>
<td>-1.34</td>
<td>0.42</td>
<td>1.73</td>
</tr>
<tr>
<td>ethanol</td>
<td>4.41</td>
<td>23.26</td>
<td>2.32</td>
<td>-1.14</td>
<td>0.57</td>
<td>1.75</td>
</tr>
<tr>
<td>methanol</td>
<td>7.54</td>
<td>22.94</td>
<td>2.00</td>
<td>-1.29</td>
<td>0.84</td>
<td>1.55</td>
</tr>
<tr>
<td>water</td>
<td>149.0</td>
<td>21.16</td>
<td>0.22</td>
<td>1.22</td>
<td>4.16</td>
<td>5.60</td>
</tr>
</tbody>
</table>

Table 4-2: kinetic rate and free energy of transfer (with formamide as reference value) data for the reaction between triethylamine and ethyl iodide [56].
Figure 4-3: Comparison of experimental and reaction fingerprint reaction rates for the Menschutkin reaction between triethylamine and ethyl iodide.

Dotted lines indicate one order of magnitude error bounds.

Figure 4-3 includes solvents which do not exhibit high polarizability, such as aromatic solvents, or strong specific interactions, such as polyhalogenated solvents. The average error is 0.59 orders of magnitude for this set of solvents.

Figure 4-4 shows the results of the calculation for the full set of solvents. From the data, it is seen that for most solvents the calculation yielded reaction rates which are within one order of magnitude of the experimental values. Diethyl ether, toluene, chlorobenzene, ethyl benzoate, bromobenzene, dichloromethane and 1,1-dichloroethane are the solvents whose reaction rate error comes close or exceeds one order of magnitude. Diethyl ether illustrates the difficulties of the reaction fingerprint method to model solvents with low polarity, in which entropic effects, which are consistently underestimated by the method - because the radial distribution is a simple function of the pairwise electrostatic potential -
are comparable to the enthalpic ones, resulting in a calculated free energy of solvation that is smaller than expected. Another source of error for low polarity solvents is the fact that short range interactions such as van der Waals forces, which are not taken into account, play a comparable role to that of electrostatic forces.

The other “misbehaved” solvents represent three groups: aromatic, polyhalogenated compounds and protic solvents. These have also been experimentally shown to have anomalous behavior when compared to other solvents with similar charge distribution properties [56]. To explain the behavior of such solvents it is necessary to understand the nature of the transition state.

![Figure 4-4: Comparison of experimental and reaction fingerprint reaction rates for the Menschutkin reaction between triethylamine and ethyl iodide. Dotted lines indicate one order of magnitude error bounds.](image)

The average absolute error is 0.69 orders of magnitude and the maximum error is 1.55 orders of magnitude.
The reaction between triethylamine and ethyl iodide can generate either a pair of ions (Et₃N⁺ and I⁻) or an ion pair complex (Et₄N⁺.I⁻). To determine which of the two species the transition state resembles most, Abraham and Grellier [56] also examined the correlation between the transfer free energy of solvation for the transition state, ΔG₀(TS) and free energy of transfer data for the pair of ions and the ion pair complex, ΔG₀(Et₄N⁺ + I⁻) and ΔG₀(Et₄N⁺.I⁻). For “normal” aprotic solvents (excluding aromatic and polyhalogenated solvents) a good linear correlation was found between free energy data for the transition state and the ion pair complex. However, little correlation was observed between the data for transition state and the pair of ions, and therefore the transition state is closer to the ion pair complex than the pair of ions along the reaction coordinate. Moreover, taking into account the correlation between transition state and reactants’ free energies of transfer [57], it is possible to infer a charge separation of 0.4 and show that the transition state lies closer to the reactants than the product ion pair complex. This is confirmed by experimental and theoretical evidence, including reaction field theory [58], solvent effects on kinetic rates [59] and kinetic isotope effects [60] [61]

From the three last columns in Table 4-1 one can see that for most aprotic solvents, transition state effects prevail over reactant effects, again with the exception of aromatic and polyhalogenated compounds. This is also substantiated by reaction field theory calculations [58]. The anomalies introduced by aromatic and polyhalogenated solvents result from a lower transition state free energy than expected considering other aprotic solvents with similar electrostatic makeup. Hydroxylic solvents also have considerable transition state effects, although they unexpectedly raise the transition state free energy, and consequently reduce the reaction rate more than aprotic solvents with similar dielectric constants.
### Table 4-3: Reaction fingerprint calculation results for the Menschutkin reaction between triethylamine and ethyl iodide.

<table>
<thead>
<tr>
<th>Solvent</th>
<th>Calculated Rate $\log_{10}(10^6 \times K_{exp})$ (1 mol(^{-1}) s(^{-1}))</th>
<th>Experimental Rate $\log_{10}(10^6 \times K_{calc})$ (1 mol(^{-1}) s(^{-1}))</th>
<th>Error $\log_{10}(10^6 \times K_{exp}) - \log_{10}(10^6 \times K_{calc})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 diethyl ether</td>
<td>0.69</td>
<td>-0.44</td>
<td>-1.13</td>
</tr>
<tr>
<td>2 1-butanol</td>
<td>0.76</td>
<td>0.22</td>
<td>-0.54</td>
</tr>
<tr>
<td>3 1-propanol</td>
<td>0.83</td>
<td>0.40</td>
<td>-0.43</td>
</tr>
<tr>
<td>4 1,1,1-trichloroethane</td>
<td>0.48</td>
<td>0.50</td>
<td>0.01</td>
</tr>
<tr>
<td>5 2-propanol</td>
<td>1.06</td>
<td>0.52</td>
<td>-0.54</td>
</tr>
<tr>
<td>6 toluene</td>
<td>-0.44</td>
<td>0.53</td>
<td>0.97</td>
</tr>
<tr>
<td>7 ethanol</td>
<td>1.07</td>
<td>0.64</td>
<td>-0.43</td>
</tr>
<tr>
<td>8 bromocyclohexane</td>
<td>0.50</td>
<td>0.79</td>
<td>0.28</td>
</tr>
<tr>
<td>9 methanol</td>
<td>1.35</td>
<td>0.88</td>
<td>-0.47</td>
</tr>
<tr>
<td>10 ethyl acetate</td>
<td>0.38</td>
<td>0.89</td>
<td>0.51</td>
</tr>
<tr>
<td>11 tetrahydrofuran</td>
<td>1.21</td>
<td>1.07</td>
<td>-0.14</td>
</tr>
<tr>
<td>12 chlorobenzene</td>
<td>0.29</td>
<td>1.29</td>
<td>1.00</td>
</tr>
<tr>
<td>13 ethyl benzoate</td>
<td>0.18</td>
<td>1.38</td>
<td>1.20</td>
</tr>
<tr>
<td>14 chloroform</td>
<td>0.70</td>
<td>1.48</td>
<td>0.78</td>
</tr>
<tr>
<td>15 bromobenzene</td>
<td>0.04</td>
<td>1.54</td>
<td>1.49</td>
</tr>
<tr>
<td>16 2-butanone</td>
<td>1.64</td>
<td>1.60</td>
<td>-0.04</td>
</tr>
<tr>
<td>17 dichloromethane</td>
<td>0.35</td>
<td>1.90</td>
<td>1.55</td>
</tr>
<tr>
<td>18 1,1-dichloroethane</td>
<td>0.68</td>
<td>1.97</td>
<td>1.29</td>
</tr>
<tr>
<td>19 propionitrile</td>
<td>1.34</td>
<td>2.07</td>
<td>0.73</td>
</tr>
<tr>
<td>20 water</td>
<td>1.38</td>
<td>2.17</td>
<td>0.79</td>
</tr>
<tr>
<td>21 benzonitrile</td>
<td>1.38</td>
<td>2.18</td>
<td>0.80</td>
</tr>
<tr>
<td>22 acetonitrone</td>
<td>1.34</td>
<td>2.21</td>
<td>0.88</td>
</tr>
<tr>
<td>23 nitrobenzene</td>
<td>2.50</td>
<td>2.26</td>
<td>-0.24</td>
</tr>
<tr>
<td>24 acetonitrile</td>
<td>1.49</td>
<td>2.36</td>
<td>0.87</td>
</tr>
<tr>
<td>25 nitromethane</td>
<td>2.77</td>
<td>2.52</td>
<td>-0.24</td>
</tr>
</tbody>
</table>

All these anomalies point to specific solvent-solute interactions, and can be explained by a transition state with low charge separation - as it is closer to the reactants than the product ion pair - and thus high polarizability. Polarizable nonelectrolytes have been shown to have their free energy increased by hydroxylic solvents whereas aromatic solvents have the inverse effect [57] when compared to aprotic, aliphatic solvents of comparable polarity. In this light, the “anomalies” caused by protic and aromatic solvents are actually rather common effects observed in a variety of solutes.
The reaction fingerprint model is able to capture specific interactions in the case of hydroxylic solvents (although no hydroxylic solvents were included in the test group used to derive the properties of the reaction fingerprint). However, it fails to capture effects due to specific interactions in aromatic solvents; with the exception of nitrobenzene, which is highly polar (its dipole moment is almost twice as large as that of water) and whose electrostatic interactions possibly outweigh polarizability effects, the calculated rates for toluene, chlorobenzene, ethyl benzoate, bromobenzene and acetophenone underestimate the experimental values by close to or more than one order of magnitude.

Two factors contribute to this discrepancy in results: the first factor pertains to the geometric modeling of the solvents; the spherical geometrical model fits aliphatic hydroxylic solvents much better than aromatic ones, whose molecular structure is largely planar. This leads to an overestimation of the molecular radius, and consequently the molecular dipole appears more occluded than it actually is, leading to lower potential energy of interaction and lower rates. The other geometrical consideration is that, in the calculation of the interaction energy, dipoles can interact only in parallel and longitudinal configurations; these are not necessarily the most favorable configurations in the case of aromatic, low polarity and apolar molecules and for these molecules more sophisticated models of interaction including a larger number of configurations can possibly lead to more accurate results. However, the most important limitation of the method is not geometric but physical: in apolar and especially in aromatic molecules, instantaneous dipole-induced dipole interactions, also known as dispersion forces, have an important role in the determination of liquid structure, whereby a polar molecule in the vicinity of another with low polarity will deform the electronic cloud around the latter creating an induced dipole. The magnitude of this effect is dependent on the mobility of electrons in electronic clouds, i.e. the polarizability of the molecule. Systems with \( \pi \) electrons exhibit polarizability, which is further enhanced in conjugated \( \pi \) systems such as aromatic molecules. As a consequence, the appearance of an induced dipole will contribute to lower the free energy and thus increase the rate, and this contribution is not captured by the reaction fingerprint. Currently, the reaction fingerprint does not take into account...
dispersion forces; the main difficulty in modeling polarizability is that it is not an intrinsic molecular property, but an emergent characteristic arising from the interaction of solute and environment. However, the mobility of a set of electrons in an atomic orbital can possibly be modeled by the magnitude of weights of the conjugation operators applying to that orbital. At this point, neither the geometric or the physical limitations of the reaction fingerprint model seem insurmountable.

Polyhalogenated solvents pose a different sort of problem: their net effect is a reduction in the energy of the transition state, but this effect is not observed with other polarizable nonelectrolytes, so dispersion force effects are not the cause of their anomalous behavior. However, polyhalogenated solvents do significantly reduce the energy of \( \text{R}_4\text{N}^+\text{I}^- \) ion pairs as compared to other solvents of similar dielectric constant [62], which points to some form of solvent-solute specific interaction, possibly involving the incipient iodide ion in the transition state, which has not been clearly elucidated. Because the reaction fingerprint model strips the details of structure and electronic distribution in reactants and transition state down to their most fundamental description, it cannot account for such interactions; one can imagine reaction fingerprint models with more complex descriptions including critical charge distribution points in both reactants and transition state which could account for these effects, although that would also increase the number of adjustable parameters and experimental data requirements.

### 4.1.1. Incorporation of Anomalous Solvents in the Calculation

In order to determine how well the reaction fingerprint model can adapt to anomalous effects such as the ones described above, the case study was revisited, this time using diethyl ether, ethyl acetate and nitromethane from the past calculation and including “anomalous” solvents water and 1,1-dichloroethane as test solvents. The calculated reaction fingerprint were:
\[ q = 0.03 \times 10^{-19} \text{ C} \]
\[ l = 3 \, \text{Å} \]
\[ R = 2.2 \, \text{Å} \]

and the reference rate was \( k_0 \) is \( 0.93 \times 10^6 \text{ mol}^{-1} \text{ s}^{-1} \). Results for the full set of solvents are shown in Figure 4-5.

From Figure 4-5 it can be seen that a reduction in the overall error can be achieved (average absolute error for \( \log_{10}(k) \): 0.598) by introducing solvents with polarizability and specific interaction effects. The reaction fingerprint parameters are not very different from those in the first calculation as is evidenced in the similar structure of the data when compared to Figure 4-3. The largest difference is in the reference rate \( k_0 \), which allows one to conclude that specific interactions are not incorporated in terms of the parameters of the reaction fingerprint but rather as an error term in the reference rate. This shows that the reaction fingerprint as it is cannot represent specific interactions except as an error correction through \( k_0 \). This calls for a reaction fingerprint model that incorporates polarizability and specific interaction effects, but it also shows that good extrapolation performance can be obtained if a wide range of solvent behavior and charge distribution characteristics is represented in the test solvent set, even in the absence of such models.
Figure 4-5: Results of reaction fingerprint calculation using solvents with anomalous behavior.

Dotted lines indicate one order of magnitude error bounds.

4.1.2. Response Surfaces

To analyze the dependence of the fit of the solution on the reaction fingerprint parameters, the response surface was generated for the test solvent set whereby the reaction fingerprint parameters are mapped to the RMS error between the logarithms of experimental and calculated rates. The response surfaces are shown in Figure 4-6, Figure 4-7 and Figure 4-8.
Comparing Figure 4-6 and Figure 4-7 with Figure 4-8, it can be inferred that the fit of solution is more strongly dependent on charge than on the other parameters, given the large "plain" of sets of $l$ and $R$ parameters with similar fit. Another interesting aspect in Figure 4-6 is that, although there is a unique minimum point, there is a "valley" across the $q$-l surface where a number of points with similar dipole moments to that of the solution have fits only slightly worse than the original solution. This implies that although the dipole length parameter is needed to provide geometric information, the reaction fingerprint model tries to select a dipole moment which, coupled with the geometric parameters, fits the test data.
Figure 4-7: Response surface for Case Study 1 - Least Squares error as a function of charge and sphere radius parameters
4.1.3. Sensitivity Analysis

To determine how the behavior of individual solvents is affected by changes in the reaction fingerprint’s parameters, the reaction rates of the four test solvents were calculated by changing one of the reaction fingerprint parameters and keeping the two remaining ones at their solution values. Figure 4-9, Figure 4-10, Figure 4-11, Figure 4-12 and Figure 4-13 show the sensitivity of diethyl ether, ethyl acetate, acetonitrile and nitromethane to variation in the reaction fingerprint’s charge, dipole length and sphere radius parameters around the solution values.
The reaction rate increases monotonically with the charge parameter as shown in Figure 4-9. As the reaction fingerprint charge increases, so do the potential electrostatic energy $U$ (which has a negative sign) and consequently the radial distribution function $e^{-U/kT}$. This leads to large negative free energies of solvation for the reaction fingerprint and thus the reaction rate increases with the charge parameters. The fact the $U$ has always a negative sign for positive values of the charge parameter is a consequence of considering...
only the most favorable docking configurations between the reaction fingerprint and the solvent. It is seen that only charge values around $0.1 \times 10^{-9}$ C or smaller yield reasonable values for the reaction rate. These values are rather low compared with reasonable molecular charge values (around $1 \times 10^{-9}$ C) and result from the use in the calculation of the free energy of solvation of a constant value for the number density of $8.8 \times 10^{27}$ molecules/m$^3$ (see Chapter 3) which is rather high when compared to that of water (approximately $3.3 \times 10^{27}$ molecules/m$^3$). Smaller values for the number density bring the range of physically feasible charge values to the expected region.

![Figure 4-10: Menschutkin reaction rate as a function of the sphere radius parameter.](image)

$q = 0.04 \times 10^{-16}$ C, $l = 0.7$ Å

- $x =$ diethyl ether
- $+$ = ethyl acetate
- $o =$ acetonitrile
- $.$ = nitromethane
Figure 4-10 shows the calculated reaction rates as a function of the sphere radius $R$. The rates are monotonically decreasing with respect to $R$ for two reasons: first, a larger sphere increases the distance between charges in the reaction fingerprint and the solvent decreasing the potential electrostatic energy and second, $R$ represents the hard boundary of the reaction fingerprint; therefore, the integration of the convolution of the potential energy with the radius distribution function starts at a radial position $R$ from the center of the reaction fingerprint. Increasing $R$ will decrease the value of the integral and increase the (negative) free energy of solvation resulting in a lower rate.

![Graph showing calculated reaction rates as a function of dipole length](image)

**Figure 4-11**: Menschutkin reaction rate as a function of the dipole length parameter.

$q = 0.04 \times 10^{-19}$ C, $R = 3\text{Å}$

- $x$ = diethyl ether
- $+$ = ethyl acetate
- $o$ = acetonitrile
- $.$ = nitromethane
Figure 4-11 shows the rate dependence on the dipole length parameter $l$. The more polar solvents are much more sensitive to variations in $l$ than the low polarity ones. Although in this case the rate is monotonically increasing with $l$, the effect of $l$ on the rate is not as straightforward as that of the other two parameters since two kinds of docking configurations are considered in the calculation of the electrostatic potential energy.

To decouple the effects of each configuration in the calculation of rates, the sensitivity of the rate to $l$ was computed using only one docking configuration at a time.

![Graph showing rate dependence on dipole length]

Figure 4-12: Menschutkin reaction rate as a function of the dipole length parameter with potential energy calculated using only the parallel dipole configuration.

$q = 0.04 \times 10^{-19} \text{ C}, R = 3\text{Å}

x = \text{diethyl ether}

+ = \text{ethyl acetate}

o = \text{acetonitrile}

. = \text{nitromethane}
Figure 4-12 shows the rate dependency on \( l \) when only the parallel dipole docking configuration is considered. In this model, the dipoles of solvent and reaction fingerprint are aligned in parallel with their centers lying on a perpendicular line to the axes of the dipoles. All other parameters remaining constant, in this configuration the electrostatic potential energy will have a minimum point and thus the rate will have a maximum. Although the minimum point lies in this case around \( l = 25 \) Å for all the solvents, which is not realizable in physical terms, it can lie within the feasible region for other solvents and reaction fingerprint parameters. The rate dependency on \( l \) is shown for \( l \) in the range between 0 and 100 Å to illustrate the full behavior of the model, even though the physically reasonable range of \( l \) is between 0 and 10 Å. Comparing Figure 4-12 with Figure 4-11 with respect to the sensitivity of the rate to \( l \), it is clear that for the considered solvents the parallel configuration does not play a large role in determining the rate.

Examination of Figure 4-13, which shows the sensitivity using the longitudinal model and is identical to Figure 4-11 confirms that the rate in this case is determined exclusively by the longitudinal model. In this model, the solvent and reaction fingerprint dipoles are aligned longitudinally, so that the negative charge in one is closest to the positive charge of the other resulting in a negative (attractive) value of the electrostatic potential energy. The dominance of the longitudinal docking configuration is due in part to the geometry of the solvents, but also to the fact that in a spherical reaction fingerprint, the minimum distance between charges is \( R \), the sphere radius in the parallel docking model whereas in the longitudinal model it is \( R - l/2 \) and as such charges can be closer in the longitudinal model resulting in larger sensitivity to the dipole length parameter. A range of values of \( l \) from 0.1 to 8 Å is shown in Figure 4-13 was used for clarity and also because of discontinuities in the potential energy expression for the longitudinal configuration.
Figure 4-13: Menschutkin reaction rate as a function of the dipole length parameter with potential energy calculated using only the longitudinal dipole configuration

\[ q = 0.04 \times 10^{-19} \, \text{C}, \, R = 3 \, \text{Å} \]

- \( x \) = diethyl ether
- + = ethyl acetate
- o = acetonitrile
- . = nitromethane

4.1.4. Full Atomic Charges Model

To verify if the model results could be improved, the reaction fingerprint parameters for the Menschutkin reaction was calculated using the full three-dimensional structure of the solvents (as opposed to just the dipoles) with the corresponding atomic charges. This allows for the use of apolar solvents and provides a more detailed picture of the solvent
distribution. The calculation of the potential energy was performed using only the more favorable configurations (longitudinal and parallel) by aligning the whole molecule along its dipole moment with respect to the reaction fingerprint, as shown in Figure 4-14. Although the dipolar representation of the solvent does take into account the position of the charges within the molecule and thus can account to a certain extent for the accessibility of these charges, the full atomic charge representation allows for more accurate modeling of charge shielding and could potentially improve the results of the model with little extra computational cost.

![Figure 4-14: Schematic representation of the interactions of the dipolar representation of a solvent (left) and its full atomic charge representation (right) with the reaction fingerprint.](image)

Apolar solvents were included in the calculation involving the full atomic charge. The test solvents were diethyl ether, ethyl acetate, water and nitromethane (aromatic and halogenated solvents were not included in the test set). The calculated reaction fingerprint parameters are:

\[ q = 0.05 \times 10^{-19} \text{ C} \]
\[ l = 2.8 \text{ Å} \]
\[ R = 1.5 \text{ Å} \]
The reference rate \( k_0 \) is \( 1.38 \times 10^{-6} \) mol\(^{-1}\) s\(^{-1}\). Results of its application to the full set of solvents are detailed in Figure 4-15 and Table 4-4.

Figure 4-15: Results of reaction fingerprint calculation for the Menschutkin reaction using full atomic charge description. Apolar solvents are included. Dotted lines indicate one order of magnitude error bounds.

The average absolute error is 0.82 orders of magnitude and the maximum absolute error is 2.63 orders of magnitude when all the solvents are included and 0.67 orders of magnitude when apolar solvents are excluded. This data, combined with Table 4-3 and Figure 2-1 show that apolar solvents (hexane, cyclohexane, tetrachloromethane, 1,4-dioxane, 1,2-dichloroethane, 1,1,2,2-tetrachlorethane) are the largest source of error due to the fact that charge effects cannot account for the short range interactions and the polarizability effects observed in these solvents and low-polarity diethyl ether. As shown before, the model does not handle well polyhalogenated and aromatic solvents such as 1,2-dichloroethane, bromobenzene (note however the good performance in the case of...
<table>
<thead>
<tr>
<th>Solvent</th>
<th>Calculated Rate</th>
<th>Experimental Rate</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\log_{10}(10^6 \times K_{exp})$</td>
<td>$\log_{10}(10^6 \times K_{calc})$</td>
<td>$\log_{10}(10^6 \times K_{exp}) - \log_{10}(10^6 \times K_{calc})$</td>
</tr>
<tr>
<td>diethyl ether</td>
<td>1.05</td>
<td>-0.44</td>
<td>-1.49</td>
</tr>
<tr>
<td>1-butanol</td>
<td>1.05</td>
<td>0.22</td>
<td>-0.83</td>
</tr>
<tr>
<td>1-propanol</td>
<td>1.12</td>
<td>0.40</td>
<td>-0.72</td>
</tr>
<tr>
<td>1,1,1-trichloroethane</td>
<td>0.71</td>
<td>0.50</td>
<td>-0.22</td>
</tr>
<tr>
<td>2-propanol</td>
<td>1.53</td>
<td>0.52</td>
<td>-1.01</td>
</tr>
<tr>
<td>toluene</td>
<td>0.14</td>
<td>0.53</td>
<td>0.39</td>
</tr>
<tr>
<td>ethanol</td>
<td>1.41</td>
<td>0.64</td>
<td>-0.77</td>
</tr>
<tr>
<td>bromocyclohexane</td>
<td>0.73</td>
<td>0.79</td>
<td>0.06</td>
</tr>
<tr>
<td>methanol</td>
<td>1.92</td>
<td>0.88</td>
<td>-1.04</td>
</tr>
<tr>
<td>ethyl acetate</td>
<td>0.94</td>
<td>0.89</td>
<td>-0.05</td>
</tr>
<tr>
<td>tetrahydrofuran</td>
<td>1.81</td>
<td>1.07</td>
<td>-0.75</td>
</tr>
<tr>
<td>chlorobenzene</td>
<td>0.65</td>
<td>1.29</td>
<td>0.64</td>
</tr>
<tr>
<td>ethyl benzoate</td>
<td>0.65</td>
<td>1.38</td>
<td>0.73</td>
</tr>
<tr>
<td>chloroform</td>
<td>0.91</td>
<td>1.48</td>
<td>0.57</td>
</tr>
<tr>
<td>bromobenzene</td>
<td>0.45</td>
<td>1.54</td>
<td>1.09</td>
</tr>
<tr>
<td>2-butane</td>
<td>2.09</td>
<td>1.60</td>
<td>-0.49</td>
</tr>
<tr>
<td>dichloromethane</td>
<td>0.82</td>
<td>1.90</td>
<td>1.08</td>
</tr>
<tr>
<td>1,1-dichloroethane</td>
<td>0.86</td>
<td>1.97</td>
<td>1.11</td>
</tr>
<tr>
<td>propionitrile</td>
<td>1.44</td>
<td>2.07</td>
<td>0.63</td>
</tr>
<tr>
<td>water</td>
<td>2.44</td>
<td>2.17</td>
<td>-0.27</td>
</tr>
<tr>
<td>benzonitrile</td>
<td>1.56</td>
<td>2.18</td>
<td>0.62</td>
</tr>
<tr>
<td>acetoephonone</td>
<td>1.62</td>
<td>2.21</td>
<td>0.60</td>
</tr>
<tr>
<td>nitrobenzene</td>
<td>2.57</td>
<td>2.26</td>
<td>-0.30</td>
</tr>
<tr>
<td>acetonitrile</td>
<td>1.60</td>
<td>2.36</td>
<td>0.76</td>
</tr>
<tr>
<td>nitromethane</td>
<td>2.64</td>
<td>2.52</td>
<td>-0.12</td>
</tr>
</tbody>
</table>

Table 4-4: Comparison of experimental and calculated rates for the Menschutkin reaction between ethyl iodide and triethylamine using full atomic charges.

bromocyclohexane), dichloromethane, and 1,1-dichloroethane, suggesting that even a better charge model cannot handle polarizability and hydrogen bonding effects (although no anomalous solvents were included in the computation of the reaction fingerprint parameters) and that specific interaction models that address these effects are required for a more accurate picture. Using full atomic charges slightly improves the results when apolar solvents are not considered (average absolute error of 0.69 orders of magnitude when the dipolar model is used).
4.1.5. Comparison with Statistical Correlation Models

Figure 4-16 and Figure 4-17 show the correlation between reaction rates and the dielectric constant of the solvents and the reaction rate and the Kirkwood parameter, defined as

\[ K = \frac{\varepsilon - 1}{2\varepsilon + 1} \]  

(4-4)

where \( \varepsilon \) is the dielectric constant of the solvent. Table 4-5 shows the absolute average error and correlation coefficients for linear correlation between rates calculated by the reaction fingerprint model, dielectric constant, Kirkwood parameter and experimental reaction rates. It can be seen that the reaction fingerprint yields the best correlation and smallest error of all three methods.

![Figure 4-16: Correlation between solvent dielectric constant and reaction rates for the reaction between triethylamine and methyl iodide.](image-url)
Figure 4-17: Correlation between solvent Kirkwood parameter and reaction rates for the reaction between triethylamine and methyl iodide

<table>
<thead>
<tr>
<th></th>
<th>Reaction Fingerprint</th>
<th>Dielectric Constant</th>
<th>Kirkwood Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average absolute log(k) error</td>
<td>0.5523</td>
<td>0.6855</td>
<td>0.6927</td>
</tr>
<tr>
<td>Correlation coefficient $r^2$</td>
<td>0.4310</td>
<td>0.2887</td>
<td>0.2438</td>
</tr>
</tbody>
</table>

Table 4-5: Absolute average error and correlation coefficient for linear correlation between reaction fingerprint calculated rates, dielectric constant, Kirkwood parameter and experimental reaction rates.
4.1.6. Response Surfaces Using Full Atomic Charges

Figure 4-18, Figure 4-19 and Figure 4-20 show the average squared error, i.e. the difference between the logarithms of the experimental and calculated rates in the region around the set of reaction fingerprint parameters. Again one can see a minimum in the region of reasonable values for the reaction fingerprint parameters. In Figure 4-18, one can see there is a valley of solutions with approximately the same fit roughly following the line where the dipole moment of the reaction fingerprint is between $1.5 \times 10^{-30}$ and $2 \times 10^{-30}$ C.m, as shown in Figure 4-6 when only the solvent dipole was used in the potential energy calculations. This shows that the set of parameters of the reaction fingerprint will approximate a certain dipole moment value that provides the best fit.

![Figure 4-18: Average squared log(k) error as a function of charge and dipole length using the full atomic charge model. Radius = $1.5 \times 10^{-10}$ m](image-url)
Figure 4-19: Average squared log(k) error as a function of radius and charge using the full atomic charge model. Dipole length = 2.8x10^{-10} m

Figure 4-20: Average squared log(k) error as a function of radius and dipole length using the full atomic charge model. Charge = 0.05x10^{-19} C
4.1.7. Sensitivity Analysis Using Full Atomic Charges

When a set of solvent atomic charges are used in the calculation of potential energy instead of the solvent dipole, there is little change in the behavior of the solvents with respect to variations in the reaction fingerprint parameters, as shown in Figure 4-21, Figure 4-22 and Figure 4-23. Comparing these with Figure 4-9 to Figure 4-13 it can be seen that the range of variation in rates when full atomic charges are used is similar to those when only solvent dipoles are used. It can also be seen that apolar solvents are insensitive to the reaction

![Chart](chart.png)

Figure 4-21: Sensitivity analysis of reaction rate with respect to the reaction fingerprint’s charge parameter - Menschutkin reaction between ethyl iodide and triethylamine using full atomic charges.

fingerprint parameters, even when the full atomic charge model is used and therefore models of short range interactions, polarizability and hydrogen bonding need to be
included in the calculation of the potential energy so that the reaction fingerprint model can be applied to apolar solvents.

Figure 4-22: Sensitivity analysis of reaction rate with respect to the reaction fingerprint’s dipole length parameter - Menschutkin reaction between ethyl iodide and triethylamine using full atomic charges.
4.2. Case Study 2: Application of reaction fingerprint parameters to a different Menschutkin reaction

The reaction fingerprint model provides a description of the intrinsic characteristics of a chemical reaction in the form of a set of parameters. It would be desirable for this parametrized representation to allow for the clustering of chemical reactions according to their position in a “reaction fingerprint space” defined by the reaction fingerprint parameter. In this case study, we examine whether the set of reaction fingerprint parameters derived from data from one reaction can be applied to another reaction successfully. The reaction between triethylamine and ethyl iodoacetate [63].

\[
(C_2H_5)_3N + CH_2COOC_2H_5 \rightarrow (C_2H_5)_3NCH_2COOC_2H_5^+ + I^-
\]
has similar charge transfer characteristics as the reaction between iodoethane and triethylamine, examined in the previous section. However, ethyl iodide and ethyl iodoacetate are quite different in both their molecular structure and charge distribution and therefore one should expect a different value for the reference value of the reaction rate involving the latter.

The reaction fingerprint parameters from both the dipole and the full atomic charge models for the reaction between triethylamine and iodoethane were applied to the new reaction.

### 4.2.1. Application of Solvent Dipole Model Parameters

The reaction fingerprint parameters obtained in the previous section using the solvent dipole model are:

\[
q = 0.04 \times 10^{-19} \text{ C} \\
1 = 2.3 \text{ Å} \\
R = 3 \text{ Å}
\]

The reference rate for the new reaction was adopted as the average difference between the calculated and experimental value for the slowest and fastest solvent (p-xylene and nitromethane), in this case, \( \log(k_0) = -2.94 \). Figure 4-24 and Table 4-6 show the results for the solvent dipole model:
Figure 4-24: application of the reaction fingerprint parameters from the reaction between Et₃I + Et₃N to the reaction between CH₃COOC₂H₅ + Et₃N using the solvent dipole model.

<table>
<thead>
<tr>
<th>Solvent</th>
<th>( \log(k_{\text{exp}}) )</th>
<th>( k_{\text{exp}} ) in ( \text{mol}^{-1} \text{min}^{-1} )</th>
<th>( \log(k_{\text{calc}}) )</th>
<th>( k_{\text{calc}} ) in ( \text{mol}^{-1} \text{min}^{-1} )</th>
<th>Error ( \log(k_{\text{exp}}) - \log(k_{\text{calc}}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-xylene</td>
<td>-2.17</td>
<td>-2.90</td>
<td>0.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>toluene</td>
<td>-1.96</td>
<td>-2.79</td>
<td>0.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-butanol</td>
<td>-1.08</td>
<td>-0.67</td>
<td>-0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>acetone</td>
<td>-0.92</td>
<td>-0.40</td>
<td>-0.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>propionic</td>
<td>-0.68</td>
<td>-0.98</td>
<td>0.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>acetonitrite</td>
<td>-0.66</td>
<td>-0.98</td>
<td>0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>benzonitrite</td>
<td>-0.64</td>
<td>-0.94</td>
<td>0.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nitrobenzene</td>
<td>-0.55</td>
<td>0.21</td>
<td>-0.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>acetonitrite</td>
<td>-0.48</td>
<td>-0.83</td>
<td>0.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nitromethane</td>
<td>-0.25</td>
<td>0.48</td>
<td>-0.73</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4-6: application of the reaction fingerprint parameters from the reaction between Et₃I + Et₃N to the reaction between CH₃COOC₂H₅ + Et₃N using the solvent dipole model.
The average absolute error is 0.52 orders of magnitude. When only the subset of solvents that appears in both case studies is used the average absolute error is 0.60 orders of magnitude for the first case study and 0.49 for the second. This is evidence that solvent effects on reactions could be calculated using sets of reaction fingerprint parameters derived from different reactions of the same class.

4.2.2. Application of Full Atomic Charge Model Parameters

Reaction fingerprint parameters using full atomic charges from Case 1 were applied to the new reaction to ascertain whether any improvement could be gained from a more detailed solvent description. The results are shown in Figure 4-25 and Table 4-7:

\[ q = 0.05 \times 10^{-19} \text{ C} \]
\[ l = 2.8 \text{ Å} \]
\[ R = 1.5 \text{ Å} \]

The reference rate was calculated as the average error for the slowest and fastest solvents (p-xylene and nitromethane) and \( \log(k_0) = -2.51 \). The average absolute error is 0.51 (as opposed to 0.52 in the solvent dipole model case), although the error increased for most solvents except for aromatic ones, i.e. toluene, acetophenone and nitrobenzene (p-xylene and nitromethane are not considered as they were used to calculate \( k_0 \)). Acetone has a sharp increase in its error when compared to that of the solvent dipole model, probably due to the inclusion of its exposed carbonyl group in the calculation of the potential energy.

Both models indicate that it is indeed possible to cluster reactions of similar charge transfer properties into groups that share the same reaction fingerprint parameters. Once a large number of reaction fingerprint calculations has been performed, unsupervised learning algorithms could be deployed in “reaction fingerprint space” to identify features.
of different types of reactions with regard to their reactants, transition states and charge transfer characteristics.

![Graph](https://via.placeholder.com/150)

Figure 4-25: application of the reaction fingerprint parameters from the reaction between Et₃I + Et₃N to the reaction between ICH₂COOC₂H₅ + Et₃N using the full atomic charge model.

<table>
<thead>
<tr>
<th>Solvent</th>
<th>log(k&lt;sub&gt;exp&lt;/sub&gt;)</th>
<th>log(k&lt;sub&gt;calc&lt;/sub&gt;)</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-xylene</td>
<td>-2.47</td>
<td>-2.17</td>
<td>0.30</td>
</tr>
<tr>
<td>toluene</td>
<td>-2.46</td>
<td>-1.96</td>
<td>0.50</td>
</tr>
<tr>
<td>2-butanone</td>
<td>-0.51</td>
<td>-1.08</td>
<td>-0.57</td>
</tr>
<tr>
<td>acetone</td>
<td>0.26</td>
<td>-0.92</td>
<td>-1.18</td>
</tr>
<tr>
<td>propionitrile</td>
<td>-1.16</td>
<td>-0.68</td>
<td>0.48</td>
</tr>
<tr>
<td>acetophenone</td>
<td>-0.98</td>
<td>-0.66</td>
<td>0.32</td>
</tr>
<tr>
<td>benzonitrile</td>
<td>-1.04</td>
<td>-0.64</td>
<td>0.40</td>
</tr>
<tr>
<td>nitrobenzene</td>
<td>-0.03</td>
<td>-0.55</td>
<td>-0.52</td>
</tr>
<tr>
<td>acetonitrile</td>
<td>-1.00</td>
<td>-0.48</td>
<td>0.52</td>
</tr>
<tr>
<td>nitromethane</td>
<td>0.05</td>
<td>-0.25</td>
<td>-0.30</td>
</tr>
</tbody>
</table>

Table 4-7: application of the reaction fingerprint parameters from the reaction between Et₃I + Et₃N to the reaction between ICH₂COOC₂H₅ + Et₃N using the full atomic charge model.
4.2.3. Comparison with Statistical Correlation Models

Figure 2-1 and Figure 4-27 show the correlation between dielectric constant, Kirkwood parameter and the experimental reaction rates. Table 4-8 shows the results for a linear correlation between the above parameters and the experimental reaction rates. The correlation coefficient and the average absolute error suggest that the Kirkwood parameter correlation yields the best performance in the case. However, examination of Figure 4-27 shows that there is a very weak correlation between the Kirkwood parameter and the experimental reaction rates.

![Figure 4-26: Correlation between solvent dielectric constant and reaction rates for the reaction ICH₂COOC₂H₅ + Et₃N](image)

Figure 4-26: Correlation between solvent dielectric constant and reaction rates for the reaction $\text{ICH}_2\text{COOC}_2\text{H}_5 + \text{Et}_3\text{N}$
Kirkwood parameter, \((e-1)/(2e+1)\), \(e=\)dielectric constant

Figure 4-27: Correlation between solvent Kirkwood parameter and reaction rates for the reaction \(\text{ICH}_2\text{COOC}_2\text{H}_5 + \text{Et}_3\text{N}\)

<table>
<thead>
<tr>
<th></th>
<th>Reaction Fingerprint</th>
<th>Dielectric Constant</th>
<th>Kirkwood Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average absolute log(k) error</td>
<td>0.5467</td>
<td>0.5552</td>
<td>0.4399</td>
</tr>
<tr>
<td>Correlation coefficient (r^2)</td>
<td>0.4537</td>
<td>0.3688</td>
<td>0.6093</td>
</tr>
</tbody>
</table>

Table 4-8: Absolute average error and correlation coefficient for linear correlation between reaction fingerprint calculated rates, dielectric constant, Kirkwood parameter and experimental reaction rates.
4.3. Case Study 3 - Bromination of 1-pentene

The bromination of olefins is a class of reactions of great industrial importance. Much effort has been made to understand its mechanism and the structure of its transition state. The bromination of 1-pentene was selected as a case study for two reasons: first, it is a reaction which proceeds along a multistep mechanism including equilibrium reactions and thus offers a good opportunity to test the reaction fingerprint model in situations where multiple transition states are present. The second reason is the enormous effect solvents have on the kinetics of the reaction, whose rate spans ten orders of magnitude with increasing solvent polarity, providing a challenge to the model’s ability to encompass such a wide range of solvent effects.

Dubois and Garnier [64, 65] demonstrated that the mechanism of bromination of olefins starts with the formation of a charge transfer complex in a pre-equilibrium with the reactants, whereby the bromine molecule interacts with the $\pi$ orbital electrons in the double bond of the olefin.

\[
\begin{align*}
\text{C} & \quad \text{Br}_2 \\
\text{C} & \quad \text{C}
\end{align*}
\]

Figure 4-28: Formation of charge transfer complex between bromine and 1-pentene

From the formation of the charge transfer complex there are two plausible mechanisms for the bromination. The first involves a first order rate-limiting step whereby the charge transfer complex ionizes forming a bromonium ion followed by a nucleophilic addition with a nucleophile:
The second possible mechanism involves a bimolecular reaction involving simultaneous ionization and nucleophilic attack to the charge transfer complex in a second-order rate-limiting step:

$$\begin{align*}
\text{N}^- + \text{Br}_2 & \xrightarrow{\text{limiting}} \text{N}^-\text{C}^-\text{Br} + \text{Br}^- \\
\text{N}^- + \text{Br}_2 & \xrightarrow{\text{limiting}} \text{N}^-\text{C}^-\text{Br} + \text{Br}^-
\end{align*}$$

Figure 4-30: Second-order mechanism for bromination of 1-pentene: the limiting step is a bimolecular nucleophilic attack in concert with bromine bond cleavage.

Interestingly, solvent effects on the rate of bromination are used as partial evidence to determine the reaction mechanism [66],[67]. The rate-determining transition sate in the first-order mechanism is less bulky and more polarized than that of the second-order mechanism and as a result should be more sensitive to solvation effects. Since the transition states of the two mechanisms closely resemble those of $S_N1$ and $S_N2$ mechanisms, and the correlation between the observed rate and the ionizing power of the solvent is roughly the same as that for an $S_N1$ mechanism, there is evidence favoring the
first-order mechanism. The final proof of the bromonium ion-based, first-order mechanism is based on the stereoselectivity of the reaction [68].

4.3.1. Solvent Dipole Model

The reaction fingerprint parameters using the solvent dipole model and t-butanol, acetic acid and water as test solvents are:

\[ q = 0.4 \times 10^{-19} \text{ C} \]
\[ l = 0.4 \text{ Å} \]
\[ R = 0.2 \text{ Å} \]

with reference rate \( \log(k_0) = -0.32 \). Results of the application of the parameters are shown in Figure 4-3 and Table 4-9:

<table>
<thead>
<tr>
<th>Solvent</th>
<th>( \log k_{\text{exp}} ) (l mol(^{-1}) min(^{-1}))</th>
<th>( \log k_{\text{calc}} ) (l mol(^{-1}) min(^{-1}))</th>
<th>Error ( \log k_{\text{exp}} - \log k_{\text{calc}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>acetic anhydride</td>
<td>0.50</td>
<td>4.16</td>
<td>-3.66</td>
</tr>
<tr>
<td>t-butanol</td>
<td>1.90</td>
<td>2.82</td>
<td>-0.92</td>
</tr>
<tr>
<td>i-propanol</td>
<td>2.15</td>
<td>2.81</td>
<td>-0.66</td>
</tr>
<tr>
<td>acetic acid</td>
<td>2.84</td>
<td>2.17</td>
<td>0.66</td>
</tr>
<tr>
<td>ethanol</td>
<td>2.89</td>
<td>2.87</td>
<td>0.01</td>
</tr>
<tr>
<td>methanol</td>
<td>4.36</td>
<td>3.68</td>
<td>0.67</td>
</tr>
<tr>
<td>formic acid</td>
<td>7.80</td>
<td>2.08</td>
<td>5.71</td>
</tr>
<tr>
<td>water</td>
<td>9.15</td>
<td>8.94</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Table 4-9: Experimental rates [65] and reaction fingerprint calculation results for the bromination of 1-pentene.
Figure 4-31: Comparison of experimental [65] and reaction fingerprint reaction rates for the bromination of 1-pentene.
Dotted lines indicate one order of magnitude error bounds.

The model performs poorly for acetic anhydride and formic acid, even though the results for acetic acid are satisfactory. To understand the failure of the model, one needs to understand the nature of solvent effects on the rate-limiting step. The average absolute error is 1.56 orders of magnitude.

Solvent effects on the limiting step of the first-order mechanism are based on electrophilic assistance of the solvent to the transition state, which closely resembles the bromonium ion [65] [69], as shown in Figure 4-32. In the case of acetic anhydride and formic acid the problem is caused due to the fact that the solvent dipole model does not consider the shielding effect of large atomic charges in the carbonyl group in both compounds. The dipole model implies that the large charges are exposed to interact with the solvent, whereas in the case of acetic anhydride the electrophilic carbon atoms are shielded by oxygen atoms and in the case of formic acid the dipole is closer to the...
hydroxyl group than the nucleophilic carbon atom. This results in acetic anhydride appearing to be a better electrophile than it actually is; the opposite happens to formic acid.

![Diagram of solvent electrophilic assistance to the transition state during limiting step in the bromination of 1-pentene](image)

**Figure 4-32: Solvent electrophilic assistance to the transition state during limiting step in the bromination of 1-pentene**

### 4.3.2. Full Atomic Charge Model

To test the charge shielding hypothesis that has been raised to account for the poor performance of the solvent dipole model, reaction fingerprint parameters were calculated using the full atomic charge model. The reaction fingerprint parameters are:

- \( q = 0.5 \times 10^{-19} \text{ C} \)
- \( l = 0.4 \text{ Å} \)
- \( R = 0.6 \text{ Å} \)

with \( \log(k_0) = 0.90 \). Results are given in Figure 4-33 and Table 4-10. The absolute average error is 0.70 orders of magnitude. The error for acetic anhydride and formic acid decreases dramatically. This indicates that using the full molecular structure combined with individual atomic charges improves the overall performance of the model significantly for a modest additional computational cost, especially in reactions which are subject to large changes in charge distribution.
Figure 4-33: Experimental rates [65] and reaction fingerprint calculation results for the bromination of 1-pentene using full atomic charges.

<table>
<thead>
<tr>
<th>Solvent</th>
<th>log(k) experimental (l mol(^{-1}) min(^{-1}))</th>
<th>log(k) calculated (l mol(^{-1}) min(^{-1}))</th>
<th>Error log(k_{exp} - logk_{calc})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 acetic anhydride</td>
<td>0.50</td>
<td>1.81</td>
<td>-1.31</td>
</tr>
<tr>
<td>2 t-butanol</td>
<td>1.90</td>
<td>2.45</td>
<td>-0.55</td>
</tr>
<tr>
<td>3 i-propanol</td>
<td>2.15</td>
<td>1.87</td>
<td>-0.28</td>
</tr>
<tr>
<td>4 acetic acid</td>
<td>2.84</td>
<td>3.11</td>
<td>-0.27</td>
</tr>
<tr>
<td>5 ethanol</td>
<td>2.89</td>
<td>1.63</td>
<td>1.26</td>
</tr>
<tr>
<td>6 methanol</td>
<td>4.36</td>
<td>3.19</td>
<td>1.13</td>
</tr>
<tr>
<td>7 formic acid</td>
<td>7.80</td>
<td>7.43</td>
<td>0.37</td>
</tr>
<tr>
<td>8 water</td>
<td>9.15</td>
<td>9.52</td>
<td>-0.37</td>
</tr>
</tbody>
</table>

Table 4-10: Experimental rates [65] and reaction fingerprint calculation results for the bromination of 1-pentene using full atomic charges.
4.3.3. Comparison with Statistical Correlation Models

Figure 4-34 and Figure 4-35 show the correlation between dielectric constant, Kirkwood parameter and the experimental reaction rates. Table 4-11 shows the results for a linear correlation between the above parameters and the experimental reaction rates. Again, the reaction fingerprint model yields the lowest average absolute error and the highest correlation coefficient. There is a good linear correlation with the experimental rates for both the dielectric constant and the Kirkwood parameter for the more polar solvents, but this correlation falls apart in the low polarity region.

Figure 4-34: Correlation between solvent dielectric constant and reaction rates for the bromination of 1-pentene.
Figure 4-35: Correlation between solvent Kirkwood parameter and reaction rates for the bromination of 1-pentene.

<table>
<thead>
<tr>
<th></th>
<th>Reaction Fingerprint</th>
<th>Dielectric Constant</th>
<th>Kirkwood Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average absolute log(k) error</td>
<td>0.6965</td>
<td>0.7730</td>
<td>1.8976</td>
</tr>
<tr>
<td>Correlation coefficient $r^2$</td>
<td>0.8904</td>
<td>0.8251</td>
<td>0.1748</td>
</tr>
</tbody>
</table>

Table 4-11: Absolute average error and correlation coefficient for linear correlation between reaction fingerprint calculated rates, dielectric constant, Kirkwood parameter and experimental reaction rates.
4.3.4. Response Surfaces Using Full Atomic Charges

The response surfaces showing the squared ($L_2$) error around the reaction fingerprint parameter region are shown in Figure 4-36, Figure 4-37 and Figure 4-38. The behavior is very similar to that of the first case study with a single minimum point (for clarity, a smaller range than the one that was surveyed is shown in the figures). Again, a “valley” of reaction fingerprint parameter sets which have close electric dipoles present roughly similar fits to the experimental data, indicating the reaction fingerprint, although being a symbolic entity, does indeed represent a set of physically meaningful parameters.

Figure 4-36: Average squared log(k) error as a function of charge and dipole length using the full atomic charge model for the bromination of 1-pentene. Radius = 0.6 Å.
Figure 4-37: Average squared log(k) error as a function of radius and dipole length using the full atomic charge model for the bromination of 1-pentene. Charge = $0.5 \times 10^{-19}$ C.

Figure 4-38: Average squared log(k) error as a function of radius and charge using the full atomic charge model for the bromination of 1-pentene. Dipole length = 0.4 Å.
4.3.5. Sensitivity Analysis Using Full Atomic Charges

Figure 4-39 and Figure 4-41 show the sensitivity of the rates of the test solvents (t-butanol, acetic acid and water) to variations in parameters around the reaction fingerprint set. For these figures, the following convention applies:

\[ + = \text{t-butanol} \]
\[ *= \text{acetic acid} \]
\[ x = \text{water} \]

Figure 4-39: Reaction rate sensitivities of t-butanol, acetic acid and water with respect to the reaction fingerprint charge parameter using the full atomic charge model. \( R = 0.6 \, \text{Å}, \quad l = 0.4 \, \text{Å} \)
Figure 4-40: Reaction rate sensitivities of t-butanol, acetic acid and water with respect to the reaction fingerprint dipole length parameter using the full atomic charge model.

\[ q = 0.5 \times 10^{-19} \text{ C, } l=0.4 \text{ Å} \]

The sensitivities for the bromination of 1-pentene and the Menschutkin reaction cases show that the rates of the more polar solvents are much more responsive to variations in all of the three reaction fingerprint parameters than the less polar ones.
Figure 4-41: Reaction rate sensitivities of t-butanol, acetic acid and water with respect to the reaction fingerprint dipole length parameter using the full atomic charge model.

\[ q = 0.5 \times 10^{-19} \text{ C}, \quad R = 0.6 \text{ Å}. \]

### 4.3.6. Sampling Different Orientations Between Solvent and Reaction Fingerprint

In the previous calculations, the potential energy of interaction between the solvent and the reaction fingerprint was calculated by considering only the most favorable positioning of the solvent relatively to the reaction fingerprint, i.e. longitudinal and parallel alignment of the dipoles of both solvent and reaction fingerprint. By considering only the most favorable arrangement the potential energy is consistently underestimated and as a result the reaction rate is consistently overestimated. The underlying assumption justifying the use of the most favorable docking configurations for the calculation of the potential energy is that the error incurred by using this approximation is roughly the same for all
the solvents and thus it can be absorbed by the reference rate without affecting the relative rates between different solvents. The bromination of 1-pentene, with its dramatic solvent effects is a good reaction to test the validity of this hypothesis. This was done by sampling 576 possible orientations of the solvent around the reaction fingerprint at a fixed radial distance to calculate potential electrostatic energy of interaction between the solvent and the reaction fingerprint, as shown in Figure 4-42.

For each of the orientations, the potential energy is calculated by doing the Boltzmann averaging of the energy values for each configuration. This averaging is done at each step in the integration along the radial distance $r$, as shown in Eq. 4-4, where the subscript $i$ refers to each individual orientation of the solvent around the reaction fingerprint.

$$U(r) = \frac{\sum U_i(r) e^{-U_i(r)/kT}}{\sum e^{-U_i(r)/kT}}$$  \hspace{1cm} (4-4)
This way of calculating the potential energy of interaction is dramatically more computationally intensive as it increases the number of operations by 2 orders of magnitude. The reaction fingerprint parameters using orientation sampling are:

\[ q = 0.3 \times 10^{-19} \text{ C} \]
\[ l = 0.9 \text{ Å} \]
\[ R = 0.9 \text{ Å} \]

and the log\((k_0) = 1.38\). The test solvents were the same ones used in the two previous calculations of the case: t-butanol, acetic acid and water.

The results of the calculation using orientation sampling are shown in Table 4-12 and Figure 4-43.

Figure 4-43: Experimental rates and reaction fingerprint calculation results for the bromination of 1-pentene using full atomic charges.
Table 4-12: Experimental rates and reaction fingerprint calculation results for the bromination of 1-pentene using full atomic charges and orientation sampling.

<table>
<thead>
<tr>
<th>Solvent</th>
<th>log(k_{\text{experimental}}) (l mol(^{-1}) min(^{-1}))</th>
<th>log(k_{\text{calculated}}) (l mol(^{-1}) min(^{-1}))</th>
<th>Error log(k_{\text{exp}} - k_{\text{calc}}) (l mol(^{-1}) min(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>acetic anhydride</td>
<td>0.50</td>
<td>1.66</td>
<td>-1.16</td>
</tr>
<tr>
<td>t-butanol</td>
<td>1.90</td>
<td>2.06</td>
<td>-0.16</td>
</tr>
<tr>
<td>i-propanol</td>
<td>2.15</td>
<td>1.66</td>
<td>0.49</td>
</tr>
<tr>
<td>acetic acid</td>
<td>2.84</td>
<td>2.57</td>
<td>0.27</td>
</tr>
<tr>
<td>ethanol</td>
<td>2.89</td>
<td>1.53</td>
<td>1.36</td>
</tr>
<tr>
<td>methanol</td>
<td>4.36</td>
<td>2.58</td>
<td>1.78</td>
</tr>
<tr>
<td>formic acid</td>
<td>7.80</td>
<td>6.83</td>
<td>0.87</td>
</tr>
<tr>
<td>water</td>
<td>9.15</td>
<td>9.24</td>
<td>-0.09</td>
</tr>
</tbody>
</table>

Comparing Figure 4-43 and Figure 4-33, one can see that the structure of the data changes little when orientation sampling is used instead of the most favorable orientation. The results are essentially the same, although using orientation sampling slightly increases the average absolute error from 0.70 to 0.77 orders of magnitude. This is an indication that the computational expense incurred by sampling hundreds of orientations of the solvent around the reaction fingerprint is not compensated by better performance of the method. The calculated kinetic rates of all solvents using orientation sampling are lower than those calculated by considering only the most favorable orientation. This is due to the fact that by sampling orientations, the potential energy of interaction between solvent and reaction fingerprint is higher than that calculated by considering only the most favorable orientation, and this higher energy barrier results in a lower reaction rate.

The failure of orientation sampling to improve the results of the calculation confirms the original hypothesis that the error in the calculation of potential energy incurred by considering only the most favorable orientations is about the same for most solvents and
thus can be absorbed by the reference rate without the need for much more expensive computation.

### 4.3.7. Response Surfaces Using Full Atomic Charges

Figure 4-44, Figure 4-45 and Figure 4-46 show the average squared error of the logarithm of the reaction rate as a function of the parameters of the reaction fingerprint around the solution. These figures are coarser than the previous response surfaces due to the considerably higher computational cost when orientation sampling is used.

The response surfaces show the same pattern as in the previous case studies: in regions of low polarity one can see a flat surface, as the energy of interaction between solvent and reaction fingerprint is very small; therefore all solvents have approximately the same rate and the reference rate compensates for the error by becoming the average of the experimental rates. As the polarity of the reaction fingerprint increases, the more polar solvents respond much more markedly than less polar ones, and the error increases exponentially. Between the low and high polarity regions lies a valley where the solution (the set of reaction fingerprint parameters that yields the lowest error) can be found.
Figure 4-44: Average squared log(k) error as a function of radius and dipole length using the full atomic charge and orientation sampling model for the bromination of 1-pentene. Charge =0.3x10^{-19} C.

Figure 4-45: Average squared log(k) error as a function of radius and charge using the full atomic charge and orientation sampling model for the bromination of 1-pentene. Dipole length =0.9 Å.
4.4. Case Study 4 - Application to mixtures: solvolysis of $t$-butyl chloride

Winstein and Fainberg [70], [71], [71] examined the effect of a number of solvent mixtures on the kinetics as well as the entropy and enthalpy of activation of the solvolysis of $t$-butyl chloride, shown schematically in Figure 4-47.

During the activation step of the solvolysis, $t$-butyl chloride, a neutral molecule, has its carbon-chlorine bond ionized and as a result the dipole moment of the transition state is approximately $27 \times 10^{-30}$ C.m as opposed to $2.9 \times 10^{-30}$ C.m for the reactant. Such a high increase in polarity leads to strong solvent effects: the rate of solvolysis is increased by a factor of 335,000 when water is substituted for less polar ethanol.
Kinetic data for the solvolysis of butyl chloride were obtained for several different mixtures [71]. Based on this data, a measure of polarity termed the ionization potential $Y$ was defined by the relation:

$$\log k = \log k_0 + mY$$  (4-5)

where $k_0$ is the rate of solvolysis of $t$-butyl chloride in the standard solvent, 80% aqueous ethanol, and $m$ is the sensitivity of the reaction to the ionizing power of the solvent. Since Eq. 4-5 can be written as:

$$\log\left(\frac{k}{k_0}\right) = \frac{1}{2.303RT} (-\Delta G^\neq + \Delta G_0^\neq)$$  (4-6)

The difference between the free energies is due to the change of solvent, so it can be written as a function of the ionizing power of the solvent.

$$-\Delta G^\neq + \Delta G_0^\neq = \frac{\partial \Delta G^\neq}{\partial Y} (Y - Y_0) = \frac{\partial \Delta G^\neq}{\partial Y} Y$$  (4-7)

and thus the sensitivity factor can be expressed as
since \( m \) is a constant that depends only on the reaction it follows that the partial derivative of the free energy with respect to the ionizing power of the solvent is also constant and the free energy of activation is a linear function of the ionizing power of the function. For this reason Eq. 4-7 is called a linear free energy relationship. Fainberg and Weinstein [70] applied this relationship to the solvolysis of t-butyl chloride in 10 binary solvent mixtures and obtained a better fit than other linear correlations involving the dipole moment \( D \) or the Hildebrand parameter \((D-1)/(2D+1)\).

The reaction fingerprint model for solutions considers the interaction of the reaction fingerprint with the two solvents in a binary mixture by calculating the energy of interaction of the reaction fingerprint with each one of the individual solvents. The two calculated energies are than weight-averaged using the molar fraction as weights. With the energy value the chemical potential of the reaction fingerprint in the solvent mixture is calculated as

\[
\mu = -kT \int_{R}^{R_{\text{cut-off}}} \left( e^{\frac{-x_{1}U_{1}(r)-(1-x_{1})U_{2}(r)}{kT}} - 1 \right) 4\pi r^{2} dr
\]

where \( U_{1} \) and \( U_{2} \) are the energies of interaction of the reaction fingerprint with solvents 1 and 2 respectively and \( x_{1} \) and \( x_{2} \) are the molar fractions.

The test solvents used in the reaction fingerprint model were pure water, pure ethanol and 50\% aqueous ethanol in molar fraction. The reaction fingerprint calculation used full atomic charges for the solvent and only the 2 best solvent orientations. The calculated reaction fingerprint parameters are:
\[ q = 0.45 \times 10^{-19} \text{ C} \]

\[ l = 6 \text{ Å} \]

\[ R = 8.5 \text{ Å} \]

and \( \log(k_0) = -48 \). The results are shown in Figure 4-48.

Figure 4-48: Comparison of reaction fingerprint results and experimental results for solvolysis of \( t \)-butyl chloride in mixtures of water and ethanol.

Although the average error between experimental data (solid line) and the calculated rates (crosses) is good, the results do not capture the nonideal nature of the mixture, which manifests itself at the extremes of composition of the mixture. This is due to preferential solvation of the reaction fingerprint by one of the solvents. For instance, water molecules will aggregate themselves around a more polar reaction fingerprint while the less polar ethanol molecules will interact at the more external solvation shells. This effect is not captured by the model as the solvent molar fractions remain constant with respect to the radial distance from the reaction fingerprint. Therefore, to account for preferential
solvation effects, it is necessary to incorporate molar fractions that vary along the distance from the reaction fingerprint.
5. Conclusions and Future Work

A model for the calculation of solvent effects on reaction rates was developed. The model consists of two building blocks: solvent modeling and reaction modeling.

It was decided that the solvent should be modeled in terms of its atomic charges. The reason for this is that properties such as dipole moment or dielectric constants provide only one a one-faceted view of the complex issue of polarity that cannot be matched by the complex behavior obtained when full atomic charges are considered or even the full continuous atomic distribution of charge (which is more computationally expensive). Due to the lack of atomic charge data obtained from the same model for a large number of molecules, a model to predict atomic charges from previous results was developed. This model uses the concept of conjugation operators. Conjugation operators apply to the electrons in a chemical bond creating a localization of charge in one of the atoms in the bond. Each operator applies to a single type of bond and has a weight associated to it. Each atom in the molecule has a charge conjugation equation associated to it that involves the weights of the conjugation operators, which can be determined so as to minimize the error between ab initio charges and those calculated by the charge conjugation equation. Once the operator weights have been determined, they can be extended to molecules for which there are no atomic charge data. Conjugation operators can be described in terms of modifications to the electronic population matrix obtained in quantum mechanical calculations. Conjugation operator weights have been determined from data calculated by the Hartree-Fock model using the 6-31G** basis set and Natural Population Analysis charges for a wide variety of commonly employed solvent as well as alkanes with excellent regression and extrapolation results.

In modeling chemical reactions, the challenge was to sidestep the difficulties associated with determining the structure and charge distribution of the transition state. However, since the rate of a reaction is dependent on the difference in energy between the transition state and the reactants, it is necessary only to model this energy difference rather than the individual components. This is done by defining a molecular entity called the reaction
fingerprint whose energy of solvation is by definition equal to the difference in solvation energy of between the transition state and the reactants. The reaction fingerprint has adjustable parameters describing its polarity and the geometric distribution of its charges. These parameters are determined so as to minimize the error between experimental rates and those calculated by the model. Once the reaction fingerprint parameters have been obtained, they can be applied to the same reaction in a solvent where the rate of the reaction is not known. Since the reaction fingerprint is a molecular entity, its free energy in the solvent is calculated with the tools of statistical mechanics. The reaction fingerprint expresses the intrinsic characteristics of the reaction and the model decouples the reaction from its environment.

The reaction fingerprint model was applied to a number of reactions with different charge transfer characteristics, and rate variations of up to ten orders of magnitude due to solvent effects. A number of different implementations of the model were used. The first implementation considered only the solvent dipole and only the most favorable orientations of the solvent around the reaction fingerprint in the calculation of the energy of interaction between the two species. This implementation was not able to handle solvents which have large charges shielded by the rest of the molecule, although it handled cases with moderate solvent effects well. The second implementation used the full atomic charges of the solvent and only the most favorable orientations and predicted it was effective in all cases, even those with very large solvent effects, predicting rates with an error less than one order of magnitude in the vast majority of solvents considered. To determine if this model could be improved, the energy of interaction between solvent and reaction fingerprint was calculated using a statistical sampling of all possible orientations between the two species, but the results were virtually the same as when full atomic charges and most favorable orientations were used.

Although the reaction fingerprint model has shown to be accurate enough for process design purposes, there are difficulties when the model is applied to low polarity solvents. This is due to the fact that other interactions such as van der Waals forces, which are not included in the model at this point, start playing a role comparable to that of charge in the
interaction between solvent and solute. The model also does not perform well when polarizability effects such as induced dipoles arise, as the solvent charges are not variable.

However, the model is able to handle hydrogen bonds well, as these are basically a consequence of highly exposed charges. There is also evidence that reaction fingerprint parameters transfer well between reactions of the same type, which opens the possibility of clustering a class of reactions around the same set of reaction fingerprint parameters.

5.1. Future Work

The reaction fingerprint model, in addition to models for the calculation of phase equilibria, enables the selection of solvents that satisfy requirements not only of the operation they are involved in, but of the whole process. Thus, one of the most exciting applications of the model will be its incorporation in a system for the selection of solvents as raw materials in a chemical process. This enables the rapid screening of thousands of molecules so as to minimize waste and health hazards by selecting solvents that are easy to recover and have low environmental impact while reducing the inventory and complexity of the process by selecting solvents that can be used in several tasks.

The concept of atomic charges calculated by conjugation operators can be extended from the calculation of solvent effects on reaction rates to the calculation of phase equilibria [72]. Also, the model shows great potential in the area of molecular design, where it can be used to determine atomic charges and related physical properties of molecules whose electronic structure is not known. Another interesting problem would be to determine if the three-dimensional structure of a molecule can be determined by some description of it in terms of conjugation operators.

The reaction fingerprint model can be improved by better solvent modeling rather than by more complex reaction fingerprint formulations. This includes incorporating short range interactions such as van der Waals forces as well as polarizability effects. Since the
reaction fingerprint already has a radius parameter, it could be used in a van der Waals forcefield, with only one extra parameter being required. Since the weights of the conjugation operator are a measure of the energy barrier for charge localization, the difference between single bond operators pertaining to a molecular bond also express the polarizability of that bond. For example, if the difference between the single bond operators is close to zero it indicates high delocalization of the electronic cloud in that bond and therefore high polarizability. On the other hand, a large value indicates that the electronic cloud is highly associated to one atom in the bond and is therefore not susceptible to polarizability effects. If a molecular polarizability index can be built from the differences between single bond conjugations, then it can be easily employed in electrostatic models that calculate the energy of interaction between charge and induced dipole.

Finally, once a number of case studies has been performed for a large number of reactions, a library of reactions could be built in reaction fingerprint parameter space and clusters of reactions could be studied for similar characteristics based on their reaction fingerprint parameters.
6. References


[30] M. M. Francl, Christina Carey, Lisa Emily Chirlian, and D. M. Gange, “Charges Fit to Electrostatic Potentials. II. Can Atomic Charges Be Unambiguously Fit to


