New Solution to an Old Problem: Exploring Properties of Chemical Reactions in Condensed Phases Using Molecular Simulation

Kelly L. Fleming

A dissertation
Submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2015

Reading Committee:
Jim Pfaendtner, Ph.D., Chair
Xiaosong Li, Ph.D.
Qiuming Yu, Ph.D.
Scott Dunham, Ph.D., GSR

Program Authorized to Offer Degree:
Chemical Engineering

University of Washington

Abstract

New Solution to an Old Problem: Exploring Properties of Chemical Reactions in Condensed Phases Using Molecular Simulation

Kelly L. Fleming

Chair of the Supervisory Committee: Professor Jim Pfaendtner

Chemical Engineering

The study of chemical reactions is a foundation of chemical engineering, yet there are limited ways to examine how their properties are influenced. Our research has explored the use of molecular simulation to characterize reaction properties in varying environments.

First, we use the most common method, Quantum Mechanics (QM) with a Density Functional Theory (DFT) model to study the hydrolysis of a glyosidic bond, which is important for the breakdown of biomass. Previously unknown detailed mechanistic steps were found for the hydrolysis reaction. The reaction was subsequently simulated with a continuum model in several different acid solvents, and the reaction energetics were shown to be directly related to the inverse of the dielectric constant. However, the mechanistic details were unchanged.

Based on the trend observed between acid solvents, we set out to learn more about the way the solvents affect reaction properties. In order to accurately determine how solvent molecules affect the reaction, they must be modeled explicitly, as opposed to a continuum model. In order to facilitate fast computation in a reasonable timeframe, we overcame computational limitations stemming from the explicit solvent molecules by pairing a

multiscale modeling approach known as metadynamics with the Car Parrinello molecular dynamics method. We observed a stabilizing effect from the solvent. However, after several failed attempts to quantify the barrier heights, we discovered that this approach does not lead to reproducible or accurate estimates for reaction barriers, despite being the consensus approach in literature.

Instead of purely using metadynamics, we demonstrated for the first time that a method called “MetaRates” can be used to make estimates of a chemical reaction rate. This was an exciting addition to a method that previously had not been used to study systems with ab initio potentials. In a detailed investigation, we were able to demonstrate that this method is successfully used to predict the energy barrier heights that match with quantum mechanics. Based on our work, we believe that MetaRates can be used to model more complex reactions, such as those taking place in enzymes, on surfaces, and in complex solvents like those we are interested in for continuing biomass reaction research.

I Acknowledgements and Dedication

My progress through the PhD program has greatly hinged on the help of others in the Pfaendtner Research Group. I could not have completed the program without support from every member of our group, and most importantly Professor Pfaendtner, to develop my intuition and knowledge about computers, physical and organic chemistry, and the scientific research process. Particularly, the help and support of Vance Jaeger, Pat Burney, Mike Deighan, Kayla Sprenger, Michael Tung, and Sean Fischer were crucial in advancing me through the program. In addition to developing me as a researcher, Professor Pfaendtner has encouraged and aided in my career goals, and helped me develop into the professional I strived to be upon graduation.

Additionally I would like to dedicate this dissertation the many friends and family members, near and far, who have supported me during my graduate tenure through helping me develop as a professional and encouraging me to further my career in science. My parents have been essential in setting me up for success since I was born, and I will never be able to repay them for that except for in my efforts to continue to strive for my goals.

I especially would like to thank Alice Popejoy and Melanie Roberts for encouraging my involvement in the Graduate and Professional Student Senate and the Science Policy Committee at the University of Washington, which have been pivotal in my professional development. Not only have they given me professional advice and supported me through recommendations, but they have been there for me with the inspiring words of friends through tough times.

Table of Contents

Table of Figures

1 Preface

This document consists of the written details of research performed during my time at the University of Washington in the chemical engineering PhD program. My research goals initially were to understand and engineer reactions significant for biomass conversion into sustainable fuels. As my research progressed, it became apparent that some methods to calculate quantitative properties of reactions needed to be altered to attain reliable results with practical computational cost. This document details the results and evolution of the project into new methods to understand reaction properties, which can be applied to complex enzymatic and condensed phase systems important for biomass conversion.

I begin with a discussion of the motivation and goals for this research. Chapter 3 is an article published with Professor Jim Pfaendtner in *The Journal of Physical Chemistry-A* (JPCA) [1]. This publication is the first Density Functional Theory (DFT) study using transition state theory conducted by the Pfaendtner group. We used this method to elicit previously unknown details of a hydrolysis reaction mechanism important to biofuel processing. In this study, we increased our understanding of the energetics and reaction coordinates of the hydrolysis reaction prevalent in biomass breakdown. These details could provide information to develop new catalysts and solvents to aid in biomass breakdown, a central theme to our group's overall research.

Chapter 4 continues our studies on reactions involved in biomass breakdown in condensed phases, detailing unpublished results of studies performed using *ab initio* molecular dynamics with metadynamics on reactions involved in biomass breakdown. We were specifically interested in reactions in water and ionic liquids, another central theme for our group's research.

This research provides insight into the influence of solvent on the reaction properties that are optimal for breaking down biomass. We give qualitative results related to the influence of solvent. We also discuss the computational costs and methods that have been used previously in literature to model the transition state and calculate energetics of the reactions in complex solvents, which presented a bottleneck for these experiments after proving to be inaccurate. Based on our discoveries presented in this chapter, we devised another method, which is discussed in the subsequent chapter.

In Chapter 5, we present a new method, called "MetaRates," which we adapted and extended, developed by our collaborator, Pratyush Tiwary, to calculate the rate of a rare event. For the first time we used this approach to perform a quantitative calculation of a chemical reaction rate using metadynamics. The previously accepted method of calculating energy barriers using metadynamics was discovered to be inconsistent, and therefore an unreliable way to characterize kinetic parameters of reactions. Using MetaRates, we were able to calculate the reaction times with an acceleration factor at multiple temperatures, and use an Arrhenius-like relationship to extract kinetic information about the reaction. We verified the accuracy of calculating the activation energy with MetaRates by comparing the values to those estimated from the potential energy surface of the reaction. Based on this analysis, it is our hope that MetaRates can be used to study reactions in larger and more complex systems, like explicitly solvated models or reactions taking place at the active sites of enzymes. We also show that the previously accepted method for using metadynamics to estimate reaction barriers gives inconsistent barrier heights, with a large standard deviation. The MetaRates method was proven to give results comparable to values attained using high-cost quantum mechanics methods to extract the full potential energy surface with the same semi-empirical basis.

2 Introduction

2.1 Background

Molecular simulation is a rapidly growing and evolving method providing a powerful tool allowing for comparison to, or use in place of, experimental results at atomic and molecular level detail. From coarse-grained simulations of macroscopic molecular systems to quantum electronic structure theory, molecular simulation can predict details of structure and stability, energies at different states, and reaction processes within molecular systems [2]. For these reasons, simulations can predict properties of model systems for less cost than experiments [3]. In biomolecular systems, experimental techniques cannot easily measure dynamics of motion or the free energy surface, which are of particular interest to engineering applications. Fortunately, we can calculate these properties using molecular mechanics.

Molecular mechanics (MM) is a method that uses classical potential energy functions to model molecular systems. MM encompasses molecular dynamics, Monte Carlo simulations, ligand-docking simulations, and molecular structure refinement. MM models are usually a representation of spherical atoms connected by elastic "sticks," or springs, that represent bonds. Mathematical functions are then used to describe internal forces of the system. Classical molecular dynamics simulations (MD) use an MM approach, solving Newton's equations of motion to describe the dynamics of the system with respect to time [4].

MD was one of the first molecular simulation methods and has continued to evolve to become one of the most valuable tools in chemistry, physics, and biology [2]. MD simulations are able to describe structural conformations and movements on a time scale that is typically inaccessible using experimental methods. In biomolecular engineering, MD is used to investigate

thermodynamic and kinetic properties which contribute to our understanding of the role of dynamics and biomolecular structure in recognition and function [4]. Alone, MD is used to find atomic positions, velocities, and energies at those single points. Coupled with statistical mechanics, however, MD can be used to calculate macroscopic properties like pressure, heat capacity, and free energy of microscopic systems.

While MD is able to predict properties of condensed systems using empirical interatomic potentials such the Lennard-Jones, Morse, and Born-Mayer potentials, it is not able to provide information about electronic properties. Quantum mechanics, on the other hand, is able to model the dynamics and bonding of atoms as is defined by the electrons of the system by solving the Schrödinger equation [5] at the expense of losing the dynamic information, therefore becoming independent of time. The Hartree-Fock (HF) method is the most fundamental approach to approximate a solution to the Schrödinger equation, however it neglects the electron correlation, instead treating electrons independently as if they do not interact. Though it is efficient, HF does not provide accurate information for biological systems for this reason.

Density Functional Theory (DFT) is a method that uses the electronic density of a system, taking into account electron correlation to approximate the Schrödinger equation. This method is able to balance the inaccuracy of the HF approximation with the high computational demands of post-HF methods such as Møller-Plesset perturbation theory or composite methods. DFT is therefore able to connect theory with experiment which is useful in validating experimental conclusions and giving insight into geometric, electronic, and spectroscopic properties of systems that cannot be described using molecular mechanics [5].

While DFT can provide accurate descriptions of bonding at a very high computational cost and MD can predict properties of condensed systems of many more atoms than DFT because of its lower computational cost, there is also a need to apply DFT to large or disordered systems to calculate electronic properties as well as describe macroscopic mechanical properties that MD predicts. To meet such a need, Car and Parrinello developed an *ab initio* molecular dynamics (or first principles MD) that combines MD and DFT using a method called Car-Parrinello molecular dynamics (CPMD) [6]. CPMD extends the pair-potential approximation of MD to include accurate descriptions of covalently bonded and metallic systems, as well as making it possible to apply DFT to much larger systems [6].

We have used both ab initio molecular dynamics and quantum mechanics to model reactions in the research presented in this thesis. The next chapter of this document describes the use of DFT to model a hydrolysis reaction of a sugar molecule and describe the free energy states along the reaction coordinate, as well as provide insight into the movements of individual atoms within the system. We then extend the results of this study using CPMD to further investigate hydrolysis and esterification reactions in different solvents, which were modeled explicitly with *ab initio* MD.

Our work using CPMD enhanced with metadynamics to construct free energy surfaces of reactions proved to be a challenge. Sampling of the reactions was not sufficient for a reasonable computational cost when complex solvent atoms were present, and the kinetic data extracted for those reactions was shown to be inconsistent. Because of this challenge, we developed a new method to compute quantitative reaction parameters reactions in solvents.

Chapter 5 details work that was recently submitted for publication on the new application of a method called MetaRates, developed to correctly sample transition state barriers of reactions using ab initio molecular dynamics. Our collaborator, Pratyush Tiwary, and his colleagues, developed a method called MetaRates to gain dynamic knowledge of rare-events. Using this method, reaction times can be calculated at a variety of temperatures, and kinetic data can be studied using an Arrhenius relationship. We discuss our extension of the MetaRates method and results for a model S_N2 reaction. MetaRates can be used to model more complex reactions in condensed phases to extract reliable energy barrier values.

2.2 Motivation

2.2.1 Increasing the efficiency of the degradation of cellulosic materials to produce biofuels

Finding a solution to global climate change has become a worldwide research priority in recent years, encompassing all fields of science - from geology to quantum physics. In the field of chemical engineering, alternative energy is comprehensively researched as a potential solution. Currently, most forms of alternative energy are not provided at a low enough cost to compete with fossil fuel, requiring further development to be implemented into the current energy infrastructure. Solar energy, fuel cells, and plant-based fuel have been a long-standing priority in chemical engineering research. Though each technology is well established, developing novel methods to increase efficiency is necessary for them to be economically competitive with fossil fuel energy sources currently in use.

Transportation fuels pose the largest hurdle to implementing a renewable energy solution to largely decrease carbon dioxide emissions. Starch-based biofuels, such as those made from

corn and cane sugar, are relatively cheap to produce, however, they have been shown to contribute more carbon dioxide to the atmosphere than traditional fossil fuels and they compete with land typically used for food crops [7, 8]. Fuels derived from cellulosic materials, such as wood, switchgrass, and corn stalks, require less energy to grow and harvest and use less land that would otherwise be used to grow food. Additionally cellulosic plants regulate atmospheric carbon dioxide to a greater extent than starch based food crops, therefore reducing the net amount of carbon dioxide in the atmosphere further (see Figure 2.1) [7, 9, 10]. For these reasons, cellulosic fuels are highly sought after, however, they are not currently produced at a low enough cost to compete with traditional fossil fuels or starch-based biofuels.

The largest barrier to cost of cellulosic biofuels is the cost of enzymes that break down the cellulose and hemicellulose components of biomass into sugar monomers, which can then be processed into useable fuels like ethanol. Cellulosic materials consist of polysaccharides of varying chemical structures and compositions made up of sugar monomers linked together by a glycosidic bond [11]. The cleavage of this glycosidic bond requires a large amount of energy, and is assisted by the use of expensive enzymes typically derived from bacteria and fungi [9]. To make the production of cellulosic biofuels more economically competitive, the efficiency of this step in the process needs to be improved.

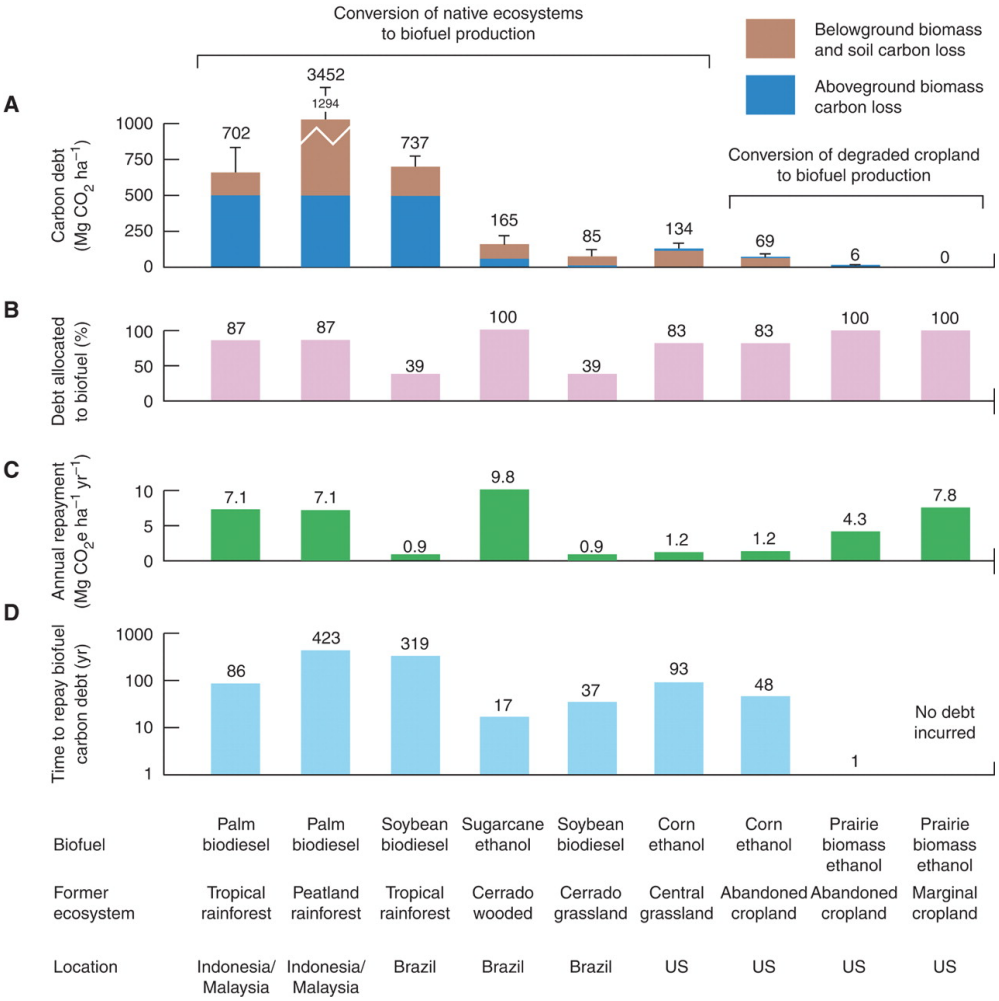

Figure 2.1. From: J Fargione et al. Science 2008;319:1235-1238: Carbon debt, biofuel carbon debt allocation, annual carbon repayment rate, and years to repay biofuel carbon debt for nine scenarios of biofuel production. Means and SDs are from Monte Carlo analyses of literature-based estimates of carbon pools and fluxes (5). (A) Carbon debt, including CO2 emissions from soils and aboveground and belowground biomass resulting from habitat conversion. (B) Proportion of total carbon debt allocated to biofuel production. (C) Annual life-cycle GHG reduction from biofuels, including displaced fossil fuels and soil carbon storage. (D) Number of years after conversion to biofuel production required for cumulative biofuel GHG reductions, relative to the fossil fuels they displace, to repay the biofuel carbon debt.

Our research group aims to better understand the interactions of cellulose and hemicellulose with the enzymes used to facilitate in the hydrolysis of the glycosidic bond, and use that information to engineer environments that enable the process to occur more efficiently. Detailed in chapter 3, we investigated the reaction mechanism of the hydrolysis of the glycosidic bond using density functional theory. In chapter 4 we investigated the interactions of solvents in assisting in the hydrolysis and esterification reactions that take place to further break down cellulosic materials and bio-derived oils into useable fuels using *ab initio* molecular dynamics with the CPMD methodology [6].

After determining the short-comings of a previously accepted method used to investigate reaction kinetics in solution using *ab initio* methods, we used a method called MetaRates, presented in chapter 5 combining metadynamics and reweighting, to adequately model rare events that occur during the transition state of a chemical reaction with *ab initio* molecular dynamics, which was previously only reliably done with quantum mechanics. This new application of the MetaRates method allows us to calculate reaction energies more accurately, with significantly lower computational costs.

3 Characterizing the Catalyzed Hydrolysis of β-1, 4 Glycosidic Bonds Using Density Functional Theory

3.1 Abstract

Unraveling the mechanistic details of biomass deconstruction at ambient conditions has remained a challenge for many years. In this study we examine a crucial step in the pretreatment of biomass: the hydrolytic cleavage of the glycosidic bond present in many forms of biomass and other oligomeric saccharides. We present the detailed mechanistic steps found using density functional theory and transition state calculations on the acid catalyzed hydrolysis of a pyranose dimer linked by a β-1,4 glycosidic bond in vacuum and various continuum solvation models. The order that the bonds in the double displacement reaction form and break was revealed along with the transition state energies and an overall intrinsic reaction pathway for the two step mechanism. The uncatalyzed hydrolysis reaction, mediated by a single water splitting event, was also determined with DFT calculations and a detailed comparison to the two-step catalyzed reaction was performed. The effects of the surrounding solvent on the reaction energetics were studied by systematically changing the dielectric strength and polarity of the solvent model. For acidic solvents, a trend was observed that related the transition state energy barrier to the inverse of the dielectric constant whereas solvents which varied slightly in dielectric but strongly in polarity (e.g., alcohols) did not significantly change the reaction energetics. The effects of the substituents on the model sugar were also studied by changing from a model pyranose dimer to

xylobiose and cellobiose. Irrespective of the solvent choice or model sugar characteristics we observed identical ordering of all bond breaking/forming in both transition states in the double displacement mechanism.

3.2 Introduction

Development of new fuel sources using plant matter has been a longstanding priority in alternative energy research [9]. Although starch-based biofuels, like corn ethanol, are cheap to produce, starchy biomass does not have a high energy density [9] and competes with crop land used for food [8]. Because of this, fuels derived from cellulosic materials, which require less energy to grow and harvest, [8, 9] are highly sought after.

Cellulosic materials are made up mostly of long polysaccharides of different structure and composition. These are long chains of sugar monomers (e.g., glucose and xylose), linked by a glycosidic bond, [9, 11] shown in Figure 3.1. Production of useful fuels or chemicals from these sugar chains requires cleavage of the glycosidic bond [11]. Strategies for achieving this include pyrolysis, acid treatment, or hydrolysis, which is desirable because it is done at lower temperatures and pressures, and is more energy efficient [9]. Limiting the widespread use of enzymatic hydrolysis, however, is the expensive cocktail of enzymes that are produced in bacteria and fungi [8].

Despite a wealth of published experimental and computational work on the hydrolysis of ether linkages in carbohydrates, [8, 9, 11-13] there is still much to be learned. Of particular importance is coupling the detailed mechanism of the reaction with the structural and energetic properties of the substrate and surrounding environment. Enzymatically, it is well established [14] that that the reaction mechanism takes place through either a retaining double displacement

reaction, with two oxocarbenium-ion-like transition states, [11, 15-17] called the Koshland Mechanism [14] or a single displacement inverting reaction, with one transition state [11]. This reaction is facilitated through the enzyme class known as glycoside hydrolase (GH). There are many families of GH enzymes that break the β-1,4 glycosidic bond. The majority of these families follow the retaining mechanism to hydrolyze the glycosidic linkage, however a small number of them use the inverting mechanism. The ordering and energetic relationships within the individual breaking and forming of bonds are not known in the retaining double displacement mechanism, which encompasses families of GH enzymes that hydrolyze cellulose and hemicellulose through the retaining mechanism[11]. Previous studies have uncovered some of the mechanistic details of the first step of the double displacement GH reaction using a quantum mechanics/molecular mechanics hybrid method and metadynamics [17].

The Koshland mechanism [14] for the hydrolysis of glycosidic bonds proposes two concerted transition state complexes in which four bonds are simultaneously broken or formed. A logical next step in better understanding this reaction is to understand the specific order of bonds that break and form in the reaction [8]. Quantum mechanics calculations and transition state theory are an ideal choice for tackling this challenge. Previous studies have found that when the glycosidic bond breaks, a hydrogen from the acid catalyst transfers onto the oxygen to produce glucose, while the other monomer bonds to the nucleophilic acid residue [11]. Likewise, for the second step of the reaction, it has been shown that a water molecule cleaves and the monomeric sugar breaks from the acid residue, where the hydrogen proton transfers onto the nucleophilic residue and the hydroxide transfers onto the monomer, producing another glucose monomer, with the same acid residues in the reactants [11]. However, the specific ordering of these the events and their connection to notable characteristics on the reaction energy surface are

yet unknown. Additionally there is not any information on the bond distances or energies found in the GH reaction for a wide variety of conditions.

To determine details of bond breaking and forming in the GH reaction, we used density functional theory (DFT) to model each of the two reactions in the mechanism. The DFT study was initially done on a model system with a simplified pyranose dimer and carboxylic acid moieties, which serve to model the active site residues. The pyranose dimer has no substituent groups and the carboxylic acids are capped with methyl groups, shown in Figure 3.1. This model was used because it was the simplest system where the relevant chemistry is still preserved. Additionally, the model system is unique in that it could serve as an analog to enzymatic studies (if the acid moieties are conceived of as acidic amino acid sidechains) or as an analog for homogeneously catalyzed hydrolysis in acidic media. Connections to both are provided in the results and discussion. This mechanism was then analyzed for several system variables that are relevant to biomass processing, and compared against each other. This revealed the relationships between the structures and the specific reaction characteristics of the system based on the residues, substituents, and solvents.

3.3 Methods

3.3.1 Systems and Structures Modeled

We first carried out characterization of the reaction using a pyranose dimer due to complexity in precisely determining the relevant transition states while the system undergoes hydrolysis in a more realistic sugar oligomer. The initial system that was used to find the mechanism is shown in Figure 3.1. No hydroxide substituent groups were included, and the catalytic acidic moieties were modeled as acetic acid and an acetate ion. This system represents

the smallest model system that contains all of the relevant chemical functionalities within the vicinity of the transition state (TS) for both steps in the double displacement retaining mechanism. The initial TS characterization was used to extend the substrate complexity and study solvation effects.

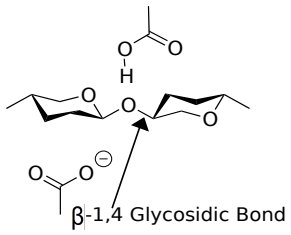

Figure 3.1. Pyranose dimer system containing acetic acid and acetate moieties.

When the reaction mechanism was identified for the pyranose dimer, we systematically varied the solvent effects with an implicit solvent model (described below) as well as the structural features of the substrate. Substituent hydroxide and methyl groups were added to the pyranose dimer to construct xylobiose and cellobiose, which are representative of majority components in biomass feedstocks. The starting structure for cellobiose was based on the work of Mayes and Broadbelt [18] who have already identified the minimum energy conformer with B3LYP calculations. The model substrates we investigated in this study are shown in Figure 3.2.

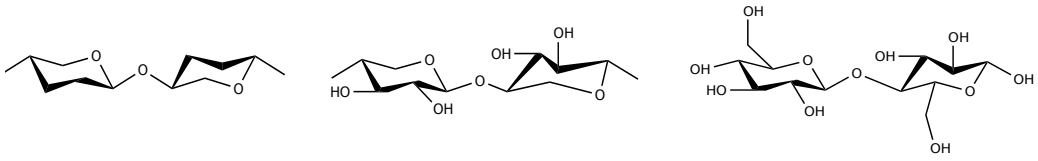

Figure 3.2. Three sugar systems that were modeled.

After the first step in the double displacement reaction mechanism, a monomeric sugar is liberated and a glycol-acidic intermediate remains. To model the second reaction step, we removed the liberated sugar and inserted a water molecule in the vicinity of the region where the reaction is presumed to happen. In total we performed characterization of 12 pyranose two-step reactions (vacuum, water, 10 additional solvents) and additional calculations of xylobiose and cellobiose reactions in the aqueous phase. Additional calculations including the uncatalyzed reaction and others are described below.

3.3.2 DFT calculations and solvation model

All density functional theory (DFT) calculations were completed using the software package Gaussian 09 [19]. Subsequent analyses to determine geometric parameters (e.g., distances and angles) were performed in GaussView and the package PLUMED [20]. All calculations were performed at the B3LYP/6-31+G(d,p) level of theory. Based on results from Mayes and Broadbelt, [18] we concluded that Becke's three parameter hybrid functional (B3LYP) provided a reasonable balance between accuracy and computational cost for size and number of systems we modeled. A diffuse function (+) was added to the basis set to help model the charge distribution arising from the acetate anion. Comparing the results with different sized basis sets changed only trivially any of the observed geometries, did not change any of the

qualitative mechanistic conclusions, and changed only minimally any of the observed energy differences (presumably due to cancellation of errors).

We hypothesized that the presence of the acetate anion would lead to unrealistically large long-range electrostatic effects in the gas phase. Therefore, we employed the implicit self-consistent reaction field with the conductor-like polarizable continuum model (CPCM) [21-23] to account for the charge screening effect of aqueous solvents. We performed both single point calculations and geometry optimizations with the CPCM model to compare the effect of both types of calculations and help assess whether the single point geometries and energies in water, obtained from optimizing in the gas phase, would be very different if re-optimized with implicit solvation. To study the effect of the ionic strength of the solvent, we used the simple model system (pyranose dimer) and studied many different solvent types with a wide range in the system's dielectric constants. These results were used to relate the dielectric constant to the relevant geometric features and the reaction energies in the double displacement mechanism.

3.3.3 Locating energy minima and transition states

Minimum energy stationary points were identified by geometry optimization with the Berny algorithm [24]. We verified the minimum energy position of the substrate/acid complex by performing constrained geometry optimizations searching the torsional angle that defined the axis of rotation between each acid moiety and the ether linkage. The minimum energy structure was then used for all subsequent calculations. Transition states were obtained by scanning previously suggested reaction coordinates [11] with constrained geometry optimizations to identify a putative guess for the transition state. The exact transition states were then found using a quasi-synchronous transit method (QST3) [25, 26]. Transition states were confirmed by verifying that the vibrations had only one imaginary frequency, corresponding to the bond

breaking and forming at the first order saddle point on the energy curve [27]. The correct transition state was confirmed using an intrinsic reaction coordinate (IRC) scan [28], stepping down the energy surface to the true reactants and products connecting the TS. Electronic energies presented in this study are corrected with the zero point energy (ZPE) but due to the presence of many hindered internal rotations and their significant (and inaccurate) contribution to the estimation of vibrational entropy [29], we do not include estimates of the free-energies of reaction or activation.

3.4 Results and Discussion

3.4.1 A fully atomistic view of the double displacement reaction mechanism

Using the ensemble of structures captured from the IRC scan, the detailed mechanism of the hydrolysis of β–1,4 glycosidic bonds was determined. Figure 3.3 and Figure 3.4 follow the distances of key bonds that are breaking or forming in the two reactions, in conjunction with the relative energy as a function of the reaction progress (for states in the vicinity of the TS, determined by scanning the IRC). This analysis is particularly useful for determining the specific order of events in the entire double displacement reaction mechanism and in coupling the reaction energetics with the order of the bonds breaking and forming. We note that the tight range of reaction coordinate scanned in the IRC calculations, shown in Figure 3.3 and Figure 3.4, keeps the system in relatively close proximity to the TS. Therefore it is important to compare the energy scale shown in these two figures to the overall energy scale of the entire reaction pathway.

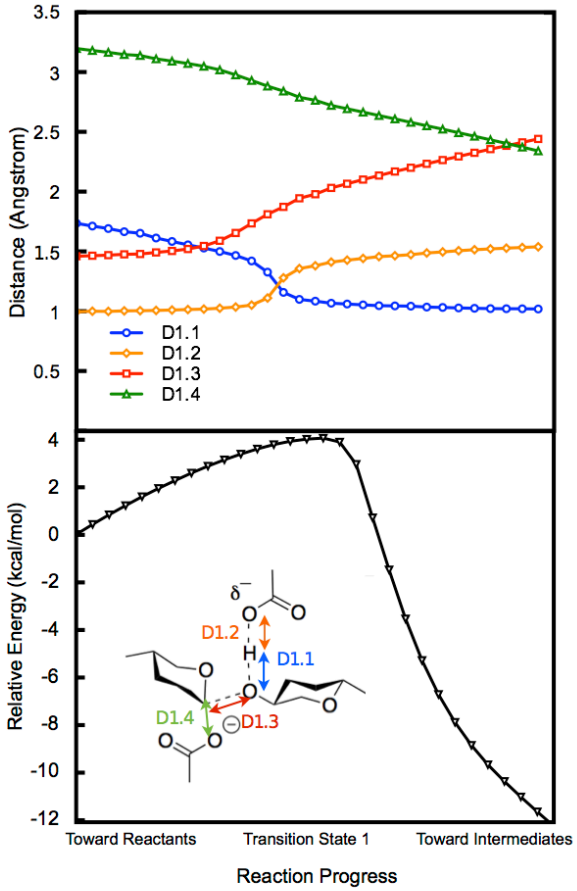

Figure 3.3. A plot of the reaction progress and corresponding bond distance (top) and relative energy (bottom) for the intrinsic reaction coordinate in the vicinity of transition state 1. Key bonds that break or form in the reaction pathway are labeled on the model structure shown in the bottom. The points are equally spaced in terms of the fundamental Cartesian mass-weighted IRC coordinates.

The reactant dimer starts off with both of the rings in the minimum energy chair conformations. As seen in Figure 3.3, the first step in the first reaction is a concomitant increase in D1.3 (red, the glycosidic bond) with a decrease in the distance between the acetate ion and carbenium-like center, D1.4 (green). Next, a proton is donated to the ether linkage from the acid moiety, shown by a sharp increase in length D1.2 (orange) centered at the actual transition state and a decrease in length D1.1 (blue), indicating the protonation has occurred. At this step the

cationic ring is in a skew position in the transition state, and the glucose ring that was formed remains in a chair conformation. Finally, in the vicinity past the specific transition state the ether linkage breaks (D1.3, red) with concomitant decrease in the carboxylate/cation center distance (D1.4, green). In this intermediate step, the rings both go back to the minimum energy chair conformations. The reaction coordinates we found using this simplified system confirm those proposed by Biarnes, et al [17], which modeled the enzyme 1,3-1,4-β-glucanase using metadynamics. Their study found that the glycosidic bond is elongated and there is a simultaneous formation of the hydrogen bond between the Glu09 residue (which we represent as a carboxylic acid) and the glycosidic oxygen. The transition state is reached when the glycosidic bond breaks and the nucleophilic attack on the cationic monomer occurs along with the protonation of the glycosidic oxygen. Our quantum calculations show that the maximum energy on the transition state pathway is reached at the same point, when the protonation is occurring and the glycosidic bond is elongated. The reaction energy in their study begins going downhill and then the bond forms between the residue and nucleophile. This step is represented with the bond forming between the nucleophilic acid moiety and the anomeric carbon ring in our system.

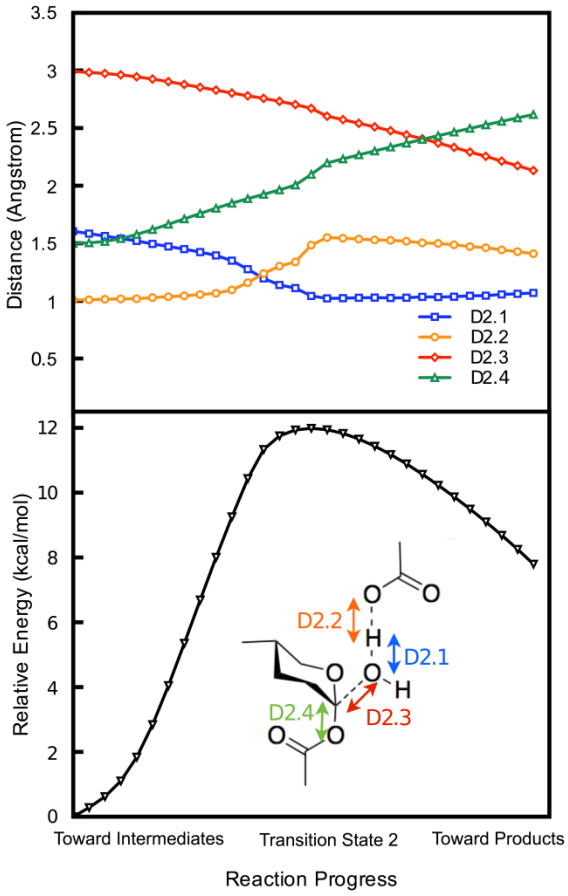

Figure 3.4. A plot of the reaction progress and corresponding bond distances (top) and relative energy (bottom) for transition state 2 along the intrinsic reaction coordinate.

In the second reaction, increases in the bond length between the glucose ring and the carboxylate moiety (D2.4, green) lead the beginning of the reaction progress. The glucose ring begins in the chair conformation. This occurs nearly simultaneously with a decrease in the distance between the oxygen water molecule and the cationic carbon (D2.3, red). As in TS1, the second transition state is associated with a proton transfer step in which the liberated hydrogen from the water molecule restores the original acid moiety (D2.2, orange), and the ring moves into the boat conformation. A distinct difference between the two transition states, however, is the relative sharpness of the proton transfer in TS1 compared to that seen in TS2. After the reaction

has passed through the second barrier there is a simultaneous transfer of the OH group to the sugar ring (decrease in D2.3), and an increase in the distance between all three independent molecules (acid, carboxylate, simple pyranose). The minimum energy ring is again in the chair conformation for the products. The complete mechanism is shown in Figure 3.5, which adds enumeration of the order of the bonds breaking and forming compared to commonly found mechanisms in the literature. Both transition states in this reaction mechanism are concerted reactions with four individual bonds being created or destroyed in each step. Our analysis reveals that the individual event most closely associated with the energy barrier is the transfer of the proton from the acid residue onto the glycosidic oxygen in the first reaction, and the transfer of the hydrogen from the water to the nucleophilic acid residue in the second reaction. We also confirmed that this process is a double displacement reaction with a retaining chemistry, as shown in previous studies. Finally, we note that irrespective of substrate or solvation effects (described in the following sections), we observed the same reaction mechanism steps in each system that was modeled. The complete mechanism found using our studies is shown in Figure 3.5.

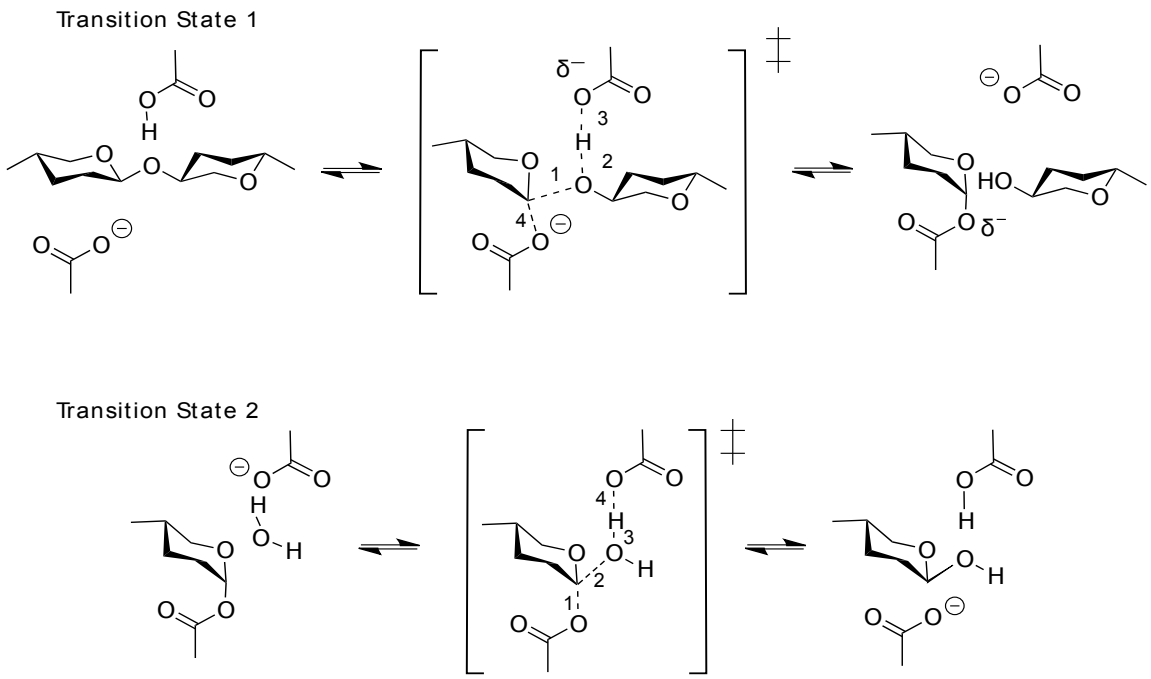

Figure 3.5. The complete double displacement retaining mechanism for the hydrolysis of β-1,4 glycosidic bonds. The bonds breaking and/or forming in the transition state step are labeled with a number corresponding to the order of events described in Figure 3.3 and Figure 3.4.

3.4.2 Comparison to the uncatalyzed reaction mechanism

The role of the acid moieties in catalyzing hydrolysis can be better understood by comparing the acid catalyzed reaction to the uncatalyzed reaction in which a water splits simultaneously with ether cleavage resulting in the passing from reactants to products through a single transition state. The pyranose dimer was modeled with the CPCM model and a single water molecule (which replaces the carboxylic acid catalysts) to facilitate the hydrolysis reaction. Determining the TS for the uncatalyzed reaction was not trivial. First we performed an extensive two dimensional potential energy scan on the simplest uncatalyzed ether hydrolysis (dimethyl ether (DME) to methanol). Using insights from the proposed geometry of the DME

decomposition TS, we scanned similar proposed reaction coordinates with semiempirical PM3 and eventually B3LYP levels of theory in our pyranose model system. Using suggested TS structures from these studies we correctly found a 1st order saddle point on the B3LYP energy surface and verified it connected the correct reactants and products with IRC calculations. The uncatalyzed reaction takes place in one step and has an energy barrier that is nearly three times as large as the catalyzed reaction in water. The energy landscape is shown in Figure 3.6.

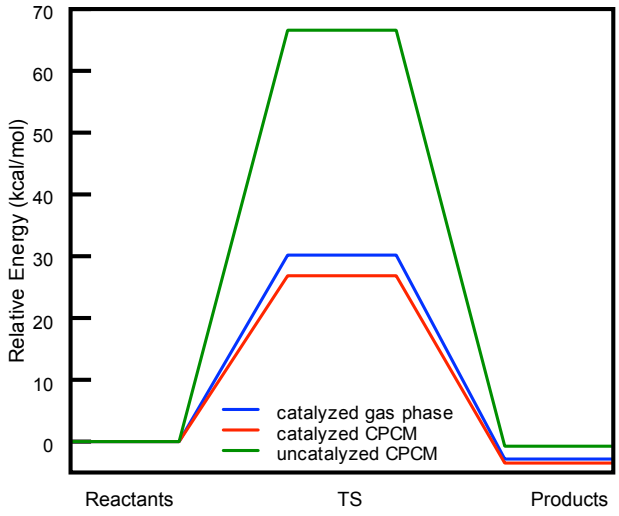

Figure 3.6. The energy landscape of the uncatalyzed reaction compared to the acid catalyzed reaction's first step.

The mechanism is similar to the acid catalyzed mechanism with respect to the transition state structures. The glycosidic bond elongates, corresponding to the initial increase in energy. The water molecule then releases a proton, still corresponding to an increase in energy. The maximum transition state energy corresponds to the proton transferring from the water and onto the glycosidic oxygen. The transition state in each of the two reactions in the catalyzed mechanism is also characterized by a proton transfer onto the glycosidic oxygen. The reaction

energy decreases as the hydroxide bonds to the empty carbon molecule, and the products yield two glucose molecules.

3.4.3 Effects of solvation

The results described above are for a full geometry optimization performed with an implicit solvation model. Using three different TS methods (gas-phase optimization, gas-phase optimization plus single point solvation correction with the CPCM model, and full CPCM optimization) we also calculated the energy landscape for the entire reaction pathway beyond the vicinity of the TS. Comparison the three energy landscapes in Figure 3.7 reveals significant differences between all three cases. Of particular note is the presence of both qualitative and quantitative differences between the two approaches for estimating solvation effects. That is, the addition of the solvation model during geometry optimization differentially changes each barrier height, where TS1 increases with the more expensive calculation, and TS2 decreases. Comparing the gas-phase and full solvation-optimization calculations, we observe that the transition state energy in the first reaction decreases from 30.2 kcal/mol in the gas phase to 26.8 kcal/mol (3.4 kcal/mol lower), which shows that the water has an expected stabilizing effect on the system from solvent effects due to shielding of the electrostatic interactions. Also observed is that the transition state energy of the second reaction decreases from 34.8 kcal/mol to 27.2 kcal/mol (7.6 kcal/mol lower) in water, at the water splitting step. We approximated the $T\Delta S^{\ddagger}$ term from the vibrational frequency calculations on TS1 as 1.2 kcal/mol, which (when combined with the $\Delta U^{\ddagger}$ estimate from our calculation) shows a favorable comparison to the Helmholtz free-energy barrier estimated by Biarnes, et al. [17] of 32 kcal/mol for the same reaction with a GH16 enzyme. Interestingly, this close comparison is obtained between our simple model system and

their detailed study of glycoside hydrolysis in the enzyme 1,3-1,4-β-glucanase with a full QM/MM+metadynamics approach.

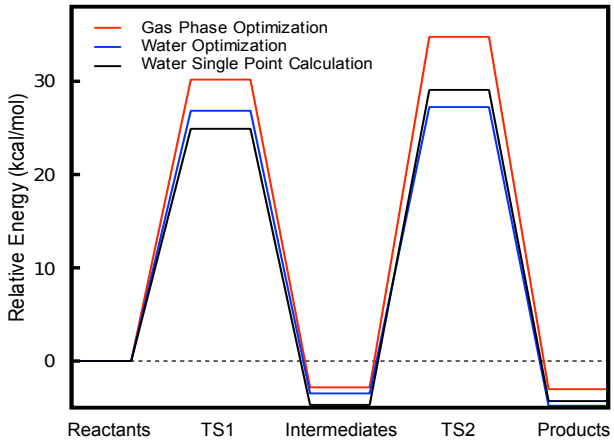

Figure 3.7. The energy landscape of glycosidic hydrolysis in a pyranose dimer in the presence of acetic acid and acetate anion in the gas phase, optimized water phase, and single point energy calculation in water.

Transition state barriers decrease when the CPCM correction is used. This is likely because the water solvent stabilizes the system due to the reactant complex around the nucleophile becoming polar [30]. To better understand the observed trends we closely scrutinized the features of the transition states and the mechanistic details as revealed by the IRC analysis. Although we observed that the mechanistic details described in the previous section remained unchanged in both sets of optimization (gas vs. CPCM), we did observe changes in the actual TS geometry. The same set of geometric features discussed above were determined also for the gas-phase calculation; the comparisons are shown in Table 3.1. The biggest differences are seen in the difference between the optimized bond distance between the assisting carboxylate oxygen atom and the carbon with which it forms an intermediate (D1.4). The next largest changes between the two calculations are seen in the distances related to hydrogen transfer (D1.1 and D2.1). Both of them are significantly closer extended (compared to their source) and closer

to the atom they eventually bond with when the optimization is performed with the implicit solvation model (CPCM). In contrast to the other three lengths, the C-O distance in the ether linkage is changed a comparitively small amount.

Table 3.1. Bond lengths and their degree of difference with respect to the vacuum phase. Numbered lengths correspond to Figure 3.3 for TS1 and Figure 3.4 for TS2.

Reaction	Bond Length (Ångstrom)			
TS1	D1.1	D1.2	D1.3	D1.4
vacuum	1.08	1.38	1.98	2.76
CPCM	1.04	1.50	2.02	3.19
% change	4.60%	-8.02%	-2.11%	-13.4%
TS2	D2.1	D2.2	D2.3	D2.4
vacuum	1.09	1.37	2.05	2.68
CPCM	1.04	1.48	2.10	3.10
% change	4.77%	-7.85%	-2.45%	-13.5%

3.4.4 Effects of varying the solvent type

From these observations, and based on the fact that many non-enzymatic pre-treatment methods of biomass take place in acidic environments [9], we hypothesized that acidic solvents might have an even greater stabilizing effect on the reaction, further lowering the transition state barrier energies. We tested this by varying the solvent in the CPCM model on the pyranose dimer and repeating the transition state calculations (including re-optimization of all relevant states).

The Gibbs free energy is related to the inverse of the dielectric constant according to the Born equations, although due to the complexity of interactions with the solvent a simple trend is not always expected [31]. Figure 3.8 shows the inverse of the dielectric constant of the solvent versus the energy barriers of the reaction and the overall reaction energy, with corresponding

energy values in Table 3.2 and Table 3.3. A minimum energy is found with a dielectric constant between the values of 5 and 10, corresponding to acetic acid, with an energy barrier around 10 kcal/mol. Table 3.2 and Table 3.3 also list the relevant geometric distances in the transition states. We also note there is a qualitative relationship between the organic acid's pKa value, the dieletric constant and the energy barrier heights. The acids with the highest activation energies (C3 and C4 acids) have the lowest dielectric constant and highest pKa values.

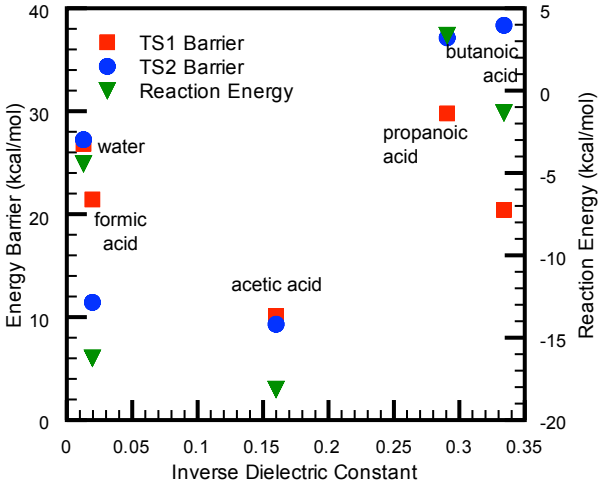

Figure 3.8. The inverse dielectric constant of acid solvents plotted against energy barriers for both transition states (left y axis) and the overall reaction energy (right y axis).

Table 3.2. The first hydrolysis reaction with the pyranose dimer in varying solvents with the TS1 barrier energy, overall reaction energy, and relevant TS1 distances (see Figure 3.3 for bonds).

solvent	dielectric constant	TS1 barrier (kcal/mol)	reaction energy (kcal/mol)	D1.1	D1.2	D1.3	D1.4
water	78.4	26.8	-4.8	1.04	1.50	2.02	3.19
formic acid	51.1	21.4	-16.6	1.04	1.50	2.02	3.18
acetic acid	6.3	10.1	-18.5	1.05	1.48	2.08	3.08
propanoic acid	3.4	29.8	3.0	1.05	1.47	2.09	3.07
butanoic acid	3.0	20.4	-1.7	1.05	1.46	2.09	3.06

Table 3.3. The second GH reaction with the pyranose dimer in varying solvents with the TS2 barrier energy, overall reaction energy, and relevant TS2 distances (see Figure 3.4 for bonds).

solvent	dielectric constant	TS2 barrier (kcal/mol)	reaction energy (kcal/mol)	D2.1	D2.2	D2.3	D2.4
water	78.4	27.2	-4.8	1.04	1.48	2.10	3.10
formic acid	51.1	11.4	-16.6	1.04	1.48	2.10	3.09
acetic acid	6.3	9.3	-18.5	1.05	1.48	2.11	2.98
propanoic acid	3.4	37.1	3.0	1.05	1.46	2.11	2.91
butanoic acid	3.0	38.3	-1.7	1.05	1.46	2.11	2.89

There is a direct correlation between the dielectric constant and a key bond length in both TS1 and TS2 as shown in Table 3.2 and Table 3.3. The stabilizing intermediate bonds D1.4 and D2.4 in both TS1 and TS2 increase from the lowest dielectric to water, and follow a strongly linear correlation of 1/ε vs D1.4 or D2.4, shown in Figure 3.9.

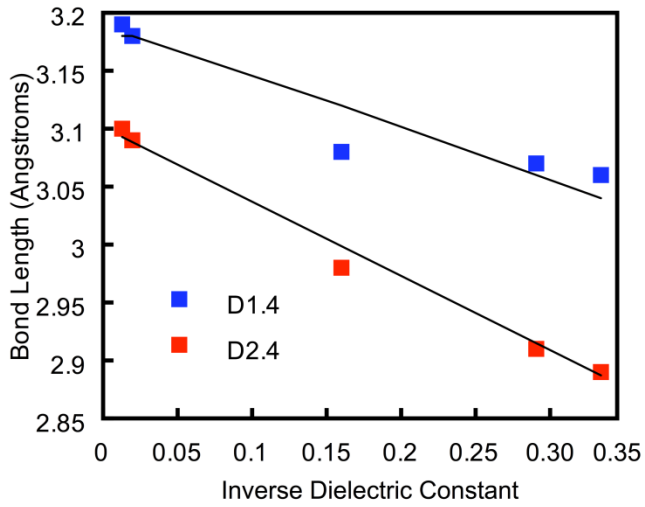

Figure 3.9. Bond lengths D1.4 and D2.4 versus the inverse of the dielectric constant. Linear fits of each data set (black) are shown to guide the eye.

The systems in solvents with high dielectric constants (water and formic acid) have shorter distances between the hydrogen and glycosidic oxygen, and the glycosidic oxygen and

ring from which it broke (D1.3). Transition state 2 also has a smaller distance between the hydrogen and nucleophile it bonds to. This suggests that increasing the dielectric of the reacting medium leads to an increase in the product stability earlier along the intrinsic reaction pathway. To test what effects other types of solvents with various dielectric coefficients might have, four alcohol solvents and formamide were also tested. We found that those solvents did not show an effect on the system energy related to the dielectric constant, and all of the barriers remained around 23 kcal/mol while the dielectric constants varied from 12 to 108. The polarity of acidic solvent has a stabilizing effect on the sugar hydrolysis, which lowers the reaction barriers and helps facilitate the reaction. This result can be understood in context of the fact that more parameters than dielectric constant alone define the properties of the implicit solvent, however it is the most influential and easy to define [19].

3.4.5 Effects of ring substituent groups

To test the relationship between the structure of the model system and the reaction energy landscape, we varied the structural features of our model pyranose dimer, increasing substitution to create dimers of xylobiose and cellobiose. Predicting the changes on the barrier heights and reaction energies a priori was difficult, as the changes to the structure of the substrate are relatively far from the ether linkage itself (c.f., Figure 3.2), but could potentially interact with the acidic moieties assisting hydrolysis. A comparison of the substituent effects on the energy landscape of the three systems is shown below in Figure 3.10. Barrier energies were 26.8 kcal/mol for the pyranose dimer, 23.6 kcal/mol for the xylobiose, and 40.6 kcal/mol, and for cellobiose for TS1; and 27.2 kcal/mol, 18.8 kcal/mol, and 35.7 kcal/mol, respectively, for the second reaction. Although direct comparison with published data is difficult since many studies report a single activation energy, we can observe that values are all within the 10 kcal/mol range

of studies found in literature [17, 32-34] for the transition state energies of cellobiose and xylobiose. With respect to the overall GH families, however, there is a much larger range of barrier energies published for the retaining hydrolysis mechanism [35-37]. Different families of GH have a larger range in transition state barrier energy, which is a result of the expected significant effect on the energy landscape of the reaction provided by each unique substrate and enzyme structure combination. It is clear that the structure of the substrate undergoing hydrolysis can make an impact in the observed barrier heights and reaction energies. However, the direct relationship between substrate and energy landscape is unclear. The observed barrier heights do not correlate with the calculated reaction energies as one might expect from a typical linear free-energy relationship [38].

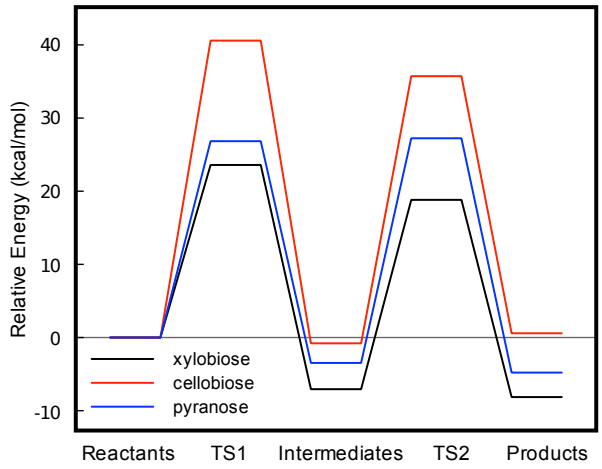

Figure 3.10. The relative energy versus the reaction progress for the three sugar dimers shown in Figure 3.2 with with geometry optimization and energy calculations performed with the solvation model enabled.

The geometric features of the two observed transition states are reported in Table 3.4, with the four key bond distances compared to the pyranose dimer for reference. From the observed differences we hypothesize that each of the different sugar systems' substituent groups

play a relevant role in the energy landscape as they are changing the relevant transition state geometry significantly.

The transition state barrier for the glycosylation step is higher for both cellobiose and xylobiose, but for pyranose the deglycosylation transition state barrier is slightly larger. Table 3.4 shows that in the glycosylation step (transition state 1), pyranose has the shortest bond distances that are being broken in the first step of the reaction. However, in the deglycosylation step (transition state 2), pyranose has similar bond distances as xylobiose in the transition state. This suggests that the substituent groups could influence energy barrier in the first step of the reaction more than it does in the second step of the reaction. This could be due to the extra bulk around the anomeric carbon that is being attacked by the proton in the first step of the reaction.

Table 3.4. Bond distances in TS1 and TS2 for the three systems modeled with the CPCM model on.

state	bond length (Angstrom)			
TS1	D1.1	D1.2	D1.3	D1.4
pyranose	1.04	1.50	2.02	2.90
cellobiose	1.04	1.53	2.05	2.15
xylobiose	1.02	1.55	2.19	2.79
TS2	D2.1	D2.2	D2.3	D2.4
pyranose	1.04	1.48	2.10	3.13
cellobiose	1.03	1.50	2.04	2.75
xylobiose	1.04	1.48	2.12	2.71

3.5 Conclusions

The mechanism of the hydrolysis of β-1,4 glycosidic bonds was investigated using DFT calculations. The pyranose dimer was modeled in the gas phase, water, and four acid solvents with the CPCM correction. Hydrolysis of xylobiose and cellobiose and was modeled in water. Irrespective of the solvent or sugar substitution, the specific order of bonding events in the full

two-step reaction mechanism was preserved for all of the systems. The reaction mechanism for a single step uncatalyzed (single water splitting) reaction was determined and quantitatively compared to the two-step reaction mechanism. The reduction in over 40 kcal/mol in the controlling energy barrier due to the presence of the acid catalysts can be primarily attributed to the stabilization of the intermediate species with the nucleophilic residue. One acid works to protonate the glycosidic oxygen while the other one performs a nucleophilic attack. The first acid then can act as a base to break a proton from the water molecule [11]. In contrast, the uncatalyzed reaction is uncharged and there is no facilitation for protonation or charge transfer between the glycosidic bond and hydrogen.

Our studies of the first reaction (cleavage of the glycosidic bond) supports a previous study [32] that found the same order of events for a full QM/MM model of glycoside hydrolase enzyme. Going further, we also provide quantitative molecular scale information about how the second reaction in which the 2^{nd} sugar is liberated via a deglycosylation from the 2^{nd} assisting acidic moiety. In addition to preserving the mechanistic details, our energy calculations showed remarkable similarity (TS1 energy barriers) to a fully enzymatic model [32] in which the energy barriers were estimated with metadynamics enhanced sampling. This is particularly interesting given that the only similarity to our calculations and an enzyme active site is the presence of two acid molecules that function as catalysts in the retaining GH mechanism.

Addition of a solvent yielded distinct trends. The addition of an acidic solvent to our model system changes the geometry and system energy, and is related to the dielectric constant. Solvation lowers the transition state energy barriers, although more for acid solvents than others. This is a promising result because the cellulose pretreatment method can be fine-tuned using this information to find a solvent that lowers the barrier energy the most for the hydrolysis step. This

should further be confirmed with experiments that test the level of cellulose breakdown with each solvent. It may also be possible to use this information to direct mutagenesis experiments or further calculations in enzymes, perhaps achieving changes in turnover rate by altering the enzyme active site to change the effective dielectric constant.

4 Ab initio molecular dynamics with metadynamics to study reactions involved in biomass conversion in solvents

4.1 Abstract

Using *ab initio* molecular dynamics and DFT we probed how solvents affect the energetic and geometric features of two important reactions in biomass degradation, hydrolysis and esterification. Hydrolysis of breaks cellulose into individual sugar monomers which can be fermented into ethanol; esterification transforms oil derived from plants into butyl acetate, or diesel. Both reactions require the assistance of enzymes, which are often more efficient in complex solvents.

The esterification of butanol was modeled with quantum mechanics using Gaussian 09 [39]. The transition state structure was found in vacuum and a water continuum model, with reaction energies of -10.3 kcal/mol and -7.2 kcal/mol, respectively, and a barrier height of approximately 60 kcal/mol. The reaction was then modeled using CPMD enhanced with metadynamics in the vacuum phase and with explicit water solvent. The transition state geometry found using this method agrees with quantum mechanics findings, however the energy barriers do not agree when calculated with previously accepted methods to calculate it from the first passage barrier height. The geometric features and energetics are compared to gain a greater understanding of the influence of the solvent molecules on the reaction.

The hydrolysis of dimethyl ether (DME) was also modeled using quantum mechanics with Gaussian in vacuum and water. Both the catalyzed and uncatalyzed transition states were found using this method. The transition state step is characterized by the transfer of a proton from water to the DME oxygen, which agrees with our previous findings. The energy barrier for

the uncatalyzed reaction in vacuum is 75 kcal/mol. The acid-catalyzed reaction energy barrier was found to be 30 kcal/mol by the work done by Liang, et al. using CPMD [40]. We modeled our system using the same parameters as Liang, et al. in vacuum and water with CPMD enhanced with metadynamics. The reaction was also modeled using AMBER enhanced with metadynamics, and the results agree with our findings from CPMD. The energy surface with respect to the CVs shows a distinct correlation to the bonds forming and breaking. The results using both CPMD and AMBER were extended to water and select ionic liquids to compare the structural and energetic differences.

We found that CPMD simulations coupled with metadynamics using the accepted standard method in literature was not an accurate or efficient way to gain quantitative information about the reaction energy barriers for either of the reactions studied. Therefore, results presented in this chapter can be used as a qualitative insight, and development on a new method for calculating reaction properties is presented in the following chapter.

4.2 Introduction

Choice of solvent has been shown to dramatically change the efficiency of reactions taking place at enzymatic sites, specifically, those which catalyze the breakdown of biomass. Recently, ionic liquids have been of particular interest because of their unique properties. Ionic liquids (ILs) are a class of salts, composed entirely of ions, that are liquid at ambient temperature or below [41, 42]. The “ionic” shape is asymmetrical, causing a non-uniform charge distribution, reducing the cohesive energy and melting point temperature of the salt. Typical ILs are composed of an organic cation, usually an alkyl imidazolium or pyridinum or a quaternary ammonium ion, with an inorganic anion listed in Table 4.1 [42].

Table 4.1. Table taken from M. Moniruzzaman *et al.* describing ionic liquid anions.

Anion	Full name	Abbreviation
Cl^-	Chloride	[Cl]
Br^-	Bromide	[Br]
BF_4^-	Tetrafluoroborate	[BF_4]
PF_6^-	Hexafluorophosphate	[PF_6]
$(CF_3SO_2)_2N^-$	*bis*((trifluoromethyl)sulfonyl)imide	[Tf_2N]
$(FSO_2)_2N^-$	*bis*(fluorosulfonyl)imide	[FSI]
$CF_3SO_3^-$	Trifluoromethylsulphonate	[TfO]
$C_4F_9SO_3^-$	Perfluorobutylsulfonate	[NfO]
$CH_3SO_3^-$	Sulphonate	[$MeSO_3$]
$CH_3OSO_3^-$	Methylsulphate	[MS]
CH_3COO^-	Acetate	[AcO]
CF_3COO^-	Trifluoromethylacetate	[TFA]
$CF_3(CF_2)_2COO^-$	Heptafluorobutanoate	[HB]
$C_6H_5COO^-$	Benzyl acetate	[BzO]
$HOCH_2COO^-$	Glycolate	[glycolate]
$CH_3CH(OH)COO^-$	Lactate	[lactate]
$HO_2CCH_2C(CO_2H)–(OH)CH_2COO^-$	Citrate	[citrate]
NO_3^-	Nitrate	[NO_3]
$C_8H_{17}SO_4^-$	n-octylsulfate	[$OcSO_4$]
$(CN)_2N^-$	Dicyanamide	[dca]
$(C_2H_5)_2PO_4^-$	Diethylphosphate	[$(Et)_2PO_4$]

Recently, ionic liquids have gained attention for their potential use as an alternative solvent in applications like biofuels, electrochemistry, pharmaceuticals, and many others. Because ILs are non-volatile, a vacuum can be used to remove the volatile products of the reaction making them an environmentally attractive alternative solvent for traditional volatile organic compounds (VOCs) frequently used as solvents for enzymatic processes [43]. In addition, ILs have no vapor pressure, so the equilibrium of the reaction can easily be driven towards the products. ILs have shown an ability to dissolve both polar and non-polar organic compounds, inorganic compounds, and polymeric compounds [43]. Several studies have also indicated that enzymes exhibit enhanced selectivity in ILs, including substrate, regio- and enantioselectivity [42]. Enzymes have an increased thermal stability in some ILs, so bioprocesses can be performed at relatively high temperatures for biological systems, in contrast to organic solvents.

Furthermore, ILs make excellent solvents because the anion-cation combination can be uniquely engineered depending on the application of the solvent based on desired physiochemical properties such as density, viscosity, hydrophobicity, and solubility, rendering them "designer solvents" [42]. The diversity in properties that can be specifically altered allows ILs to be applicable solvents for systems that cannot be dissolved in organic solvents. Because of the unique properties observed in ILs they have a wide range of applications as environmentally friendly media for processes such as synthesis and catalysis, inorganic synthesis, separation, nanomaterial synthesis, enzymatic reactions, and biotransformations [42].

A number of enzymes have shown increased stability in certain ILs compared to traditional solvents, however they have also been shown to deactivate in ILs with strongly coordinating anions like nitrate or acetate [42]. Literature provides few details into the molecular mechanism of how ILs alter the enzyme activity, which has generated significant interest in exploring how ILs change the interactions of enzymes and substrates along with the kinetics involved [44]. Studies have also shown that concentration of the IL has a significant effect on the activity and selectivity of each individual enzyme and IL combination; some perform more efficiently in anhydrous conditions while others are more efficient in partially hydrated IL solutions [45].

Our group is particularly interested in ILs' application to catalysis for reactions involved in biomass breakdown. One of our studies explored how solvation in ILs affect the structural and dynamic features of the *Candida rugosa* Lipase 1 (CRL) enzyme, which has been studied in a variety of ILs [44]. Additionally, as described in the previous chapter, we have gained new insight into the reaction mechanism of cellulosic and hemicellulosic breakdown. We are

concerned with how ILs influence the energetics and catalytic properties of the esterification and hydrolysis reactions that are important to breaking down cellulosic and hemicellulosic biomass.

4.2.1 Esterification reactions in ionic liquids

Lipases are among the most commonly studied enzymes in ILs because they are typically used in organic solvents and are well-known to be highly tolerant to such conditions [45], making for a convenient comparison. Lipases catalyze a wide variety of reactions including esterification, transesterification, alcoholysis, aminolysis, hydrolysis, and polymerization reactions [42]. Lipases are particularly useful in catalyzing the production of biodiesel through transesterification. ILs have been shown to increase yields in some lipase-catalyzed reactions [46].

Enzyme activity has been shown to be preserved or enhanced in specific ILs but not in others [44]. For example, *Candida rugosa* Lipase 1 (CRL1) was shown to retain its activity in the IL 1-butyl-3-methyl-imidizolium hexafluorophosphate (BMIM-PF_6) but is inactive in BMIM nitrate (BMIM-NO_3) [20, 47]. Particularly, many studies have investigated lipase-catalyzed transesterification reactions and cellulose-catalyzed hydrolysis reactions.

The *Candida rugosa* lipase B (CALB) enzyme has been shown to catalyze a wide range of reactions in ILs, including transesterification [42]. To understand the molecular details of the esterification reaction, we have chosen a model reaction based on esterification reactions that have been widely studied. To precisely model the effects of solvent on the reaction at an atomic level, we chose a representative reaction that can be simulated at a high level of theory. This reaction is a simplified version of an esterification of geraniol, which is catalyzed using CALB and has been widely studied in ILs [45, 48]. The reactants are acetic acid and trans-crotyl

alcohol. CALB acts as an acid catalyst to perform the esterification resulting in a trans-crotyl acetate and water. The reaction is shown below in Figure 4.1.

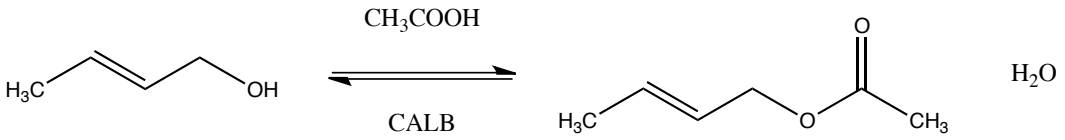

Figure 4.1. Esterification of crotyl alcohol with acetic acid and an acid catalyst to produce crotyl acetate and water.

To avoid the potential for crotyl structures to produce resonance isomers, the double bonded carbonyl groups were replaced with a single bond, and the esterification of *n*-butanol was also modeled to produce butyl acetate, a common solvent used in industry. The reaction is shown in Figure 4.2. This has been studied before by Kim, et al. [49] who investigated the synthesis of butyl acetate using the CALB enzyme for esterification in IL solutions. Their experimental results showed that imidizolium based ILs are suitable solvents for the synthesis of butyl acetate. In particular, they showed that BMIM-TfO solvated *n*-butanol expressed the highest reaction rate and conversion when compared to the organic solvent *tert*-butanol, which is conventional solvent used for this esterification process.

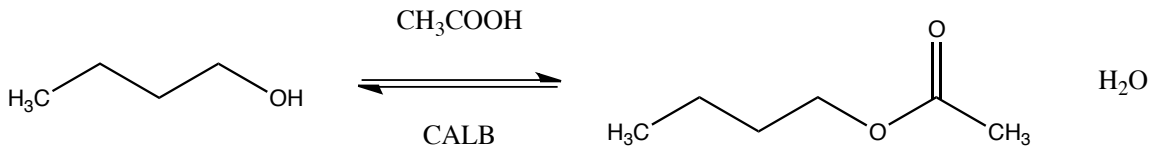

Figure 4.2. The esterification of *n*-butanol with acetic acid and an acid catalyst to produce butyl acetate and water.

Both of these reactions take place through a Fischer-Speier esterification reaction [50], in which an acid and alcohol react in the presence of an acid catalyst (CALB in this case). CALB

has shown increased product yield, activity, or conversion for esterification reactions of fatty-acids in the following ILs or IL/water mixtures BMIM-Cl was modeled as the IL system.

Barahona, et al. Studied the esterification of geraniol in BMIM-PF_6 finding that the presence of the IL increased conversion, however, decreased the reaction rate compared to the organic solvent hexane [48]. Gamba, et al. studied the conversion of soybean oil to biodiesel via transesterification finding that the *Pseudomonas cepacia* lipase in BMIM-Tf_2N was the most effective, comparing the conversion rate to that of ethanol and methanol solvents at varying concentrations [46]. Both of these studies have found that an increase in the concentration of water content decreases the reaction rate and equilibrium constants, despite the increased conversion rate.

We set out to gain more understanding of the mechanism for the esterification reaction as well as get insight into how the solvent affects the equilibrium conversion and reaction rate at the by calculating the free energy profile and comparing the structural features first with water, and then across multiple solvents using an explicit solvent model, with an overall goal of characterizing reaction kinetics as a function of solvent properties.

4.2.2 Hydrolysis reactions in ionic liquids

Researchers have aimed to find a more efficient way to derive fuel from cellulosic material for decades [9]. The hydrolysis of the ether bond, the glycosidic linkage in biomass, is the limiting step in obtaining fuels from cellulosic material due to the high cost of the enzymes required for this step in the reaction [11]. Findings presented in the previous chapter provide insight into the reaction mechanism of the hydrolysis of the glycosidic bond as well as an understanding into how solvent properties affect the energy profile of the reaction [1]. Our DFT

studies with an implicit solvation model showed that acidic solvents provide a stabilizing effect on the hydrolysis reaction involving the hydrolysis of the β-1,4 glycosidic bond. However, more can be learned from modeling the explicit solvation of a hydrolysis reaction to gain a greater understanding of the influence of solvent properties on structural dynamics and energetics of the reaction. Many studies have shown that ILs are effective at dissolving cellulose for pre-processing and aid in the hydrolysis reaction [42].

Explicit solvation is more accurate and detailed than the implicit solvation models used in quantum mechanics. Explicit solvent models use individual solvent atoms to model each individual solvent molecule in the system being studied. Implicit solvation models use a continuum model represented by a continuous field described only by macroscopic parameters, which is not accurate for short-range interactions. However, it is extremely impractical to use explicit solvation with DFT electronic structure calculations because the many degenerate states with slightly different solvent configurations in the ground state of the potential energy surface must be considered. This creates a high computational cost for a system of this size.

Dimethyl ether (DME) provides a simplified representation of the ether linkage found in cellulose and hemicellulose. The hydrolysis of the ether linkage is the same basic chemistry as the hydrolysis of the linkage found between glucose monomers, and includes all of the reacting atoms.

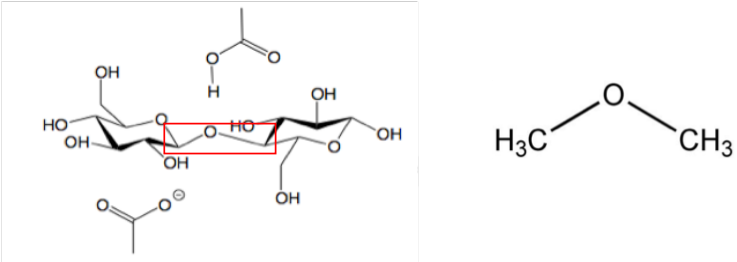

Figure 4.3. Cellobiose (left) with a box around the atoms participating in hydrolysis involves the same atoms shown in dimethyl ether (right).

Using *ab initio* molecular dynamics (AIMD) along with explicit solvation we aimed to determine which solvation properties are most favorable for ether bond hydrolysis. We extend our results from the previous chapter to include ionic liquids as solvents of interest. The catalyzed and uncatalyzed reactions are shown below in Figure 4.4. Comparing the solvent properties for the catalyzed and uncatalyzed reactions was the initial goal, by modeling the torsional angles, bond distances, and free energy surfaces as a function of solvent properties.

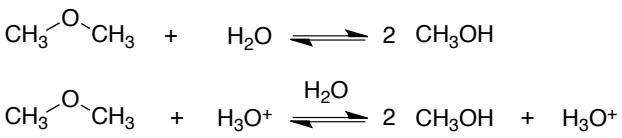

Figure 4.4. DME hydrolysis to yield two methanol molecules. The top reaction is uncatalyzed and the bottom reaction is acid-catalyzed.

Liang, et al. [40] have examined the acid-catalyzed hydrolysis of DME using *ab initio* DFT simulations. For comparison, we attempted to replicate their results and extend them further to the uncatalyzed reaction as well as model an IL in addition to water.

4.3 Methods

4.3.1 *Ab initio* molecular dynamics with metadynamics to study reactions in solvents

To accurately study hydrolysis and esterification reactions in explicit solvent models we will turn to molecular dynamics, more specifically *ab initio* molecular dynamics (AIMD). Classical molecular dynamics (MD) is one of the most commonly used methods for finding molecular details in condensed phases using explicit solvation [3]. MD uses an algorithm to solve the Newtonian equations of motion from a specified initial state using boundary conditions. The forces used to solve this equation are specified from the "force field," which do not include electronic polarization effects, therefore cannot be used to study chemical reactions. Chemical bonding is defined by a molecular topology, so classical MD cannot extend to the study of chemical reactions and we must turn to *ab initio* MD.

AIMD uses a DFT framework to solve the Kohn-Sham equation, Equation 4.1, in terms of the electron density:

Equation 4.1

$$\varepsilon_i \phi_i(r) = \left(-\frac{\hbar^2}{2m} \nabla^2 + v_s(r) \right) \phi_i(r)$$

where ε_i is the orbital energy of the orbital ϕ_i, $v_s(r)$ is the fictitious potential (Kohn-Sham potential), $\hbar$ is Plank's constant, m is the mass of the particle, and r is the distance between non-interacting particles. The Kohn-Sham equation is the Schrodinger equation of a fictitious system of non-interacting particles (electrons).

Forces obtained from the electronic structure calculations for atoms performed "on the fly" are used in AIMD to give finite-temperature dynamical trajectories. This implies it can accurately model the bonds breaking and forming that account for electronic polarization in reactions. For computationally expensive studies of chemical reactions, most commonly DFT is used to obtain the electronic energy by solving the Kohn-Sham equations. Specifically, we will use the Car-Parrinello MD (CPMD) approach [6]. The basic idea behind CPMD is to use the difference in time scale between the electronic and nuclear motion, by transforming it into a classical system. CPMD introduces fictitious dynamical variables to the degrees of freedom that explicitly track the time evolution of the coordinates. [51]

CPMD was performed with metadynamics to gain information about the potential energy surface with respect to the chosen collective variables. Metadynamics [52] applies a time-dependent bias to collective variables (CVs) that describe changes in the system. Each time step, a small Gaussian is deposited to the potential energy surface with a height w and width σ.

Equation 4.2

$$V(s(r),t) = w \sum_{t'=\tau_G,2\tau_G}^{t'<t} \prod_{i=1}^{N_{CV}} exp\left[\frac{-(s_i(r) - s_i(r(t')))^2}{2\sigma_i^2}\right]$$

In the bias potential shown by Equation 4.2, the values of the N_{CV} collective variables are defined by *s(r)*, which is a functional mapping relating the CV to the geometry of the system (not shown). The characteristic height (*W*, energy units) and width (σ*i*, CV units) define each Gaussian "hill", which is added every t_G time steps. If the simulation is performed properly [53], the cumulative bias from the simulation can be inverted to obtain the underlying free-energy surface (FES) as projected onto the CV [54]– a major advantage of the metadynamics method.

Our work section 4.3.2 and section 5 uses the well-tempered variant of metadynamics, which was recently introduced to achieve smooth convergence of the FES [81]. Within well-tempered metadynamics, the hill height, *W'*, exponentially decreases according to an adjustable parameter ΔT, which is related to the characteristic barrier heights in the system. The net effect is that for regions of phase space already visited, the additional bias potential added becomes exponentially smaller according to the following relationship:

Equation 4.3

$$W' = \omega e^{\left[-\frac{V(s,t)}{k_B \Delta T}\right]}$$

Using CPMD with DFT combined with metadynamics provides a powerful tool for exploring the energetic landscape of molecular systems in solvents by allowing the use of an explicit solvation model. However, previously accepted methods are not a reliable way to accurately calculate the energy barrier of transition states, which is needed to calculate kinetic parameters of the reactions within reasonable computational cost [51,55].

4.3.2 Modeling esterification reactions in gas, water, and ionic liquid mixtures

The characterization of the esterification reaction was first carried out in the vacuum phase with crotyl alcohol and acetic acid as reactants. DFT calculations were performed using the software package Gaussian 09 [39]. Analysis to determine geometric properties were carried out using GaussView and PLUMED [56]. Optimization calculations were performed with the complete basis set method [57-64] (CBS-QB3), which is a very accurate way to compute energies. The method is a complex energy computation involving pre-defined calculations on a specified system.

Minimum energy structures were found using geometry optimizations with the Berny algorithm [24] for each species in the products and reactants of the esterification reaction. Each of the species was optimized alone in vacuum and water with the CPCM [21, 22] implicit solvation model as well as with the ionic solvent PCM model, which uses the integral equation formalism model (IEFPCM) to perform a reaction field calculation [65]. The optimizations were performed once again with the reactant species and product species together ensuring that the energies were the same.

Transition state structures were obtained using the quasi-synchronous transit method (QST3) [25, 26]. The transition state was confirmed by ensuring that the vibrations had only one imaginary frequency, corresponding to the first order saddle point on the potential energy curve [27].

The minimum energy structures found using Gaussian 09 [39] were converted to CPMD input files. An NVE optimization was run on the system at a BLYP level of theory [66, 67] with an *ab initio* molecular dynamics simulation using CPMD. The system temperature was set at 300K with a 4 au time step. The cutoff value in the Kohn-Sham equations was set at 70 Ry. A fictitious mass was set at 800 amu. Kleinman-Bylander pseudopotentials were used [68] with the QUENCH-BO option, which converges the wavefunction at the beginning of the molecular dynamics simulation [69].

Based on the mechanism and the energies of the structures we found using Gaussian 09, we assigned CVs to real reaction coordinates, the bonds breaking and forming in the reaction, so that we could monitor energy as a function of the bond length in each case. This enabled us to assign a transition state to a particular bond forming or breaking, further confirming our findings

using transition state theory with Gaussian and observing how the explicit solvation model changes the energy and structural characteristics of the reaction.

4.3.3 Modeling hydrolysis reactions in gas, water, and ionic liquids

Building from the work done by Liang, et al. [40] we further explore the hydrolysis of DME using CPMD with metadynamics. We set out to compare a variety of solvents' effects on the reaction properties. To begin, we ran CPMD to optimize DME alone in vacuum and water. From this calculation we compared the torsional angle and bond distances between solvation states. Then we used metadynamics to manually break the C-O bond and again compare the bond distances and angles between solvation models.

The system was initially constructed and minimized using classical MD. Solvent boxes with periodic boundary conditions were added using the Packmol software package [70]. The system geometry was minimized using the Gromacs MD software [71]. A TIP3P force field model was used for water molecules and the IL force field was provided by a code written by my colleague, Vance Jaeger [72]. The DME molecule was initially optimized using Gaussian 09 [19] with DFT at the B3LYP/6-31+G(d,p) level. When the system was minimized, it was run in CPMD. A microcanonical ensemble (NVE) simulation was run at 300K with a 0.1 femtosecond timestep. The system is a 10-angstrom (Å) cubic box with one DME molecule and either one water molecule or one hydronium (H_3O^+) molecule. The water simulation contained 113 water molecules and the IL simulation contained 56 water molecules, 8 molecules of BMIM and 8 molecules of Cl. This yields a concentration of 1.66 mol/L for each of the simulations of DME.

Density functional AIMD was performed using the CPMD program [6]. Goedecker pseudopotentials [73] were used along with the Becke, Lee, Yang, and Parr (BLYP) functional

[66, 67]. The cutoff value in the Kohn-Sham equations was set at 70 Ry. A fictitious mass was set at 800 amu. The CPMD simulation was minimized using a NVE ensemble. Subsequently, the simulation was run using a canonical (NVT) ensemble at 300 K with a Nosé-Hoover thermostat [74]. Metadynamics was performed with the PLUMED package [20]. Simulations involving the breaking of the C-O bond used the bond length as the biased CV while also monitoring the torsional angle (see Figure 4.5). The hill height was set at 0.005 kcal/mol with a 0.5 Bohr Gaussian width and 30,000 MD steps.

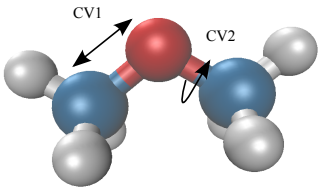

Figure 4.5. The non-reacting DME system simulated with CPMD and metadynamics. CV1 (bond length) and CV2 (torsional angle) are indicated.

For comparison we have run similar calculations using AMBER and the semi-empirical PM3 Hamiltonian which is an intermediate level of theory between the types of DFT typically used with AIMD studies and classical MD. Bonds can still break and form, but the barrier heights are known to be quite inaccurate [75] However, comparative studies within a reaction class can be performed to gain new insights about chemical properties.

The DME system was initially set up in vacuum, water, and a 50% mixture of water and BMIM-Cl. The solvated simulations were modeled using a quantum mechanics/molecular mechanics (QM/MM) technique where the atoms reacting, DME and water in this case, are modeled using a quantum or *ab initio* level of theory and the solvent is modeled using a molecular mechanics level of theory. We have modeled the atoms with PM3 and the solvent with

TIP3P [76] for water and TIP3P and the generalized AMBER force field (GAFF) [77] for BMIM-Cl. Simulations were run at 300K in an NVT ensemble using a Langevin thermostat and a damping coefficient of 10 ps^{-1} [78]. An 8 angstrom cutoff distance was implemented for non-bonded atoms and the simulation was run with a 0.2 fs timestep.

Metadynamics was used to simulate the reaction occurring by biasing the bond distances that are breaking and forming, as described in the previous section. The metadynamics parameters were a hill height of 0.20 kcal/mol, and a stride length of 1000 MD steps. The width of the Gaussians deposited was 0.1 Bohr for each of the CVs.

4.4 Results and Discussion

4.4.1 Solvation effects on the physical and chemical properties of esterification reactions

This reaction is a simplified version of an esterification of geraniol, which is catalyzed using CALB. The reactants are acetic acid and butyl alcohol. The mechanism was found from optimizing the geometry of each of the species involved in the esterification of crotyl alcohol with one of the bonds broken to find out which group the OH comes from by comparing the enthalpies of reaction. The results of this were assumed for the esterification of butanol. Based on these findings, a transition state structure was guessed and the transition state analysis described in section 4.3.2 was performed. The complete mechanism is shown in Figure 4.6.

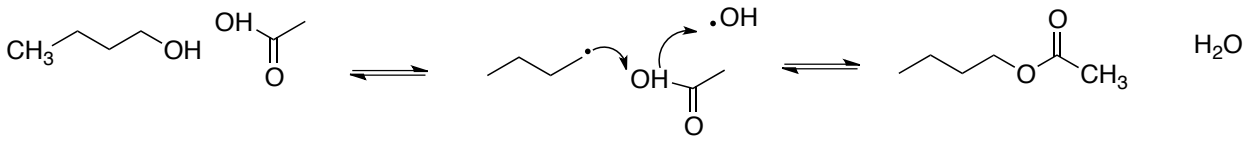

Figure 4.6. Complete mechanism of butanol esterification to butyl acetate, found using TST.

Results based on the optimized IRC structures indicate that the transition state barrier is approximately 60 kcal/mol. Results based on the CBS-QB3 optimizations of the products and reactants show that the energy of reaction is approximately -10.3 kcal/mol in vacuum and -7.2 kcal/mol in both the water and ionic solvated models.

The bond distances being broken and formed at the TS were chosen as CVs in the metadynamics run for the CPMD calculations. The CVs chosen are shown in Figure 4.7. Bonds forming are shown as solid lines and bonds breaking are shown as dashed lines.

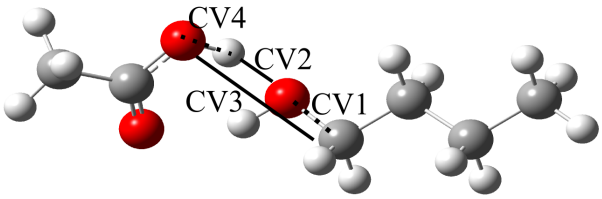

Figure 4.7. The collective variables used in the CPMD with metadynamics simulation of the esterification of butanol with acetic acid. CV1 and CV4 measure the distance between breaking bonds, indicated by dashed lines. CV2 and CV3 measure the distance between forming bonds, indicated by a solid line.

The collective variables were biased using metadynamics in the PLUMED package [56]. Initially CV1 was biased while the other three CVs were monitored. The bond distances over the time of the simulation are shown in Figure 4.8. Corresponding structures are shown on the distance plot. Using this initial calculation, we were able to observe the breaking of the C-O

bond. It can be observed that as the C-O bond breaks, the hydrogen atom spontaneously breaks from the OH group in the acetic acid molecule. However, the simulation was not able to overcome the energy barrier required to form the new C-O bond. This process agrees with the maximum transition state energy point that we found using quantum mechanics. Based on these results we biased CV4 in addition to CV1 in the subsequent simulation to overcome the energy barrier required to form the bond that results in butyl acetate.

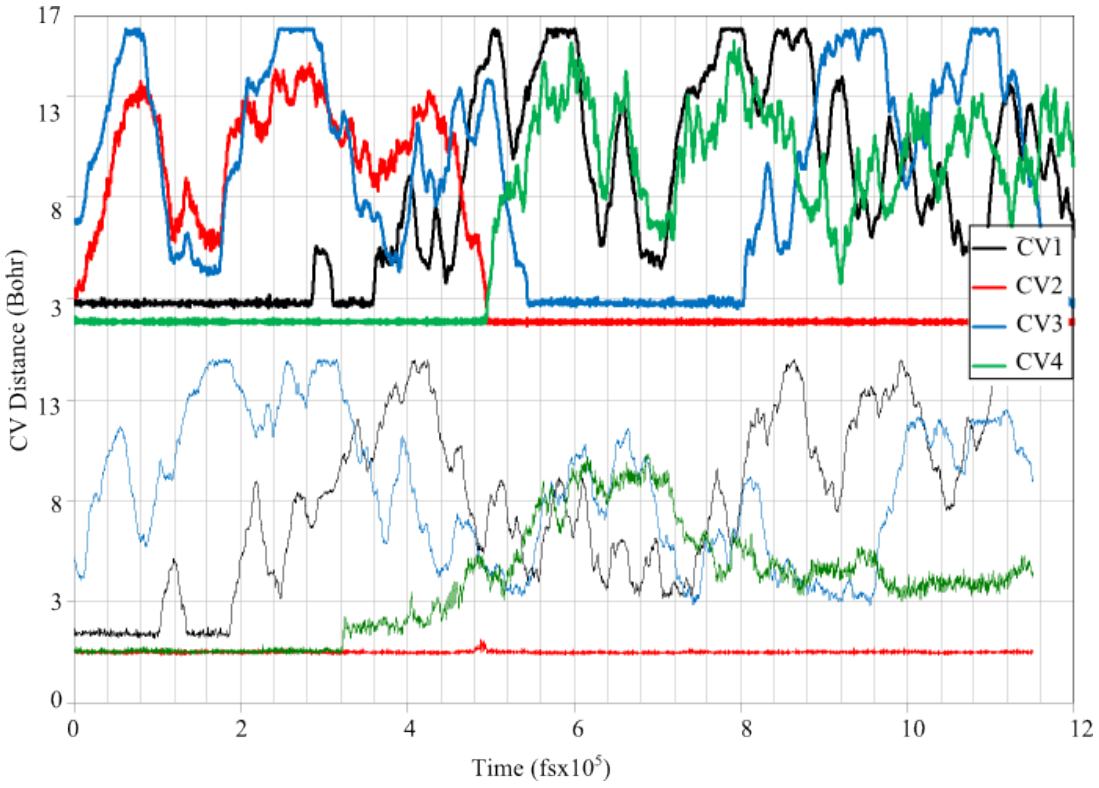

Figure 4.8. The bond distance in Bohr as a function of the time step (.002 fs) as a result of a metadynamics run with a biased CV1. CVs correspond to those shown in Figure 4.7. The top plot is in the gas phase, bottom plot is in water solvent.

When both CVs are biased the reaction is observed in the trajectory, with the products of butyl acetate and water forming. The complete free energy surface of this reaction as a function of CV1 and CV3 are shown in Figure 4.9 in the gas phase and in water. The energy wells are

labeled A, B and C. A corresponds to the reactants and C corresponds to the products. The transition state is also labeled, but is not indicated by specific characteristic across trials. According to our Gaussian calculations the saddle point, which indicates the transition state, occurs at a CV1 value of approximately 6 Bohr and a CV3 value of approximately 5 Bohr in both phases.

It can also be observed that the energy surfaces further confirm that the water solvent has a stabilizing effect on the esterification reaction. Similarly to the hydrolysis reaction, since the transition state is characterized by the transfer of an ion, the water solvent provides a stabilizing complex around the reaction, facilitating the reaction and lowering the energy barrier.

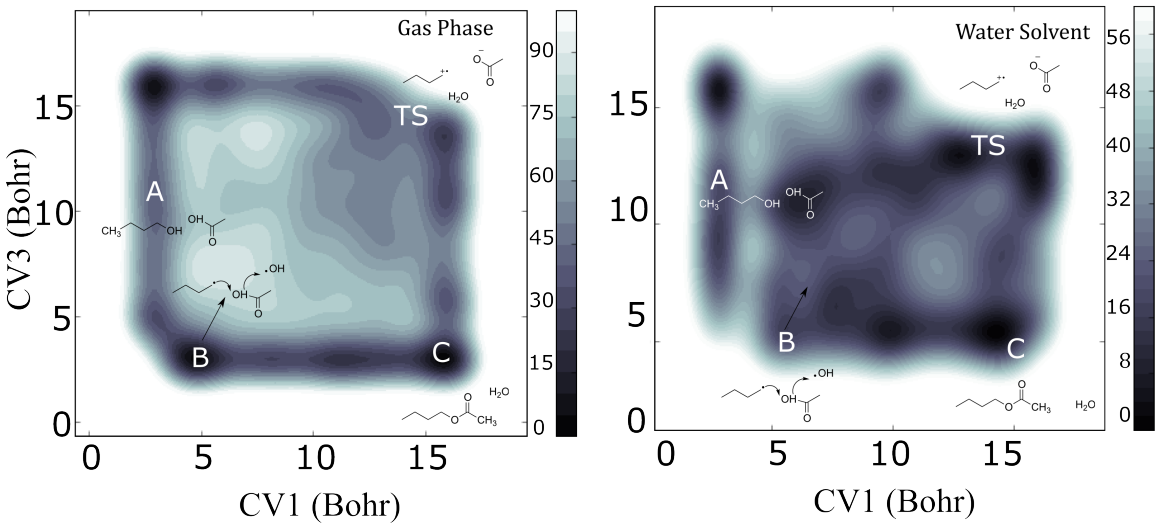

Figure 4.9. The free energy surface of the complete esterification reaction of butanol as a function of CV1 and CV3. The plots show CV1 versus CV3 with corresponding energies in the key to the right, with the reaction in the gas phase on the left and the reaction in explicit water solvent on the right. Energy wells are labeled A, B, and C, and the values are in kcal/mol. The first order saddle point indicating the transition state is labeled TS.

To examine the reliability of the methods described above, the simulations were replicated from new starting geometries. The initial input files were identical to each other, and the geometry minimization was performed five times each in vacuum and four times in water. Interestingly, the energy surfaces varied a significant amount. Not only are the quantifiable energies inconsistent, the required amount of simulation time for the reaction to occur varies widely. Full results showing the discrepancy between the trials with and the reaction and energy surface that was fully sampled is shown in Table 4.2.

Table 4.2. Table of free energy values calculated from identical starting files with random initial velocities. The table on the left is in vacuum and the table on the right is in explicit water solvent.

	Gas Phase		Water	
Trial	**Free Energy Barrier (kcal/mol)**	**Time for Reaction to Occur (fs x 103)**	**Free Energy Barrier (kcal/mol)**	**Time for Reaction to Occur (fs x 103)**
1	132.9	12.14	56.7	3.05
2	101.8	6.13	56.3	2.55
3	131.8	12.70	86.4	4.28
4	124.5	9.82	66.6	2.74
5	105.2	7.14		
Average	119.2	9.59	66.5	3.16
Standard Deviation	14.8	2.92	14.1	0.78

These results show that the free energy varies significantly between trials of identical input that is re-minimized and assigned random initial velocities. The values of these barriers are calculated using the first pass barrier height method; the simulation is monitored to continue running until the reaction occurs, determined by a limit on the specified CV. For this case we used a 4.0 Bohr limit for CV1, and then found the free energy values from that point. After truncating the HILLS file, the free energy values are calculated, and the difference between the maximum free energy value and the minimum point is determined to be the first passage barrier height. These results were replicated 5 times in vacuum and 4 times in water, and as you can see vary widely. In vacuum the average barrier height is taken to be 119.2 kcal/mol with a standard deviation of 14.8, and in water it is 66.5 kcal/mol with a standard deviation of 14.1 and the range is ~30 kcal/mol between the highest and lowest barrier heights in both vacuum and water.

In addition to the discrepancy between the values calculated using this method, the method itself is flawed in correctly identifying the important energy characteristics like first

order saddle points of chemical reactions. This explains the discrepancy of 80 kcal/mol between our calculation from CPMD and our transition state calculation using the CBS-QB3 basis set.

The amount of free energy space that is explored is directly related to the amount of time the simulation is allowed to run indicating it is non-random (Poisson) and can be affected by metadynamics parameters. As you can see in Table 4.2, the amount of time required for the reaction to happen is also variable. The standard deviation in vacuum is 2900 fs and in water it is 800 fs. When the relationship between the amount of time it takes for the reaction to happen and the energy barrier calculated with the first passage method, there is a direct linear relationship, which is very distinct in vacuum. This relationship is shown in Figure 4.10. This trend further indicates that the energy barrier calculated with this method is dependent on the simulation parameters, and is not a reliable way to calculate energy barriers for chemical reactions.

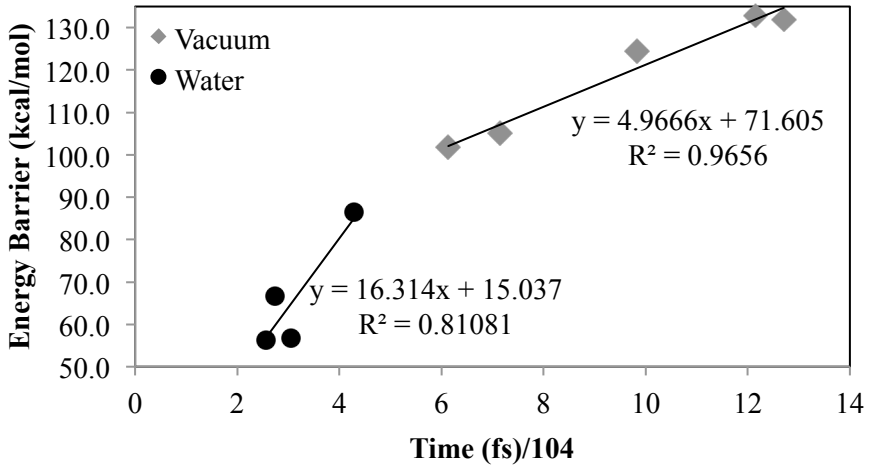

Figure 4.10. The energy barrier calculated with the first passage method compared to the amount of time required for the reaction to occur in the simulation.

4.4.2 Solvation effects on physical properties of hydrolysis reactions to break down biomass

4.4.2.1 Analysis of the uncatalyzed DME hydrolysis reaction

A 2D constrained geometry optimization was performed on the DME system using the bond distances between the proton and oxygen of the water and the carbon and oxygen of the DME molecule. This was performed at the PM3 level of theory. The three-dimensional energy surface for the uncatalyzed reaction found using QM is shown in Figure 4.11. Using the saddle points from this energy surface, we were able to input a transition state structure guess for the transition state analysis using QST3.

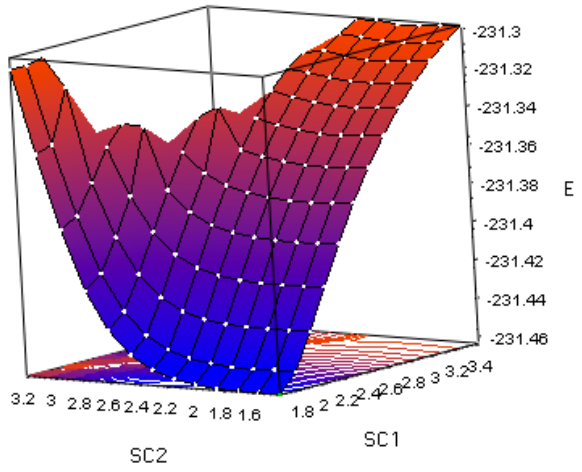

Figure 4.11. Full energy surface for the uncatalyzed hydrolysis of DME. SC1 is the distance between the proton and oxygen, SC2 is the distance between the carbon and oxygen that are separating.

A transition state analysis was done on DME hydrolysis using DFT at the B3LYP/6-31+G(d,p) level of theory, as was described in section 3.4.2. We found that DME is hydrolyzed in a single step reaction. The transition state step is characterized by the transfer of a proton from water to the DME oxygen, which agrees with our results from section 3.4.2. The energy barrier

for this step is 75 kcal/mol. The acid-catalyzed reaction energy barrier was found to be 30 kcal/mol by the work done by Liang, et al. using CPMD [40].

4.4.2.2 Solvation effects of energetic and geometric features of dimethyl ether hydrolysis

Using QM/MM studies on the DME system with AMBER we have observed the uncatalyzed reaction occurring using metadynamics with the bond distances as CVs. The bond distances were biased at a sigma value of 0.1 kcal/mol. Using this methodology we were able to observe the reaction occurring in all 3 phases tested. The free energy surface with the corresponding CVs for the vacuum simulation are shown in Figure 4.12. Based on this simulation we can estimate that the barrier height in the vacuum phase is approximately 100 kcal/mol. The same reaction is observed in water and in BMIM-Cl. The results of these simulations show the same trend in energy surface. The estimated barrier heights are approximately 85 kcal/mol and 89 kcal/mol in water and BMIM-Cl, respectively. It should be noted that PM6 does not provide accurate binding energies, but we can observe the trend in energy for the type of solvation model used.

From Figure 4.12 we can observe that the largest energy barrier corresponds to the smaller CV2 values and is spread across a range of CV1 values. This is promising because it shows that the transition state is likely located along the coordinate where the hydrogen from the water is bonding to the nucleophilic CH_3 group, which agrees with our findings using DFT.

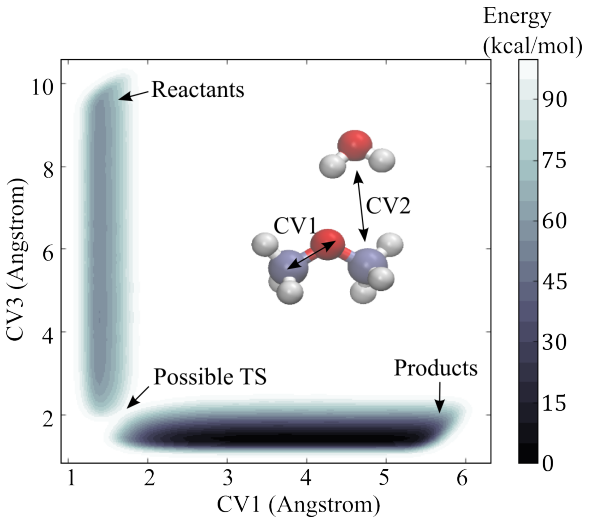

Figure 4.12. The free energy surface as a function of the DME hydrolysis reaction coordinates.

4.5 Conclusions

Based on our quantum mechanical calculations using Gaussian 09 [39] we observed that the transfer of the proton onto oxygen is the rate limiting step for the hydrolysis of DME, which is in agreement with our findings from chapter 3 and the shape of the energy surface we found using MD in the AMBER package. We also found that the rate-limiting step for esterification reactions is the transfer of the acidic oxygen nucleophile to the carbon atom to form an ester. We used metadynamics to further explore the energy surfaces of each of these reactions in multiple solvents.

We successfully set up both of our reaction systems in vacuum to observe energetic and structural trends, and extended these systems to water using CPMD. We compared our preliminary CPMD results for DME hydrolysis with results found using classical MD. Using our results for DME with AMBER, we ran the same metadynamics calculations in CPMD to attempt get an accurate value for the barrier heights. However, what we found from our esterification study was that using these established methods, CPMD and metadynamics alone are not an

adequate way to learn about chemical reaction properties because they do not correctly calculate barrier heights of reactions (this is further discussed in Chapter 5). Typical practice to find barrier heights in literature has been to use metadynamics to force the reaction to happen, the way we did with the bond distances as our biased CVs. Typically this is replicated 3-5 times with new random initial velocities, and then the average first passage barrier height resulting from those calculations is taken to be the reaction activation energy. Our calculations show significant variance using this method, both in vacuum and with explicit solvent, with a standard deviation of 14.8 kcal/mol in vacuum and 14.1 kcal/mol in water. Based on these revelations about using this method, we determined that a new way to model reactions in explicit solvents is needed. Chapter 5 discusses a method called MetaRates that we have developed based on the work of Tiwary et al. [79].

Despite the inaccuracy of the energy barrier values, we are able to get a clearer picture of how the solvent affects the reaction taking place at the atomic level with a lower level of theory, like PM6. If we can use the MetaRates approach to make the simulations both computationally efficient and quantitatively reliable, we can model reactions in complex ionic liquids as well as reactions taking place within enzyme active sites (i.e. the full hydrolysis mechanism of cellulase) to get quantitative results related to the reaction kinetics.

5 A New Approach for Investigating Reaction Dynamics and Rates with Ab Initio Calculations

Copyright: This chapter was written in the form of a manuscript, which has been submitted to a scientific journal for publication and is under review at the time of the submission of this dissertation.

Acknowledgement of coauthors: Dr. Pratyush Tiwary and Dr. Jim Pfaendtner supervised this project and assisted in the analysis and production of figures. The following is the full text of our manuscript.

5.1 Abstract

Herein we demonstrate a convenient approach to systematically investigate chemical reaction dynamics using the Metadynamics (MetaD) family of enhanced sampling methods. Using a symmetric S_N2 reaction as a model system we applied 'MetaRates', a theoretical framework based on acceleration factors, to quantitatively estimate the rate of reaction from biased and unbiased simulations. A systematic study of the algorithm and its application to chemical reactions was performed by sampling over 5000 independent reaction events. Additionally, we quantitatively reweighed exhaustive free-energy calculations to obtain the reaction potential energy surface and showed that MetaRates works to effectively determine Arrhenius-like activation energies. Exact agreement with unbiased high-temperature kinetics is also shown. The feasibility of using the approach on actual ab initio molecular dynamics calculations is then presented by using Car-Parrinello MD+MetaD to sample the same reaction using only 10-20 calculations of the rare event. Owing to the ease of use and comparatively low-

cost of computation, the approach has extensive potential applications for catalysis, combustion, pyrolysis and enzymology.

5.2 Introduction

Computational investigation of chemical kinetics is an active area of research from industrial and theoretical perspectives. Having matured over the past half-century, there is now a vast array of choices for using computer simulations to make an estimate of various kinetic quantities for virtually every type of reaction, and obtain useful insight with high spatio-temporal resolution into the various factors at play. Broadly speaking, a convenient taxonomy for computational kinetics is to subdivide whether a particular approach uses 1) a time-dependent or stochastic sampling (e.g., molecular dynamics (MD) or Monte Carlo) or 2) calculations on a static potential energy surface (PES) combined with a theoretical model for dynamics of reaction barrier crossing. In the former category, a canonical example is the Car-Parrinello MD (CPMD) method (a subset of ab initio MD or AIMD), which in principle permits direct simulation of chemical bond-breaking and forming events due to the use of ab initio potentials. In the latter category harmonic transition state theory (TST) is one of many theoretical approaches for prediction of detailed reaction rates. An exhaustive compilation of the full canon of methods and their merits and limitations is far beyond the scope of this article.[80] However, it is generally true that the system size and complexity of the chemical transformation govern the choice of method and ultimate accuracy of the model used to estimate barrier heights or rates.

Among available choices in the class of methods using MD, it is almost universally true that additional methods (so-called “enhanced sampling”) are also needed to overcome the high barriers associated with chemical reactions and significantly extend the typically accessible simulation times through unbiased MD. In particular with chemical reactions, these time scale

limitations in MD and especially AIMD can be severe, e.g., even a barrier height of a few kcal/mol represents an effectively unattainable goal for modern AIMD approaches. One example, out of many possible choices, of a popular enhanced sampling approach is the metadynamics (MetaD) method [52, 81]. It involves building a time-dependent bias as a function of carefully chosen low-dimensional collective variables, that discourages the system from getting trapped in low lying energy states, and encourages exploration of configuration space in an enhanced and controllable manner. In the long time limit, the net result of a typical metadynamics simulation is an estimate of the true underlying free energy distribution [82]. The MetaD approach has found wide application in studies of chemical reactions [83] owing to the simplicity of implementation, the intuitive nature of the algorithm, and the huge amount of freely available codes. Detailed reviews of the application of MetaD to chemical reactions are available elsewhere coving a huge number of use cases and practical tips for implementation [55, 84, 85]. However, because AIMD calculations remain quite expensive, the application of MetaD to chemical and biochemical reactions is generally still in the qualitative or semi-quantitative realm. An additional challenge has been the difficulty in interpreting the thermodynamic results obtained from a MetaD calculation in light of kinetic observables such as rates or activation energies.

A recent extension of the MetaD approach by Tiwary and Parrinello [79] demonstrated for a range of systems from model potentials to actual protein-ligand systems [79, 86, 87], that it is possible to recover unbiased dynamic information from biased MetaD simulations. In a later work it was also shown as to how could one establish the reliability of the dynamics so constructed [88]. This algorithm, which we refer to herein as “MetaRates”, is based on the assignment of acceleration factors [89], calculable from the instantaneous value of the applied

MetaD biasing potential, at a particular moment in time. As described below, the rate of the rare event is calculated post-priori using the bias potential and biased trajectory. The method is an exciting addition to the MetaD family of methods and there are already indications that application to chemical reactions is possible (using semi-empirical potentials) [90], However, several challenges remain in using this approach for studying chemical reactions. First, the temperature dependence of the observables determined by MetaRates needs to be fully understood. Second, demonstration on a real AIMD simulation (and the concomitant tuning of the required MetaD parameters) is important to show feasibility for applications that do not involve semi-empirical force fields.

In this paper we demonstrate how to apply MetaRates to problems in chemical kinetics. We have carried out over 5000 MetaD simulations of a model reaction using the PM6 Hamiltonian and used the results to optimize and demonstrate feasibility for application of MetaRates using CPMD simulations. We also demonstrate how to extract the PES for a reaction from a converged MetaD simulation, thereby enabling a direct comparison between the activation energy and the temperature dependence of the MetaRates output with the Arrhenius relationship. Another key result of this work is a quantitative insight into the transition state ensemble for chemical reactions [91], and a clear illustration of how tricky it can be to accurately sample this ensemble even with enhanced sampling, thereby possibly introducing very significant errors into rate calculations based on direct barrier energy calculations. Instead, with a careful application of the approach of Ref. 5, validated via rigorous statistical analysis[88], one can obtain direct and statistically accurate insight into the dynamics of chemical reactions. The methodology illustrated here should be applicable to a vast array of complex reactions for

instance in enzymes and condensed phase reactions solvents, without the need to use semi-empirical force fields.

5.3 Methods

This study used the symmetric S_N2 nucleophilic substitution reaction of chloromethane and a chlorine: $CH_3Cl + Cl^- \Leftrightarrow CH_3Cl + Cl^-$ as a model system due to its simplicity and low energy barrier. Three different methods were used to study the reaction. AIMD simulations were carried out with semi-empirical PM6 [92] in the AMBER program [93]. Static energy calculations were carried out with PM6 and the Gaussian 09 [39] program. Finally, CPMD [6] simulations were carried out at the BLYP [68] level of theory. Further details of the MD and enhanced sampling are provided below.

MD simulations with PM6/AMBER were performed with a 1 fs timestep on an isolated system of one methylchloride and chloride ion (-1 total charge). A cutoff of 1.0 nm was used and the system was verified to be energy conserving in the microcanonical (NVE) ensemble prior to biased simulations that used a thermostat. Biased NVT MetaD simulations used a Langevin thermostat with collision frequency of 10 ps^{-1}. For the biased simulations, harmonic restraints were used on the C-Cl distances to keep the heavy atoms within 0.6 nm.

The plane-wave DFT system was based on the Becke exchange[94] and Lee-Yang-Parr correlation (BLYP) [95] functional used with the Kleinman-Bylander norm conserving pseudo-potentials[68] for chlorine, carbon, and hydrogen atoms. MD simulations with BLYP/CPMD [6] were performed on the same isolated system used for PM6 MD. We used a timestep of 4 a.u. (~0.1 fs) and a fictitious electron mass of 800 a.u. The simulations were performed in the canonical ensemble (NVT) in an isolated 1.0 nm cubic box. Temperature was controlled with

Nose−Hoover thermostats [96, 97] of chain length three with frequencies for the atoms and fictitious electron kinetic energy of 3132 and 10000 cm^{-1}, respectively. All parameters were obtained and/or verified with NVE simulations, which were also used to verify that they were energy conserving.

Simulations were biased with the MetaD framework, which biases the system to selectively explore phase space along coarse descriptors of rare events – so called collective variables (CVs, *s*). Additional terms are added to the system's potential energy in the form of Gaussian "hills". The hills are dimensional in the number of biased CVs, and the added terms result in applied force on the atomistic coordinates used to define the CVs. A very large number of review articles and tutorials are available covering the theory, proof and use of MetaD simulations. Our biased MetaD simulations were performed with PLUMED [20] and used the well-tempered variant[98]. The so-called biasfactor (γ) was set to 9. The two distances between carbon and chlorine were used as biasing CVs with Gaussian widths of σ=0.0025 nm. Unless noted otherwise, the initial hills height was 0.3 kJ/mol. A variety of "stride" (τ) values were used, as discussed below, representing the time between subsequent deposition of MetaD Gaussian "hills".

The simulations were post-processed with the MetaRates algorithm.[79] Briefly, in the MetaRates approach the transition time over a barrier is calculated through the use of acceleration factors derived from a MetaD simulation. An individual transition time (t^{eff}) is calculated as:

$$t^{eff} = \sum_{i=0}^{N} \Delta t_i^{MetaD} e^{\beta V_{Bias}(s,t)}$$

where the "MetaD time step (Δt)" refers to the timestep in the simulation and V_{Bias} is the instantaneous value of the MetaD potential at the i^{th} point in the transition trajectory. The transition rate (ν) is an ensemble average computed from a set of individual transition times, provided the set follows the conditions described below. Note that for the MetaRates algorithm to work, the stride has to be larger than the typical time a successful reactive trajectory spends in the barrier region while crossing over from one basin to another. Furthermore, the biased CV must be able to distinguish between all relevant stable states in the system. The p-value analysis [88] allows to check *a posteriori* whether these two important requirements were satisfied or not. In this work, this summation is carried out over a single escape trajectory (rare event) and many individual escape times are collected from independent simulations. According to the recommendation of Salvalagio et al. [88], the ensemble of escape times are verified to be uncorrelated and to follow a Poisson distribution, as quantified through a Kolmogorov-Smirnov analysis with p-value >0.05 signifying that the distribution is Poisson. The inverse of the mean escape time $<\nu>$ is deemed the "rate" of a particular process. Further details of the implementation and tuning of parameters are discussed below.

5.4 Results and Discussion

5.4.1 Sampling the S_N2 reaction rates on a semiempirical energy surface.

Using the PM6 Hamiltonian, we studied the temperature dependence of the chemical reaction rates determined with MetaRates. Additionally, the use of a simple gas-phase chemical reaction with one characteristic low-barrier height event permitted us to also use high temperatures as a means to accelerate sampling and provide an independent validation of the accelerated sampling at low-temperature by comparing against an Arrhenius type fit. In order to ensure that the MetaRates data were independent, each simulation was stopped after a single

event was observed and a new simulation was initiated with random initial velocities. The reactions were sampled in increments of 75K between 300 and 600K. Initially, we used extremely long (for typical MetaD simulations) wait periods between deposition of consecutive Gaussian hills (τ) in the bias potential (also commonly called the "stride"). Values of 20 and 100 (ps between hill addition) for τ were used. With only one exception (due to very long simulation times at the lowest temperature), a minimum of 100 escape times were collected for each of the biased simulations.

Table 5.1 provides the mean escape time data for the S_N2 reactions. Owing to the simplicity of the reaction, there is a direct correspondence to the chemical reaction rate and the escape from the reactant free-energy basin. A total of 5812 reaction events were sampled. For each of the groups of simulations, we verified that the population of escape times followed a Poisson distribution by verifying the p-value obtained from the Kolmogorov-Smirnov test was greater than 0.05 according to established procedure.

Table 5.1. Mean escape time data for the S_N2 reaction.

T(K)	ν (ns)[A]	p_{KS}	Number of events	τ[B] (ps)
600	0.457(8.6e-4)	0.909	631	20
600[C]	0.675(1.3e-3)	0.968	480	20
525	0.915(2.0e-3)	0.942	393	20
450	3.21(8.6e-3)	0.995	270	20
375	10.3(3.8e-3)	0.940	122	20
300	184(7.9e-1)	0.906	133	20
300[C]	157(6.0e-1)	0.971	142	20
600	0.486(8.6e-4)	0.765	720	100
525	1.21(2.2e-3)	0.250	697	100
450	3.43(1.1e-2)	0.987	197	100
375	17.7(5.4e-2)	0.898	226	100
300	167(8.8e-1)	0.320	78	100
375	35.2	N/A	1	
450	3.77(4.2e-2)	0.975	13	
525	1.46(1.3e-2)	0.872	34	
600	0.640(2.9e-3)	0.700	77	
900	0.116(2.3e-4)	0.473	432	
1200	0.0427(5.9e-5)	0.386	1170	

[A] The mean escape time (a.k.a., reaction time). The value in parentheses indicates the estimate of the standard error of the mean (SEM) through the sampling protocol described in the SI. The SEM is not calculated for data sets smaller than 50 in size.

[B] The time between successive deposition of Gaussian hills in the biased simulations. If blank, then the simulation is unbiased

[C] Repeated simulations (discussion in text)

As seen in Table 5.1, the p-values are always well above the recommended threshold of 0.05, and as such the rates are clearly uncorrelated and could be considered a Poisson process. Figure S1 shows two of the cumulative distribution functions (CDF) for the 20 and 100 ps simulations at two temperatures. The CDFs follow the expected behavior and the choice of τ (comparing 20 and 100 ps) does not change the observed result. Table 5.1 shows a repeated a campaign of simulations for the τ=20 ps/300K simulation and found quite similar results. Interestingly, the average of the two simulation series is nearly identical to the τ=100 ps/300K

simulation. When comparisons were available (T ≥ 450K), the rates from the unbiased simulations also show near perfect agreement with the biased simulations. We also repeated the τ=20 ps/600K simulations and found that the τ=100 ps value fell between the duplicate values. Using all of the data in Table 5.1, we verified Arrhenius-like behavior was found and calculated a range of possible activation energies (Figure 5.1). As shown in Table 1, the symbol size is larger than the standard error from the uncertainty analysis.

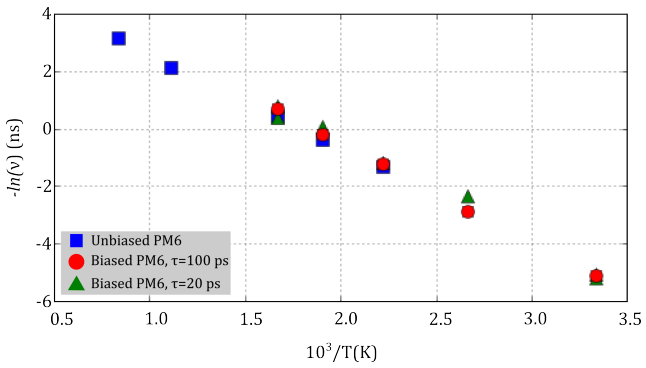

Figure 5.1. Arrhenius plot of the inverse temperature in Kelvin versus the log of the reaction rate, ν. Blue squares are unbiased MD, red circles are biased MD with 100 ps stride, and green triangles are biased MD with a 20 ps stride

As seen in Figure 5.1, the temperature dependence of the data can be modeled with an Arrhenius type relationship, and therefore an overall activation energy (E_A) value can be established. Depending on which subsets of data are selected (e.g., unbiased vs. biased vs. all data), a range of E_A values from ~6.7-7.0 kcal/mol is obtained. We next set out to determine the relationship of activation energy determined in this manner to the more common definitions of activation energy obtained through analysis of a reaction's potential energy landscape.

5.4.2 The S_N2 free- and potential energy landscapes.

We first established a well-converged MetaD bias potential (on the PM6 energy surface) using the two relevant distances (both of C-Cl distances) as CVs with well-tempered MetaD. We ran the simulation for ~250 ns of biased simulation time at 300K. Over 30,000 barrier crossings were observed owing to the strongly diffusive behavior that was observed once the MetaD transient period was over (i.e., the well-tempered MetaD bias potential was filled and smoothed). This also clearly indicates strong coupling between the two CVs we chose and the actual reaction coordinate of the system. Since there is no independent reference surface with which to compare, we show in Figure S2 the free-energy surfaces (FES) at 50%, 75% and 100% of the total simulation length, which clearly establishes that the MetaD simulation is converged.

If Figure 5.1 is interpreted through the lens of classical harmonic transition state theory (TST) [31], wherein the entropy goes strictly into the prefactor [99], the slope of the plot would strictly be related to a activation enthalpy closely corresponding to the potential energy barrier, and not the free-energy barrier. Accordingly, we sought to use the biased simulations to construct an overall potential energy surface (PES). After tabulating PM6 energies (and corresponding CV values) over the whole MD trajectory, we used a reweighing technique[79, 100] to construct a three-dimensional FES: F(CV1,CV2,E_{PM6}), where the CVs are the same distanced biased in the MetaD simulation. Following this, the ensemble average value of the PE at each point in CV space was trivially obtained by taking a Boltzmann-weighted average over all PE values at each point:

$$\left\langle E_{PM6}(CV1,CV2)\right\rangle = \sum E_{PM6} e^{-\beta F_i(CV1,CV2,E_{PM6})}$$

Figure 5.2A shows the PES obtained in this manner. As expected, the PES obtained from reweighing is not perfectly smooth in the transition state region, owing to fact that even in a perfect randomly diffusive system this area of phase space is only sampled in proportion to the relative amount of phase space it occupies. To understand the accuracy of this PES, we used constrained geometry optimizations and calculated the PM6 energies using Gaussian09[39] on a grid of ~ 400 CV1/CV2 values. The PES obtained from Gaussian in this manner is shown in Figure 5.2B and the difference between the two plots is given in panel C. The agreement between the PES is remarkable – on average the values in the difference plot are just a fraction of a kcal/mol, in both the basins and TS region. Moreover, the barrier height on this surface is also on the order of 7 kcal/mol, which is precisely the estimate of E_A found above.

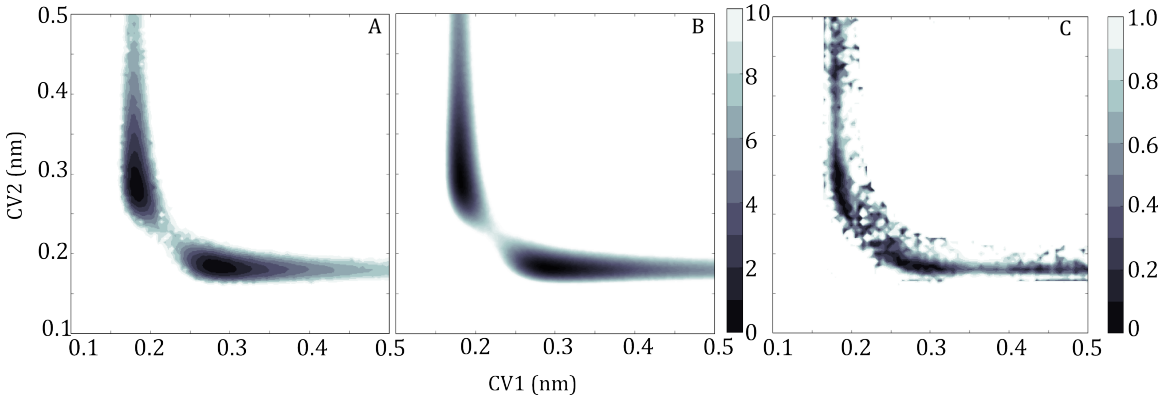

Figure 5.2. Panel A shows the PES with respect to the carbon-chlorine bond distances for the chloromethane reaction constructed from reweighted MetaD and panel B shows the PES constructed from Gaussian calculations. The energy scale (kcal/mol) for both is shown to the right of Panel B. Panel C shows the absolute difference between the two energies in panels A and B (in kcal/mol).

In summary, the results presented in Figure 5.1 and Figure 5.2 offer support of two new concepts in the use of biased ab initio calculations to study chemical reaction kinetics. First, the quantitative features of the entire PES can be abstracted through careful reweighing analysis.

Analysis of complex reacting systems, e.g., combustion and pyrolysis [101], often makes use of features of the PE landscape, not the FE landscape, so this approach holds great promise for facilitating automated analysis [102]. Second, the "MetaRates" approach is well suited for the quantification of meaningful estimates of activation energies for complex reactions. As previously discussed [85, 103], an approach often used is to estimate the transition state free energy for a reaction by truncating the MetaD bias potential exactly at the moment of 1st passage and taking an average of several 1st passage estimates. The inherent and rather arguable assumption here is that this distribution of energies is the same as the transition state ensemble (TSE) [91]. This is an alluring idea, which we also sought to explore given the huge amount of individual events sampled in this study (Table 5.1). Using the τ=20 ps/300K simulation (containing 133 events) we systematically tabulated all of the 'barrier heights' that would be obtained in this manner and compared to the final FES obtained using the exhaustive and well-converged MetaD calculation (Figure 3). Comparing the histogram of 1st passage "barrier heights" (Figure 5.3A) with the final FES (Figure 5.3B) clearly shows that even for a simple reaction it is impossible to attempt to estimate a barrier height in this manner. As seen in Figure 5.4, we also studied the properties of this distribution as they relate to an adjustable MetaD input parameter: the Gaussian deposition rate. It can be seen that the properties of this distribution shift significantly higher as the Gaussian stride is increased. This further underscores the risk in attempting to extract meaningful information from just a few 1st passage "barrier heights" for a given system. Thus even if one assumes that the first passage distribution is the same as the true TSE, this approach is not practical for an accurate estimation of rates. We also find it very interesting that the barrier height distribution is skewed with a pronounced fat left tail. In future

work we will explore the implications of this on the entropy of the transition state, and the potential errors in using a harmonic TST type model to approximate this entropy.

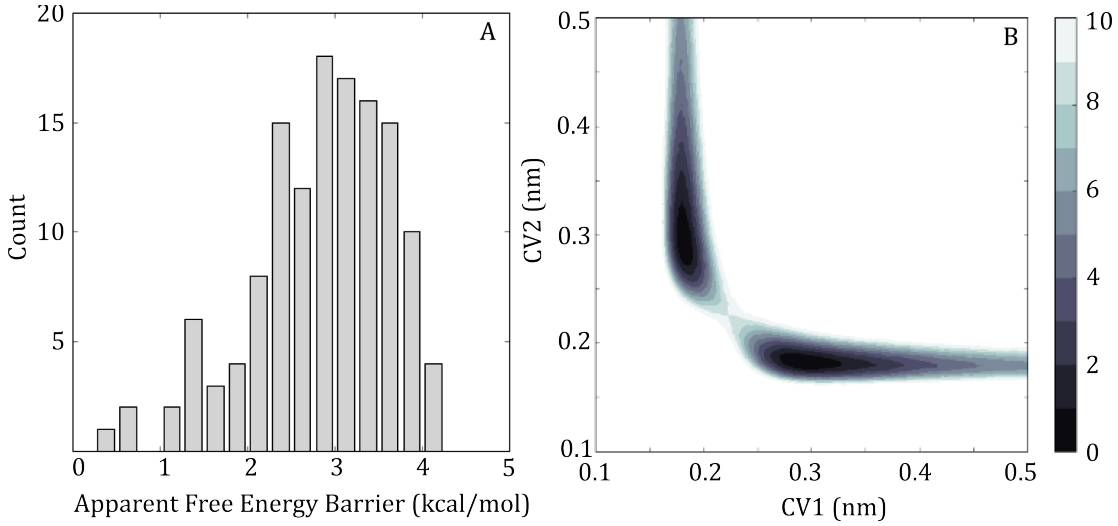

Figure 5.3. Panel A: Histogram of the 1st passage barrier heights determined from ~130 PM6 MetaD simulations initiated from different starting conditions. energy barrier calculated by metadynamics alone shown to the left. Panel B: Fully converged FES for the same reaction (kcal/mol).

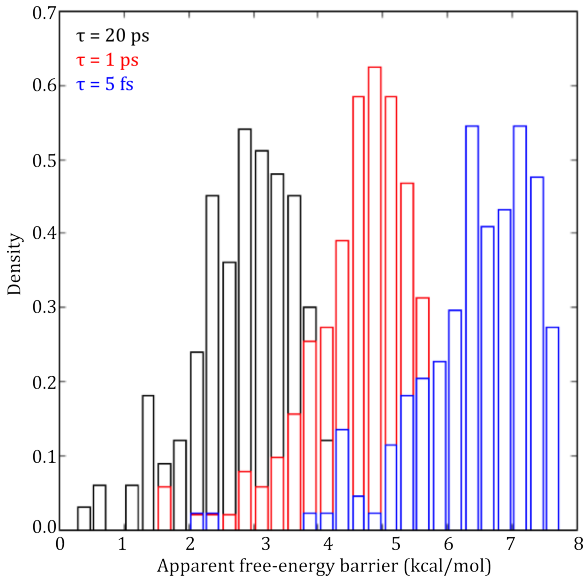

Figure 5.4. Changing the MetaD stride rate significantly changes the achieved distribution of apparent free-energy barriers. Further information about these simulations is provided in the next section and table S1. The number of sampled events for each histogram are τ=20ps (133), τ=1ps (211), τ=5fs (478).

5.4.3 Optimizing the MetaRates parameters for use with CPMD.

Although the analysis performed above using the PM6 semi-empirical model offered the opportunity to collect thousands of statistics and understand the meaning of key attributes of the rates approach. We wanted to demonstrate the feasibility of performing a similar analysis with a more accurate MD Hamiltonian. To ensure realistic studies in the future, which of course will have only a tiny fraction of the data collected above, will still be meaningful. Prior to attempting this study with ab initio MD, we studied the properties of the obtained rates and rates distributions as a function of the Gaussian deposition rate using PM6. The values used in the study above (τ=20 ps and 100 ps) were useful to ensure the results were devoid of systematic error from the sampling and construction of the bias potential. However, those values are orders of magnitude higher than what can be achieved in applications of MetaD using ab initio dynamics. We systematically varied the value of τ all the way down to the value of τ=5 fs/hill

(an extremely fast energy rate of deposition). The average rates and p-values are shown in Table S1, where it can be seen that even a value of 25 fs/hill can still result in p-value above 0.05. At the fastest deposition rate, the p-value drops precipitously and it can be seen that the rate obtained is also significantly different than the other rates. Based on the results in the table, we selected a value of τ=50 fs/hill for the ab initio MD study.

5.4.4 Sampling the SN2 reaction rates with CPMD

Similarly to the AMBER analysis, the reaction was simulated in CPMD with well-tempered metadynamics with a BLYP plane-wave basis set [104] to determine the reaction rates and their temperature dependence. Again, we used a simple gas-phase SN2 reaction of chloromethane and a chloride atom. After a single reaction event was observed, a new simulation was initiated with random initial velocities. Reactions were sampled at 300K with increments up to 500K for easy comparison to the PM6 calculations. Owing to the required decrease in simulation time step (~0.1 fs vs 1.0 fs) and the significant increase in computational cost of BLYP vs PM6, the calculations were significantly slower and we collected ~20 reaction events at each temperature. Other metadynamics parameters were also matched exactly to the PM6 calculations with the following exception – the initial hill height was increased from ~0.07 to ~0.35 kcal/mol. This change was only made in order to facilitate a reasonable total computational cost and test calculations (on the PM6 system) showed identical results on the increased hill height at the same MetaD stride value.

The mean escape times at the three temperatures (300K = 175 ns, 400K = 8.1 ns , 500K = 1.5 ns) also follow expected temperature dependence and suggest an activation energy for this PES of 7.1 kcal/mol. Note that the agreement between the BLYP and PM6 energy surfaces is completely coincidental and does not suggest anything significant about the MetaRates approach,

or about the nature of the PM6 and BLYP models. The KS p-values for the three data sets also indicate uncorrelated events (300K: p=0.61 , 400K: p=0.57, 500K: p=0.13) and demonstrate that on a drastically reduced sample size (N=20 for each temp) that the method still works as expected. To further study the potential impacts of the sample size we randomly selected (with replacement) sets of the 300K data at values of N=(5, 10, 15). At each new sample size 1000 random trials were performed and the results show that at N=5 the mean escape time was substantively different and at N=10 and 15 the mean escape time was identical to the full N=20 value (full data in Table S2).

Similar to the PM6 case, we also established a converged bias potential on the BLYP surface using identical CVs, the two C-Cl bond distances, with well-tempered MetaD. Simulations were run at 300K until convergence was achieved. The MetaD simulation was reweighed in the manner described above to obtain the potential energy surface (using the Kohn-Sham energies tabulated during the CPMD simulation) and the comparison of the PES and FES for the biased ab initio dynamics are given in Figure 5.5. As in the PM6 case, the barrier height on the PES is quite close to the Arrhenius activation energy suggested by MetaRates.

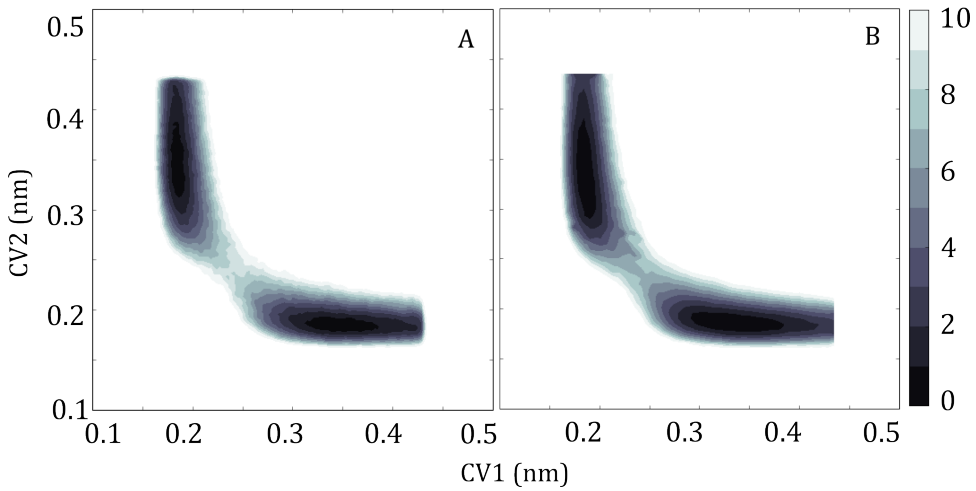

Figure 5.5. The FES is shown in panel A, rendered using CPMD with BLYP. The reweighted PES is shown in panel B. The energy scale is in kcal/mol

5.5 Conclusion

The MetaRates approach is a powerful addition to the set of computational tools offered by the Metadynamics framework. Herein, we have comprehensively shown that the method can accurately be applied to the study of chemical reactions with ab initio dynamics approaches. Several important conclusions arise from the analysis presented in this work. First, the temperature dependence of the mean rates calculated with MetaRates offer the ability to access meaningful Arrhenius-like activation energies. Exhaustive comparison between the E_A values and the underlying reaction PES confirmed this for semiempirical PM6 and identical conclusions were obtained on a computationally expensive CPMD/BLYP calculation in which far fewer statistics were available. Second, the general practice of obtaining several "1st passage" free-energy barriers from standard MetaD simulations is not a reliable way to obtain activation energies. When enough statistics are available, it is quite clear these free-energies have a wide distribution. More importantly, is that distributions of the "1st passage" barriers are strongly dependent on the MetaD parameters themselves. Finally, the equivalence of the potential energy

surface (and thus the origin of the MetaRates activation energies) with the underlying static PES obtained in the traditional manner of constrained geometry optimizations suggests that this approach can offer quantitative rate estimation for simulations of complex chemical transformations.

Specifically, we envision this approach to be particularly useful in cases such as heterogeneous or enzymatic catalysis. In these cases it is often possible to use a biasing technique like MetaD to observe several reaction events [85]. Application of MetaRates, for example to a complicated QM/MM model of an enzyme or a surface reaction with nearest-neighbor effects, is now possible. Since MetaRates still works even with aggressive energy deposition rates (c.f., Table S1) we expect that the computational cost of this approach will not be so burdensome as to prevent widespread use with realistic systems governed by accurate MD potentials.

Acknowledgement

This work was accomplished with support of AFOSR Grant FA9550-12-1-0133 monitored by Dr. Chiping Li and NSF award CBET-1150596. This work was facilitated through the use of computational, storage, and networking infrastructure provided by the Hyak supercomputer system, supported in part by the University of Washington.

Supporting Information. Brief discussions and figures verifying the methodology used in this paper can be found in the Supporting Information Appendix.

6 Significance and Contributions of this Research

6.1 Increasing the efficiency of the degradation of cellulosic materials to produce biofuels

Although the mechanism to break down cellulose has been widely studied using a broad range of methods we are the first group, to our knowledge, to elicit the precise reaction mechanism of the hydrolysis of β-1,4 glycosidic bonds described in chapter 3. The information gained about the reaction energetics, as well as how solvents affect the reaction provides valuable information for enzyme design. Results previously published on the mechanism of hydrolysis using glycoside hydrolase lacked details of the transition state and implied that the mechanism was concerted. Using intrinsic reaction coordinates we were able to correspond the exact bond distances to the relative energy associated with the reaction process. Additionally, we were able to compare the uncatalyzed reaction mechanism to the acid-catalyzed mechanism, which is has not been studied before. Our findings support previous studies that investigated the order of events for glycoside hydrolase enzymes using QMM/MM.

Additionally, we were the first group to observe distinct trends by the addition of a solvent model to the hydrolysis reaction. This has provided motivation for subsequent studies by showing a promising result for yielding new, more efficient pre-treatment methods that assist in the hydrolysis of cellulose. This result provides information that may be helpful in directing mutagenesis experiments on enzymes to alter the active site depending on the dielectric constant of the solvent being used.

Our work done in chapter 4 is largely motivated by the distinct trend reported in section 3.4.3. Work in our group has focused on the interaction of ionic liquids with enzymes used in biomass processing. My colleagues have observed the influence of ILs on the structure and

dynamics of GH enzymes [105] and Lipase enzymes [44] using molecular dynamics and found some detail on the molecular interactions at the active site are influenced by solvents. As such, we modeled the reaction mechanisms used in hydrolysis and esterification using explicit solvation and a high level of theory. To our knowledge, this has never been done for either of the reactions we are using as representative systems. An understanding of the exact interactions between solvent molecules and molecules involved in the reaction, along with how the energetics of the reaction is affected by these interactions, will provide valuable insight into engineering solvents to help make pre-treatment processes for biomass breakdown more efficient. It may also give more insight into the mechanisms that enzymes use and how they are influenced by the solvent molecules surrounding them. The results found for the energetics can be tested experimentally using advanced calorimetry techniques. Results from this study coupled with results found from my colleagues on IL interaction with lipase and xylanase enzymes will provide future researchers with valuable methodology to study other important reactions as well as an analysis that will facilitate in further investigation of biomass breakdown for their use as a fuel source.

The MetaRates approach that we have demonstrated can be used on chemical reactions is a powerful addition to the metadynamics set of enhanced sampling tools – especially those studies using ab initio dynamics. We showed that it can be used to study chemical reactions with AIMD, using the temperature dependence of the rate values. We also determined that the accepted method of using first passage free energy barriers from metadynamics calculations is not an accurate way to calculate energy barriers for chemical reactions. With this knowledge, the MetaRates approach will be useful in cases where enzymatic catalysis is being studied.

MetaRates can accurately model an enzyme or surface with the QM/MM level of theory, with a relatively low computational cost.

7 Bibliography

[1] K. L. Fleming and J. Pfaendtner, "Characterizing the Catalyzed Hydrolysis of β-1,4 Glycosidic Bonds Using Density Functional Theory," *The Journal of Physical Chemistry A,* vol. 117, pp. 14200-14208, 2013/12/27 2013.

[2] W. F. van Gunsteren and H. J. Berendsen, "Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry," *Angewandte Chemie International Edition in English,* vol. 29, pp. 992-1023, 1990.

[3] M. Karplus and J. A. McCammon, "Molecular dynamics simulations of biomolecules," *Nature Structural & Molecular Biology,* vol. 9, pp. 646-652, 2002.

[4] S. A. Adcock and J. A. McCammon, "Molecular dynamics: survey of methods for simulating the activity of proteins," *Chemical Reviews,* vol. 106, pp. 1589-1615, 2006.

[5] M. Orio, D. A. Pantazis, and F. Neese, "Density functional theory," *Photosynthesis research,* vol. 102, pp. 443-453, 2009.

[6] R. Car and M. Parrinello, "Unified approach for molecular dynamics and density-functional theory," *Physical Review Letters,* vol. 55, p. 2471, 1985.

[7] T. Searchinger, R. Heimlich, R. A. Houghton, F. Dong, A. Elobeid, J. Fabiosa*, et al.*, "Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change," *Science,* vol. 319, pp. 1238-1240, 2008.

[8] Y. Sun and J. Cheng, "Hydrolysis of lignocellulosic materials for ethanol production: a review," *Bioresource Technology,* vol. 83, pp. 1-11, 2002.

[9] G. W. Huber, S. Iborra, and A. Corma, "Synthesis of Transportation Fuels from Biomass; Chemistry, Catalysts, and Engineering," *Chemical Reviews,* vol. 106, pp. 4044-4098, 2006/09/01 2006.

[10] J. Fargione, J. Hill, D. Tilman, S. Polasky, and P. Hawthorne, "Land Clearing and the Biofuel Carbon Debt," *Science,* vol. 319, pp. 1235-1238, February 29, 2008 2008.

[11] T. Collins, C. Gerday, and G. Feller, "Xylanases, xylanase families and extremophilic xylanases," *FEMS Microbiology Reviews,* vol. 29, pp. 3-23, 2005.

[12] D. J. Vocadlo and G. J. Davies, "Mechanistic insights into glycosidase chemistry," *Current Opinion in Chemical Biology,* vol. 12, pp. 539-555, 2008.

[13] W. Deng, R. Lobo, W. Setthapun, S. T. Christensen, J. W. Elam, and C. L. Marshall, "Oxidative Hydrolysis of Cellobiose to Glucose," *Catalysis Letters,* vol. 141, pp. 498-506, 2011.

[14] D. E. Koshland, "Stereochemistry and the Mechanism of Enzymatic Reactions," *Biological Reviews,* vol. 28, pp. 416-436, 1953.

[15] J. D. McCarter and G. Stephen Withers, "Mechanisms of enzymatic glycoside hydrolysis," *Current Opinion in Structural Biology,* vol. 4, pp. 885-892, 1994.

[16] G. J. Davies, L. Mackenzie, A. Varrot, M. Dauter, A. M. Brzozowski, M. Schulein*, et al.*, "Snapshots along an Enzymatic Reaction Coordinate; Analysis of a Retaining Beta-Glycoside Hydrolase," *Biochemistry,* vol. 37, pp. 11707-11713, 1998/08/01 1998.

[17] X. Biarnes, A. Ardevol, J. Iglesias-Fernandez, A. Planas, and C. Rovira, "Catalytic Itinerary in 1,3-1,4-Beta-Glucanase Unraveled by QM/MM Metadynamics. Charge Is Not Yet Fully Developed at the Oxocarbenium Ion-like Transition State," *Journal of the American Chemical Society,* vol. 133, pp. 20301-20309, 2011/12/21 2011.

[18] H. B. Mayes and L. J. Broadbelt, "Unraveling the Reactions that Unravel Cellulose," *Journal of Physical Chemistry A,* vol. 116, pp. 7098-7106, 2012/07/05 2012.

[19] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, *et al.*, "Gaussian 09, Revision A.1," in *Gaussian, Inc.*, ed. Wallingford CT, 2009.
[20] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, *et al.*, "PLUMED: A portable plugin for free-energy calculations with molecular dynamics," *Computer Physics Communications,* vol. 180, pp. 1961-1972, 2009.
[21] V. Barone and M. Cossi, "Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model," *Journal of Physical Chemistry A,* vol. 102, pp. 1995-2001, 1998/03/01 1998.
[22] M. Cossi, N. Rega, G. Scalmani, and V. Barone, "Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model," *Journal of Computational Chemistry,* vol. 24, pp. 669-681, 2003.
[23] A. Becke, "Density-functional thermochemistry. III. The role of exact exchange," *Journal of Chemical Physics,* vol. 98, pp. 5648-5652, 1993.
[24] X. Li and M. J. Frisch, "Energy-Represented Direct Inversion in the Iterative Subspace within a Hybrid Geometry Optimization Method," *Journal of Chemical Theory and Computation,* vol. 2, pp. 835-839, 2006/05/01 2006.
[25] C. Peng and H. B. Schlegel, "Combining Synchronous Transit and Quasi-Newton Methods for Finding Transition States," *Israel Journal of Chemistry,* vol. 33, 2003.
[26] C. Peng, P. Y. Ayala, H. B. Schlegel, and M. J. Frisch, "Using redundant internal coordinates to optimize equilibrium geometries and transition states," *Journal of Computational Chemistry,* vol. 17, 1996.
[27] H. B. Schlegel, "Geometry optimization," *Wiley Interdisciplinary Reviews: Computational Molecular Science,* vol. 1, pp. 790-809, 2011.
[28] K. Fukui, "The path of chemical reactions - the IRC approach," *Accounts of Chemical Research,* vol. 14, pp. 363-368, 1981.
[29] J. Pfaendtner, X. Yu, and L. Broadbelt, "The 1-D hindered rotor approximation," *Theoretical Chemistry Accounts,* vol. 118, pp. 881-898, 2007.
[30] S. A. P. Lenz, J. L. Kellie, and S. D. Wetmore, "Glycosidic Bond Cleavage in Deoxynucleotides: Effects of Solvent and the DNA Phosphate Backbone in the Computational Model," *The Journal of Physical Chemistry B,* vol. 116, pp. 14275-14284, 2012/12/13 2012.
[31] J. W. Moore and R. G. Pearson, "Kinetics and Mechanism," in *Kinetics and Mechanism*, ed: John Wiley & Sons, Inc, 1981.
[32] X. Biarnes, A. Ardevol, A. Planas, C. Rovira, A. Laio, and M. Parrinello, "The Conformational Free Energy Landscape of Beta-d-Glucopyranose. Implications for Substrate Preactivation in Beta-Glucoside Hydrolases," *Journal of the American Chemical Society,* vol. 129, pp. 10686-10693, 2007/09/01 2007.
[33] X. Biarnes, J. Nieto, A. Planas, and C. Rovira, "Substrate Distortion in the Michaelis Complex of Bacillus 1,3,-1,4-Beta-Glucanase: Insight from First Principles Molecular Dynamics Simulations," *Journal of Biological Chemistry,* vol. 281, pp. 1432-1441, January 20, 2006 2006.
[34] I. J. Barker, L. Petersen, and P. J. Reilly, "Mechanism of Xylobiose Hydrolysis by GH43 Beta-Xylosidase," *Journal of Physical Chemistry B,* vol. 114, pp. 15389-15393, 2010/11/25 2010.

[35] L. Petersen, A. Ardevol, C. Rovira, and P. J. Reilly, "Molecular mechanism of the glycosylation step catalyzed by Golgi alpha-mannosidase II: a QM/MM metadynamics investigation," *Journal of the American Chemical Society,* vol. 132, pp. 8291-8300, 2010.
[36] C. B. Barnett, K. A. Wilkinson, and K. J. Naidoo, "Molecular details from computational reaction dynamics for the cellobiohydrolase I glycosylation reaction," *Journal of the American Chemical Society,* vol. 133, pp. 19474-82, Dec 7 2011.
[37] J. Liu, X. Wang, and D. Xu, "QM/MM Study on the Catalytic Mechanism of Cellulose Hydrolysis Catalyzed by Cellulase Cel5A from Acidothermus cellulolyticus," *Journal of Physical Chemistry B,* vol. 114, pp. 1462-1470, 2010/01/28 2009.
[38] L. J. Broadbelt and J. Pfaendtner, "Lexicography of kinetic modeling of complex reaction networks," *AIChE Journal,* vol. 51, pp. 2112-2121, 2005.
[39] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, *et al.*, "Gaussian 09, Revision B.01," ed. Wallingford CT, 2009.
[40] X. Liang, A. Montoya, and B. S. Haynes, "Molecular dynamics study of acid-catalyzed hydrolysis of dimethyl ether in aqueous solution," *The Journal of Physical Chemistry B,* vol. 115, pp. 8199-8206, 2011.
[41] C. S. P. Cinzia Chiappe, "Computational studies on organic reactivity in ionic liquids," *Physical Chemistry Chemical Physics,* vol. 15, pp. 412-423, 2013.
[42] M. Moniruzzaman, K. Nakashima, N. Kamiya, and M. Goto, "Recent advances of enzymatic reactions in ionic liquids," *Biochemical Engineering Journal,* vol. 48, pp. 295-314, 2010.
[43] A. M. J. Joel L. Kaar, Jason A. Bernerich, Roger Moulton, Alan J. Russel, "Impact of Ionic Liquid Physical Properties on Lipase Activity and Stability," *Journal of the American Chemical Society,* vol. 125, pp. 4125-4131, September 16, 2002 2003.
[44] P. R. Burney and J. Pfaendtner, "Structural and Dynamic Features of Candida rugosa Lipase 1 in Water, Octane, Toluene, and Ionic Liquids BMIM-PF6 and BMIM-NO3," *The Journal of Physical Chemistry B,* vol. 117, pp. 2662-2670, 2013.
[45] F. van Rantwijk and R. A. Sheldon, "Biocatalysis in ionic liquids," *Chemical Reviews,* vol. 107, pp. 2757-2785, 2007.
[46] M. Gamba, A. A. Lapis, and J. Dupont, "Supported ionic liquid enzymatic catalysis for the production of biodiesel," *Advanced Synthesis & Catalysis,* vol. 350, pp. 160-164, 2008.
[47] M. Sureshkumar and C.-K. Lee, "Biocatalytic reactions in hydrophobic ionic liquids," *Journal of Molecular Catalysis B: Enzymatic,* vol. 60, pp. 1-12, 2009.
[48] D. Barahona, P. H. Pfromm, and M. E. Rezac, "Effect of water activity on the lipase catalyzed esterification of geraniol in ionic liquid [bmim] PF6," *Biotechnology and Bioengineering,* vol. 93, pp. 318-324, 2006.
[49] H. Kim and Y.-M. Koo, "Effects of physicochemical properties of ionic liquids on butyl acetate synthesis using Candida antarctica lipase B," *Korean Journal of Chemical Engineering,* vol. 29, pp. 1610-1614, 2012/11/01 2012.
[50] E. Fischer and A. Speier, "Darstellung der ester," in *Untersuchungen aus Verschiedenen Gebieten*, ed: Springer, 1924, pp. 285-291.
[51] J. Hutter, "Lecture Notes Introduction to Ab Initio Molecular Dynamics," ed, 2002.
[52] A. Laio and M. Parrinello, "Escaping free-energy minima," *Proceedings of the National Academy of Sciences,* vol. 99, pp. 12562-12566, 2002.

[53] A. Laio, A. Rodriguez-Fortea, F. L. Gervasio, M. Ceccarelli, and M. Parrinello, "Assessing the accuracy of metadynamics," *The Journal of Physical Chemistry B,* vol. 109, pp. 6714-6721, 2005.
[54] G. Bussi, A. Laio, and M. Parrinello, "Equilibrium free energies from nonequilibrium metadynamics," *Physical review letters,* vol. 96, p. 090601, 2006.
[55] A. Laio and F. L. Gervasio, "Metadynamics: A method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science," *Reports on Progress in Physics,* vol. 71, p. 126601, 2008.
[56] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri*, et al.*, "PLUMED: a portable plugin for free-energy calculations with molecular dynamics," ed, 2009.
[57] M. R. Nyden and G. Petersson, "Complete basis set correlation energies. I. The asymptotic convergence of pair natural orbital expansions," *The Journal of Chemical Physics,* vol. 75, pp. 1843-1862, 1981.
[58] G. Petersson, A. Bennett, T. G. Tensfeldt, M. A. Al‚ÄêLaham, W. A. Shirley, and J. Mantzaris, "A complete basis set model chemistry. I. The total energies of closed‚Äêshell atoms and hydrides of the first‚Äêrow elements," *The Journal of Chemical Physics,* vol. 89, pp. 2193-2218, 1988.
[59] G. Petersson and M. A. Al‚ÄêLaham, "A complete basis set model chemistry. II. Open‚Äêshell systems and the total energies of the first‚Äêrow atoms," *The Journal of Chemical Physics,* vol. 94, pp. 6081-6090, 1991.
[60] G. Petersson, T. G. Tensfeldt, and J. Montgomery Jr, "A complete basis set model chemistry. III. The complete basis set‚Äêquadratic configuration interaction family of methods," *The Journal of Chemical Physics,* vol. 94, pp. 6091-6101, 1991.
[61] J. W. Ochterski, G. A. Petersson, and J. A. Montgomery Jr, "A complete basis set model chemistry. V. Extensions to six or more heavy atoms," *The Journal of Chemical Physics,* vol. 104, pp. 2598-2619, 1996.
[62] J. Montgomery Jr, J. Ochterski, and G. Petersson, "A complete basis set model chemistry. IV. An improved atomic pair natural orbital method," *The Journal of Chemical Physics,* vol. 101, pp. 5900-5909, 1994.
[63] J. A. Montgomery Jr, M. J. Frisch, J. W. Ochterski, and G. A. Petersson, "A complete basis set model chemistry. VI. Use of density functional geometries and frequencies," *The Journal of Chemical Physics,* vol. 110, pp. 2822-2827, 1999.
[64] J. A. Montgomery Jr, M. J. Frisch, J. W. Ochterski, and G. A. Petersson, "A complete basis set model chemistry. VII. Use of the minimum population localization method," *The Journal of Chemical Physics,* vol. 112, pp. 6532-6542, 2000.
[65] G. Scalmani and M. J. Frisch, "Continuous surface charge polarizable continuum models of solvation. I. General formalism," *The Journal of Chemical Physics,* vol. 132, p. 114110, 2010.
[66] A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior," *Physical Review A,* vol. 38, p. 3098, 1988.
[67] C. Lee, W. Yang, and R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density," *Physical Review B,* vol. 37, p. 785, 1988.
[68] L. Kleinman and D. Bylander, "Efficacious form for model pseudopotentials," *Physical Review Letters,* vol. 48, p. 1425, 1982.

[69] J. Hutter, "Car-Parrinello Molecular Dynamics-An Electronic Structure and Molecular Dynamics Program," *CPMD Manual,* 2000.
[70] L. Martinez, R. Andrade, E. G. Birgin, and J. M. Martinnez, "Packmol: A package for building initial configurations for molecular dynamics simulations," *Journal of Computational Chemistry,* vol. 30, pp. 2157-2164, 2009.
[71] B. K. Hess, C.; van der Spoel, D.; Linhal, E., "GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation," *Journal of Chemical Theory and Computation,* vol. 4, pp. 435-447, 2008.
[72] K. Sprenger, V. W. Jaeger, and J. Pfaendtner, "The General AMBER Force Field (GAFF) Can Accurately Predict Thermodynamic and Transport Properties of Many Ionic Liquids," *The Journal of Physical Chemistry B,* vol. 119, pp. 5882-5895, 2015.
[73] S. Goedecker, M. Teter, and J. r. Hutter, "Separable dual-space Gaussian pseudopotentials," *Physical Review B,* vol. 54, p. 1703, 1996.
[74] S. Nosé, "A unified formulation of the constant temperature molecular dynamics methods," *The Journal of Chemical Physics,* vol. 81, pp. 511-519, 1984.
[75] J. Cioslowski, *Quantum-mechanical prediction of thermochemical data*: Springer, 2001.
[76] E. Neria, S. Fischer, and M. Karplus, "Simulation of activation free energies in molecular systems," *The Journal of Chemical Physics,* vol. 105, pp. 1902-1921, 1996.
[77] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," *Journal of Computational Chemistry,* vol. 25, pp. 1157-1174, 2004.
[78] X. Wu and B. R. Brooks, "Self-guided Langevin dynamics simulation method," *Chemical Physics Letters,* vol. 381, pp. 512-518, 2003.
[79] P. Tiwary and M. Parrinello, "From Metadynamics to Dynamics," *Physical Review Letters,* vol. 111, p. 230602, 12/03/ 2013.
[80] S. J. Klippenstein, V. S. Pande, and D. G. Truhlar, "Chemical kinetics and mechanisms of complex systems: A perspective on recent theoretical advances," *Journal of the American Chemical Society,* vol. 136, pp. 528-546, 2014.
[81] A. B. Barducci, M; Parrinello M., "Metadynamics," *WIREs Comput. Mol. Sci.,* vol. 1, pp. 826-843, 2011.
[82] J. F. Dama, M. Parrinello, and G. A. Voth, "Well-Tempered Metadynamics Converges Asymptotically," *Physical Review Letters,* vol. 112, p. 240602, 06/18/ 2014.
[83] M. Iannuzzi, A. Laio, and M. Parrinello, "Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics," *Physical Review Letters,* vol. 90, p. 238302, 2003.
[84] B. Ensing, M. De Vivo, Z. Liu, P. Moore, and M. L. Klein, "Metadynamics as a Tool for Exploring Free Energy Landscapes of Chemical Reactions," *Accounts of Chemical Research* vol. 39, pp. 73-81, 2006/02/01 2005.
[85] S. Zheng and J. Pfaendtner, "Enhanced sampling of chemical and biochemical reactions with metadynamics," *Molecular Simulation,* vol. 41, pp. 55-72, 2015.
[86] P. Tiwary, V. Limongelli, M. Salvalaglio, and M. Parrinello, "Kinetics of protein–ligand unbinding: Predicting pathways, rates, and rate-limiting steps," *Proceedings of the National Academy of Sciences,* vol. 112, pp. E386-E391, 2015.
[87] P. Tiwary, J. Mondal, J. A. Morrone, and B. Berne, "Role of water and steric constraints in the kinetics of cavity–ligand unbinding," *Proceedings of the National Academy of Sciences,* vol. 112, pp. 12015-12019, 2015.

[88] M. Salvalaglio, P. Tiwary, and M. Parrinello, "Assessing the reliability of the dynamics reconstructed from metadynamics," *Journal of Chemical Theory and Computation,* vol. 10, pp. 1420-1425, 2014.
[89] A. F. Voter, "A method for accelerating the molecular dynamics simulation of infrequent events," *The Journal of Chemical Physics,* vol. 106, pp. 4665-4677, 1997.
[90] K. M. Bal and E. C. Neyts, "Merging Metadynamics into Hyperdynamics: Accelerated Molecular Simulations Reaching Time Scales from Microseconds to Seconds," *Journal of Chemical Theory and Computation,* vol. 11, pp. 4545-4554, 2015/10/13 2015.
[91] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, "Transition path sampling: Throwing ropes over rough mountain passes, in the dark," *Annual Review of Physical Chemistry,* vol. 53, pp. 291-318, 2002.
[92] J. P. Stewart, "Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements," *Journal of Molecular Modeling,* vol. 13, pp. 1173-1213, 2007/12/01 2007.
[93] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. Deolt*, et al.*, "AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules," *Computer Physics Communications,* vol. 91, pp. 1-41, 1995.
[94] A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic-behavior," *Physical Review A,* vol. 38, pp. 3098-3100, Sep 15 1988.
[95] C. Lee, W. Yang, and R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density," *Physical Review B,* vol. 37, pp. 785-789, 1988.
[96] S. Nose, "A unified formulation of the constant temperature molecular-dynamics methods," *Journal of Chemical Physics,* vol. 81, pp. 511-519, 1984.
[97] W. G. Hoover, "Canonical dynamics - equilibrium phase-space distributions," *Physical Review A,* vol. 31, pp. 1695-1697, 1985.
[98] A. Barducci, G. Bussi, and M. Parrinello, "Well-tempered metadynamics: A smoothly converging and tunable free-energy method," *Physical review letters,* vol. 100, p. 020603, 2008.
[99] G. H. Vineyard, "Frequency factors and isotope effects in solid state rate processes," *Journal of Physics and Chemistry of Solids,* vol. 3, pp. 121-127, 1957.
[100] P. Tiwary and M. Parrinello, "A time-independent free energy estimator for metadynamics," *Journal of Physical Chemistry B,* vol. 119, pp. 736-742, 2014.
[101] H. Wang and M. Frenklach, "A detailed kinetic modeling study of aromatics formation in laminar premixed acetylene and ethylene flames," *Combustion and flame,* vol. 110, pp. 173-221, 1997.
[102] S. Zheng and J. Pfaendtner, "Car–Parrinello Molecular Dynamics+ Metadynamics Study of High-Temperature Methanol Oxidation Reactions Using Generic Collective Variables," *Journal of Physical Chemistry C,* vol. 118, pp. 10764-10770, 2014.
[103] L. Alessandro and L. G. Francesco, "Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science," *Reports on Progress in Physics,* vol. 71, p. 126601, 2008.
[104] S. Goedecker, M. Teter, and J. Hutter, "Separable dual-space Gaussian pseudopotentials," *Physical Review B,* vol. 54, p. 1703, 1996.

[105] V. W. Jaeger and J. Pfaendtner, "Structure, Dynamics, and Activity of Xylanase Solvated in Binary Mixtures of Ionic Liquid and Water," *ACS chemical biology,* vol. 8, pp. 1179-1186, 2013.

II Supporting Information for 'A New Method for Investigating Reaction Kinetics using Metadynamics'

This supplement includes details of the simulations employed in the chapter 5 and plots that are further illustrative of the ideas presented here.

II.I Error analysis

For every set of simulations (i.e, one row in any of the tables) the ensemble of transition times was used for a stochastic sampling procedure to assess the error. The data were downsampled by randomly selecting (with replacement) subsets of the data. The subset size was either 50% of the total sample size or 100 samples (whichever was smaller). 1000 different subsets were sampled for each simulation. If p-value for a subset was < 0.05 it was rejected and replaced with another subset. The error values reported in the tables are the standard error for the 1000 different ν values obtained in this manner. The true standard deviation of the data set can be trivially obtained by multiplying any of the standard deviations by $\sqrt{1000}$. The rate of randomly selecting a subset with a p-value < 0.05 was very small. Of the 28,000 subsets sampled in this manner there were fewer than 400 subsets rejected. For the study conducted with CPMD, the subset size was varied to study the error estimates obtained with different subsets (since the data sets were significantly smaller).

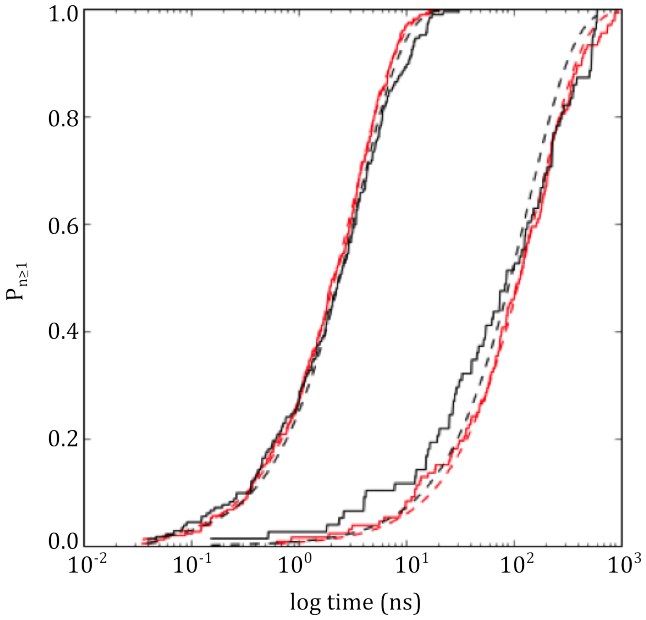

Figure S 1. Comparison of cumulative distribution functions for 450K (left curves) and 300K (right curves) simulations. The data are shown in solid lines for two MetaD stride values (black = 100ps / hill and red = 20 ps/hill). The dashed lines are the analytical CDFs obtained from the mean escape time (rate) of the simulation data.

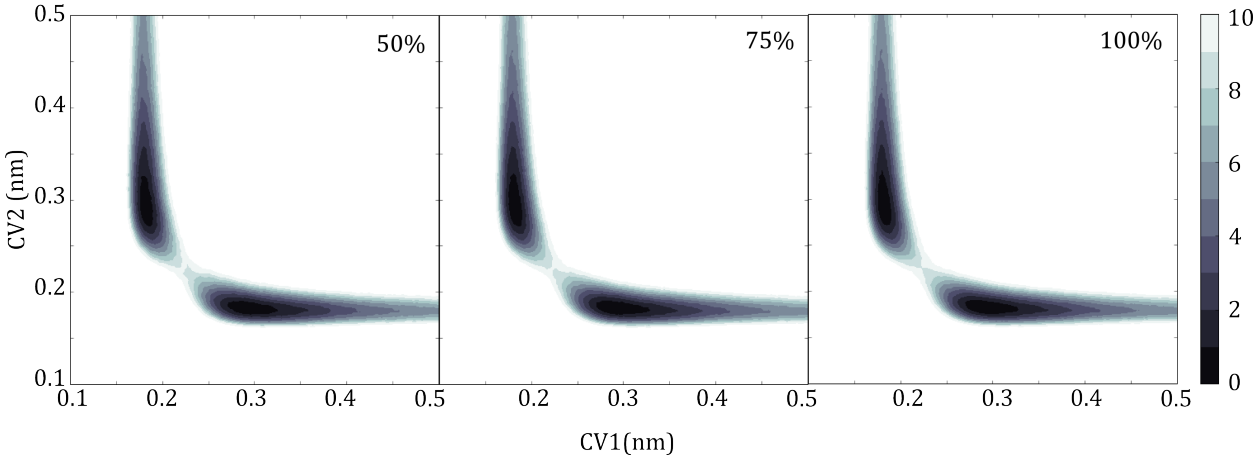

Figure S 2. The free-energy surfaces (FES) for the reaction determined at the PM6 level of theory at 50%, 75% and 100% of the total simulation length, as indicated in the top right corner of each panel.

Table S 1. Data from the stride, τ, parameter study

T(K)	ν (ns)	p_{KS}	σ (ns)	w_0 (kJ/mol)	T	# of events
300	96.9(5.2e-1)	0.99	0.0025	0.3	200000.00	51
300	167(8.9e-1)	0.32	0.0025	0.3	100000.00	78
300	157(5.2e-1)	0.971	0.0025	0.3	20000.00	142
300	184(7.9e-1)	0.906	0.0025	0.3	20000.00	**133**
300	202(8.2e-1)	0.8	0.0025	0.3	5000.00	125
300	180(5.4e-1)	0.88	0.0025	0.3	**1000.00**	**211**
300	185(8.2e-1)	0.765	0.0025	0.3	500.00	100
300	181(4.5e-1)	0.845	0.0025	0.3	50.00	374
300	95.4(3.5e-1)	0.76	0.0025	0.3	25.00	194
300	34.8	***1.50E-04***	0.0025	0.3	**5.00**	**478**

Table S 2. Study of stochastic sampling of the N=20 data set obtained with MetaRates at 300K for the S_N2 reaction. The mean escape time, standard error of the mean, and standard deviation for 1000 random samples (with replacement) are given.

N	ν (ns)	SEM (ns)	STD (ns)
5	157	2.2	69
10	170	1.7	52
15	175	1.4	43
20	175		