THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF CHEMICAL ENGINEERING

# A STUDY OF SOLVATION ENERGY USING MOLECULAR DYNAMICS

YUSHENG CAI
SPRING 2019

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Chemical Engineering
with honors in Chemical Engineering

Reviewed and approved* by the following:

Wayne Curtis
Professor of Chemical Engineering
Honors Advisor

Scott Milner
Joyce Chair and Professor of Chemical Engineering
Thesis Advisor

*Signatures are on file in the Schreyer Honors College.

# Abstract

Chemical reactions typically take place in solvents, which can interact with reacting species and products, thereby shifting the free energy of reaction. Quantum calculations using density functional theory (DFT) are widely used to compute reaction free energies, however the calculation can only be performed in vacuum or in continuum dielectrics. Thus, DFT cannot account for solvation effects with any molecular detail. We propose to improve upon the DFT results by adding the difference in solvation free energy of reactants and products. Molecular dynamics simulations can obtain the solvation free energies by computing the work to slowly turn on interactions between a solute and surrounding solvent. We apply this approach to study the preferred state of a proton dissolved in water. First, DFT calculations were performed to obtain reaction energies for the reaction $H_2O + H_3O^+ \longrightarrow H_5O_2^+$. Next, molecular dynamics simulations were performed on three aqueous solutes: $H_2O$, $H_3O^+$, $H_5O_2^+$. The reaction energy from DFT and solvation energy from simulations can then be combined to give a better estimate of the free energy of the reaction $H_2O + H_3O^+ \longrightarrow H_5O_2^+$.

# Contents

**List of Figures** **iii**

**List of Tables** **iv**

**1 Introduction** **1**

**2 Density Functional Theory** **3**

**3 Molecular Dynamics** **5**
3.1 Background Information . . . . . 5
3.1.1 Basics of Molecular Dynamics . . . . . 5
3.1.2 Basics of free energy calculations . . . . . 6
3.2 Methods . . . . . 9
3.2.1 Generate initial random system configuration . . . . . 9
3.2.2 Simulation with full solute-solvent interactions . . . . . 9
3.2.3 Thermodynamic integration job flow . . . . . 9
3.3 Water in water simulation . . . . . 10
3.3.1 Tip3p water model . . . . . 10
3.3.2 SPC/E water model . . . . . 15
3.3.3 Experimental solvation energy of water . . . . . 16
3.4 $H_3O^+$ into water simulation . . . . . 18
3.4.1 Parameter Fitting . . . . . 18
3.4.2 MD simulation for $H_3O^+$ into water . . . . . 21
3.5 $H_5O_2^+$ into water . . . . . 23
3.5.1 Parameter fitting . . . . . 23
3.5.2 MD simulations for $H_5O_2^+$ . . . . . 26
3.6 Hess's Law calculation for reaction free energy in aqueous phase . . . . . 28

**Bibliography** **30**

# List of Figures


1.1 Overall flow of the research . . . . . 2

2.1 $\Delta F_{rxn}$ with selected basis sets in the order of complexity . . . . . 4
2.2 Optimized Structures . . . . . 4

3.1 Pictorial representation of thermodynamic integration . . . . . 7
3.2 Graph of $\frac{\partial H}{\partial \lambda}$ time series in a $H_2O$ into $H_2O$ simulation . . . . . 11
3.3 $\langle \frac{\partial H}{\partial \lambda} \rangle$ at different $\lambda$ for $H_2O$ into $H_2O$ simulation using TIP3P water model . . . . . 11
3.4 Autocorrelation values for all $\frac{\partial H}{\partial \lambda}$ values in TIP3P water model simulation . . . . . 12
3.5 Linearized autocorrelation curve . . . . . 13
3.6 Integral curve for water into water simulation using Tip3P water model . . . . . 14
3.7 Histogram plot of number of configurations at different energies for $H_2O$ simulation 15
3.8 Integral curve for SPC/E and TIP3P water model simulations . . . . . 16
3.9 Scan of O-H bond around its optimal value of 0.098 nm in a $H_3O^+$ molecule . . . . 19
3.10 Scan of O-H-O angle around its optimal value of 2.1 rad in a $H_3O^+$ molecule . . . . 20
3.11 $H_3O^+$ bond and angle scan pictorial representation . . . . . 20
3.12 Time series data of $\frac{\partial H}{\partial \lambda}$ for $H_3O^+$ solvating into $H_2O$ simulation . . . . . 21
3.13 Integrand curve of $\langle \frac{\partial H}{\partial \lambda} \rangle$ vs $\lambda$ for $H_3O^+$ into $H_2O$ simulation . . . . . 21
3.14 $H_3O^+$ integral curve . . . . . 22
3.15 Scan of $H_5O_2{}^+$ dihedral . . . . . 23
3.16 Columbic interaction in dihedral scan for $H_5O_2{}^+$ . . . . . 24
3.17 $H_5O_2{}^+$ dihedral plot . . . . . 25
3.18 Central bond scan for $H_5O_2{}^+$ molecule . . . . . 25
3.19 $H_5O_2{}^+$ central O-H-O angle scan . . . . . 26
3.20 Time series for $\frac{\partial H}{\partial \lambda}$ of $H_5O_2{}^+$ solvation into $H_2O$ simulation . . . . . 26
3.21 Integrand curve of $\langle \frac{\partial H}{\partial \lambda} \rangle$ vs $\lambda$ for $H_5O_2$ into $H_2O$ simulation . . . . . 27
3.22 Integral for $H_5O_2{}^+$ solvation into $H_2O$ . . . . . 27
3.23 Hess's Law applied to calculating the aqueous reaction free energy . . . . . 28
3.24 Three water cluster with proton . . . . . 29

# List of Tables

3.1 Theoretical $\Delta F_{sol}$ compared to experimental $\Delta F_{sol}$ for different water models . . . 17
3.2 Results from Hess's Law calculation . . . . . 28

# Chapter 1

# Introduction

A close examination of the free energy behavior is often necessary to understand chemical and biochemical processes.[1] The probability of finding a molecular system in one state instead of another is determined by the free energy difference of the two states. Therefore, free energy differences can be used to obtain a wide range of important physical properties such as protein-ligand binding constants, partition coefficients and solubility. Recent developments in software and hardware have brought free energy calculations to a level of robust modelling tool and at the same time widening its applications across different fields such as chemistry, biology, material science etc.

The ability to calculate free energies from molecular simulations allows us to probe states of a system that are not accessible experimentally.[2] Moreover, free energy calculations are of great importance in the emerging fields of rational drug design, which is hard to achieve without an accurate knowledge of the associated free energy changes. In this research work, the aqueous reaction free energy of a chemical reaction is studied using molecular simulations. The reaction of interest is $H_2O + H_3O^+ \longrightarrow H_5O_2^+$. This particular reaction is chosen because we want study whether the proton in aqueous solvent will stay on a water molecule to form hydronium ion or a water cluster such as $H_5O_2^+$. Understanding where the protons prefers to stay has applications in fields such as electrochemistry and catalysis. To calculate the aqueous reaction free energy, the reaction is first taken into vacuum where DFT calculation can be performed to obtain the reaction

free energy in vacuum. Then, to obtain the aqueous reaction free energy, the reaction is solvated back into water using free energies of solvation calculated from molecular dynamics simulations. The overall flow of the research is shown in **figure 1.1**

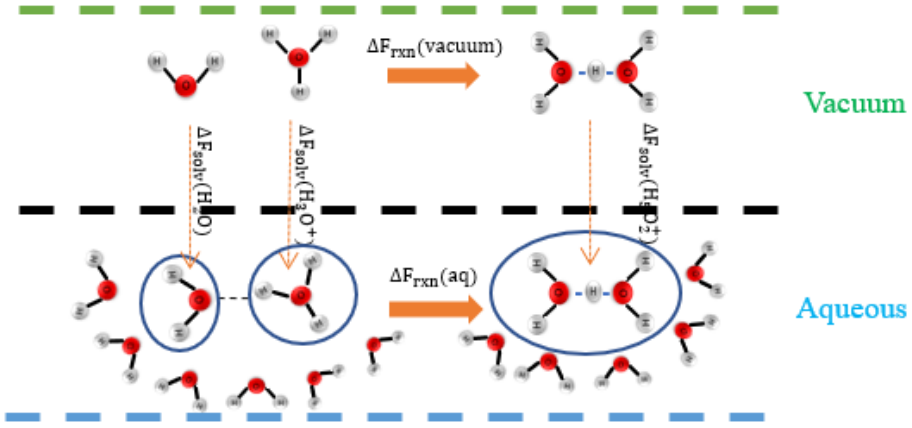



**Figure 1.1:** Overall flow of the research

# Chapter 2

# Density Functional Theory

Density functional theory (DFT) has become the most prominent tool for the calculation of ground state properties of electronic system in the recent years.[6] In the first step of this research project, DFT calculation is performed to obtain the $\Delta F_{rxn}$ of $H_2O$ + $H_3O^+$ $\longrightarrow$ $H_5O_2^+$. The dynamics of the chemical reaction is not considered in our simulation as we are performing calculation at the equilibrium state of the system. Therefore, the $\Delta F_{formation}$ of each molecule is first calculated and the $\Delta F_{rxn}$ can be calculated using **equation 2.1**

$$\Delta F_{rxn} = \sum_{product} \Delta F_{formation} - \sum_{reactants} \Delta F_{formation} \tag{2.1}$$

DFT calculations in this work are all performed in the Gaussian 09 package. In Gaussian 09 package, optimization plus frequency DFT calculation can be performed on a molecule to obtain its $\Delta F_{formation}$ as well as the optimized structure. To perform a DFT calculation, basis sets and functional form must first be specified. There are many types of functional forms in DFT calculations including non-empirical functionals and hybrid functionals. As suggested by the names, non-empirical functionals refers to the functionals that have not been fitted to any empirical results while hybrid functionals has a mix of *ab initio* and empirical parameters. In this work, the hybrid functional *B3LYP* is chosen. Basis sets refer to the mathematical description of orbitals of a system that is used to approximate theoretical calculations or modeling. There are many options of basis

sets available in Gaussian 09 package. $\Delta F_{rxn}$ is calculated for various basis sets using **equation 2.1**. The final $\Delta F_{rxn}$ is chosen when $\Delta F_{rxn}$ barely changes with increasing complexity of basis set. A plot of $\Delta F_{rxn}$ vs selected basis sets in the order of complexities are shown below in **figure 2.1**

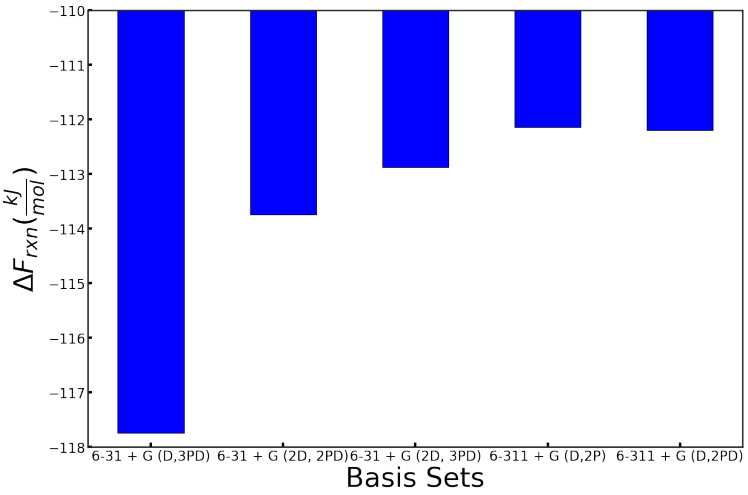
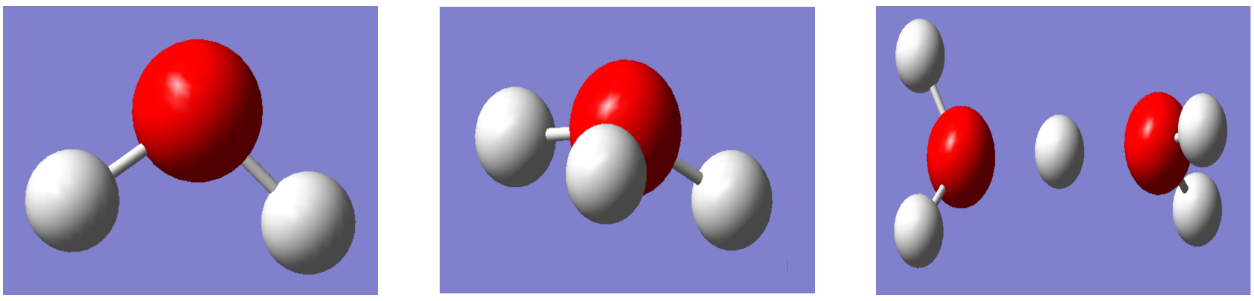



**Figure 2.1:** $\Delta F_{rxn}$ with selected basis sets in the order of complexity

From **figure 2.1**, *6-311 + G(D,2P)* level of theory is chosen and the final $\Delta F_{rxn}$ is determined to be around $-112.15\frac{kJ}{mol}$. As mentioned before, optimized structures are also generated from calculations performed above. The optimized structures of the three molecules are shown below in **figure 2.2**



**(a)** Optimized $H_2O$ **(b)** Optimzed $H_3O^+$ **(c)** Optimized $H_5O_2^+$

**Figure 2.2:** Optimized Structures

# Chapter 3

# Molecular Dynamics

## 3.1 Background Information

### 3.1.1 Basics of Molecular Dynamics

In molecular dynamics (MD) simulation, Newton's equations of motion as shown in **equation 3.1** and **equation 3.2** are solved for all $N$ particles in the system. In the two equations shown, $x_i$ represents the positions, $p_i$ represents the momenta and $U(q_i)$ represents the interaction potential of a set of particles in the system.

$$\frac{dx_i}{dt} = \frac{p_i}{m} \tag{3.1}$$

$$\frac{dp_i}{dt} = -\frac{\partial U}{\partial x_i} \tag{3.2}$$

With $N$ particles in the system and $d$ dimensions, there would be a total of $dN$ ordinary differential equations to solve. By solving these ordinary differential equations, a clip by clip time evolution or "movie" of the movement of molecules in the system can be generated. Using this "movie", various microstates available to this particular system can be generated. Energy and pressure can then be computed with the trajectory generated as they are explicit functions of coordinates and

velocities of the molecules. However, thermodynamic properties such as entropy and free energy of the system cannot be calculated directly. For example, the entropy of a system depends on the probability of different microstates appearing as shown in **equation 3.3** where k is the Boltzmann factor and $p_i$ is the probability of microstate $i$ appearing

$$S = -k \sum_i p_i \log(p_i) \tag{3.3}$$

As a result, it is necessary to develop techniques to calculate thermodynamic quantities such as free energy and entropy. Since solvation free energy is calculated in this work, the thermodynamic methods to obtain free energy using MD will be discussed.

### 3.1.2 Basics of free energy calculations

In a canonical ensemble, the Helmholtz free energy is defined as shown in **equation 3.4**

$$F = -\frac{1}{\beta} \log(Q) \tag{3.4}$$

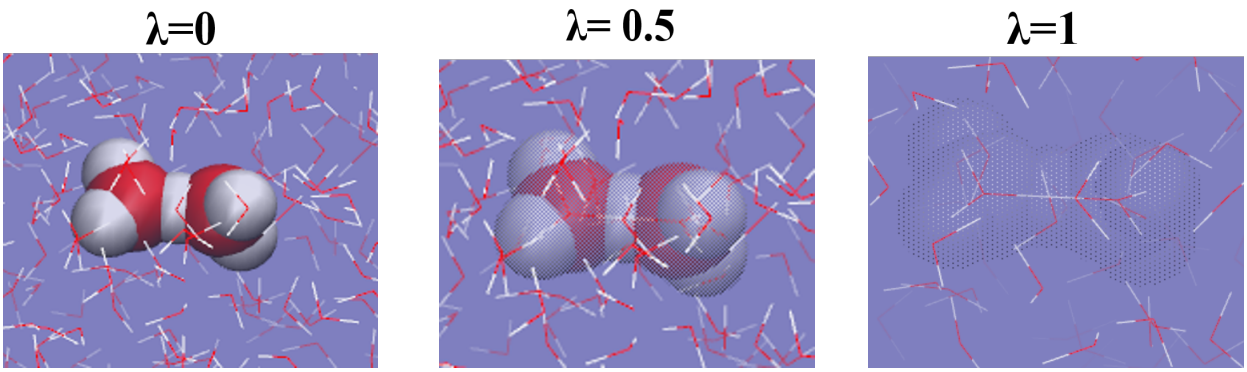
where $\beta$ is equal to $\frac{1}{k_b T}$ and $Q$ is the partition function which describes the number of microstates accessible to the system. Discrete partition function of a system is defined mathematically by **equation 3.5** where $H_i$ is the Hamiltonian or total energy of a specific state $i$

$$Q = \sum_i e^{-\beta H_i} \tag{3.5}$$

There are many useful statistical mechanical formulas and computational methods used to obtain the free energy differences between two states.[5] First of which is a direct counting method. The calculation is performed by counting the number of microstates accessible in the two corresponding states and applying **equation 3.6** where $Q_A$ refers to the number of microstates in state A and similarly $Q_B$ refers to the number of microstates in state B.

$$\Delta F = -\frac{1}{\beta}\log(\frac{Q_B}{Q_A}) \tag{3.6}$$

Direct counting method is only appropriate when the microstates of interest happen frequently in the system to obtain reliable statistics. However, there are a lot of systems where it takes a long time for the microstates of interest to happen *ie. protein folding*. If one were to perform the calculation by enumerating the microstates that happens rarely, the MD simulation will require large computation time. One alternative method to direct counting is the integration method. Integration methods determine the change in free energy of the two states of a system from the integral of the work required to go from an initial state to a final state using a reversible path. One of such techniques is called thermodynamic integration. Thermodynamic integration is a very efficient method for estimating the free energy difference between state A and state B of a system. The application of thermodynamic integration spans from computational chemistry, biology to material science. For example, it is used to calculate the free energy of morphing polymer A to polymer B as well as estimating the Flory Huggins $\chi$ parameter that quantifies the phase behavior and mesoscale structures of polymer blends.[7] In thermodynamic integration, a coupling parameter $\lambda$ is introduced to control the strength of interactions. In this work, $\lambda$ is introduced as a coupling parameter that controls the strength of interactions between solvent and solute. A pictorial representation of the time evolution of thermodynamic integration performed in this work is shown below.




**(a)** $\lambda$=0 **(b)** $\lambda$=0.5 **(c)** $\lambda$=1

**Figure 3.1:** Pictorial representation of thermodynamic integration

The equation to obtain $\Delta F_{sol}$ from thermodynamic integration simulation are derived from first principle using the relationship between partition function $Q$ and the Helmholtz free energy $F$ in **equation 3.5**. Taking the derivative with respect to $\lambda$ on both sides of **equation 3.5** yields

$$\frac{\partial F}{\partial \lambda} = -\beta^{-1} \frac{\frac{\partial}{\partial \lambda} \sum_i e^{-\beta H_i(\lambda)}}{\sum_i e^{-\beta H_i(\lambda)}} = \frac{\sum_i \frac{\partial H_i(\lambda)}{\partial \lambda} e^{-\beta H_i(\lambda)}}{\sum_i e^{-\beta H_r(\lambda)}} \tag{3.7}$$

Note that the right hand side of the equation is a weight average of $\frac{\partial H_i(\lambda)}{\partial \lambda}$ and this yields the final **equation 3.8**

$$\Delta F_{AB} = \int_{\lambda_A}^{\lambda_B} \left\langle \frac{\partial H_i(\lambda)}{\partial \lambda} \right\rangle_{\lambda} \partial \lambda \tag{3.8}$$

As a result, the Helmholtz free energy difference between A and B can be calculated by integrating the average values of $\frac{\partial H_i}{\partial \lambda}$ curve with respect to $\lambda$

## 3.2 Methods

### 3.2.1 Generate initial random system configuration

The optimized structures of $H_2O$, $H_3O^+$ and $H_5O_2^+$ generated using *ab initio* calculations as shown in **figure 2.2** are solvated with randomly placed solvent water using an appropriate water model. TIP3P and SPC/E water models were chosen for the solvent water and will be discussed at length later. After the molecule of interest is solvated, a simulation with interaction between solute and solvent turned on ($\lambda = 0$) is first performed on the system.

### 3.2.2 Simulation with full solute-solvent interactions

Firstly, energy minimization is performed on the system to reduce the magnitude of forces acting on the atoms in the initial system configuration which was randomly generated. Steepest descent algorithm is used in this part of for energy minimization since it is robust and easy to implement. Once energy is minimized, temperature and pressure of the system also need to be equilibrated. NVT (constant number of particles, volume and temperature) and NPT (constant number of particles, pressure and temperature) equilibrations were performed on the system. NVT and NPT runs are essentially shorter versions of the full MD run at around $100$ ps that ensure the system is at constant temperature and pressure. Finally a $1$ ns MD simulation is performed to provide the $\frac{\partial H}{\partial \lambda}$ values and the coordinates of the molecules in the system.

### 3.2.3 Thermodynamic integration job flow

Thermodynamic integration used in the early stages of this work has the following procedure for each $\lambda$

- Perform energy minimization on the system to reduce magnitudes of forces.
- $100$ps NVT and NPT run to ensure the system is at constant temperature and pressure.
- Perform a $1$ns MD run

In the later stage of this research, a slightly more efficient job flow was adopted that takes in the equilibrated system coordinates from previous $\lambda$ as the input file and perform only the 1ns MD simulation. The algorithm of the job flow goes like following

- Extract the coordinates output from the $\lambda_{i-1}$ simulation
- Perform the simulation at $\lambda_i$ using the coordinates output from the last simulation
- Repeat **step 1 and step 2** $n-1$ times where n is the length of the $\lambda$ list

## 3.3 Water in water simulation

One of the important aspects of performing this simulation is to use an appropriate water model. There are various water models available that are determined from quantum mechanics, molecular mechanics and experimental results. The two that were chosen for this simulation were SPC/E and TIP3P. Results from the two water models will be compared and evaluated. Future simulations of $H_3O^+$ and $H_5O_2{}^+$ into water will use the water model that produces the more accurate result. How the accuracy of results are determined will also be discussed in later section.

### 3.3.1 Tip3p water model

In an equilibrium simulation, the pressure and energy fluctuate around an average value. Shown in **figure 3.2** is the time series of $\frac{\partial H}{\partial \lambda}$ at an arbitrarily chosen $\lambda$. The value of $\frac{\partial H}{\partial \lambda}$ fluctuates around its average value which is a good indication that the simulation has reached equilibrium.

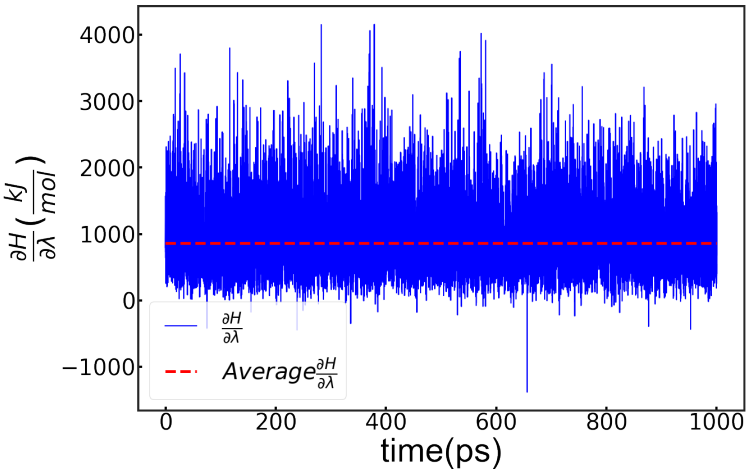



**Figure 3.2:** Graph of $\frac{\partial H}{\partial \lambda}$ time series in a $H_2O$ into $H_2O$ simulation

The $\langle \frac{\partial H}{\partial \lambda} \rangle$ (average value of $\frac{\partial H}{\partial \lambda}$) is calculated for all $\lambda$ and plotted. An interpolation is then performed to obtain a smooth curve for integration as shown in **figure 3.3**. $\langle \frac{\partial H}{\partial \lambda} \rangle$ for this simulation decreases very rapidly to $0$ and dips below $0$ at a $\lambda$ value of $0.5$. Note that there are multiple paths for the system to take from state A to state B, therefore the $\langle \frac{\partial H}{\partial \lambda} \rangle$ integrand curve could take on many different shapes. However, since Helmholtz free energy is a state function which is path independent, only the integral is physically meaningful.

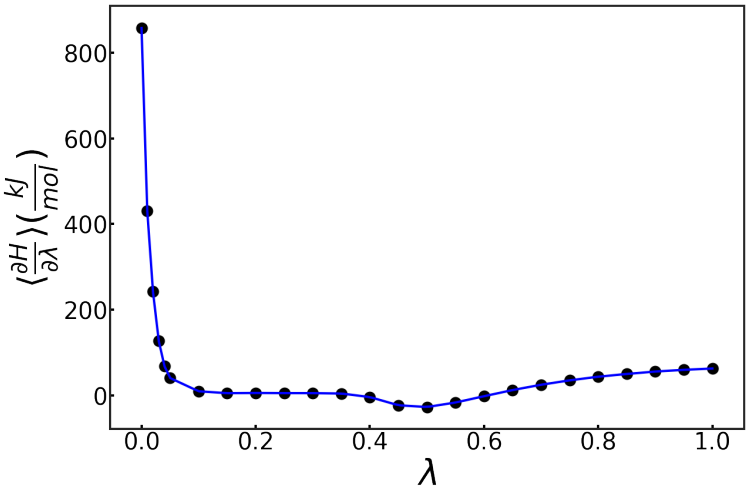



**Figure 3.3:** $\langle \frac{\partial H}{\partial \lambda} \rangle$ at different $\lambda$ for $H_2O$ into $H_2O$ simulation using TIP3P water model

In an MD simulation, the only way to get the exact value for some average quantity is by running the simulation for an infinite amount of time. Therefore, for any simulation there is always some

error in the average, a deviation from the true value. In this simulation, we seek to find the expected error in $\Delta F_{sol}$ which according to **equation 3.8** is the area under the integrand curve shown above in **figure 3.3**. To find the standard error of $\Delta F_{sol}$, standard deviation of $\langle \frac{\partial H}{\partial \lambda} \rangle$ must first be determined. Standard deviation of the $\langle \frac{\partial H}{\partial \lambda} \rangle$ can be found by the following **equation 3.9** where $n$ is the number of independent measurement and $\sigma$ represents the standard deviation.

$$\sigma(\langle \frac{\partial H}{\partial \lambda} \rangle) = \frac{\sigma(\frac{\partial H}{\partial \lambda})}{\sqrt{n}} \tag{3.9}$$

In time series data, two data points may still be correlated within a certain time frame. Therefore, the number of independent trials in a time series is found by dividing the total time by the time it takes for one point to "forget" about what happened previously. This time lag can be found by applying the autocorrelation function on the time series. The autocorrelation function of a time series with some lag **k** is defined as

$$r_k = \frac{\frac{1}{N-k}\sum_{i=1}^{N-k}(Y_i - \bar{Y})(Y_{i+k} - \bar{Y})}{\frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y})^2} \tag{3.10}$$

Using the above equation at different lags, a plot of autocorrelation values $r_k$ versus lag numbers $k$ can be generated for all $\lambda$ as shown in **figure 3.4**

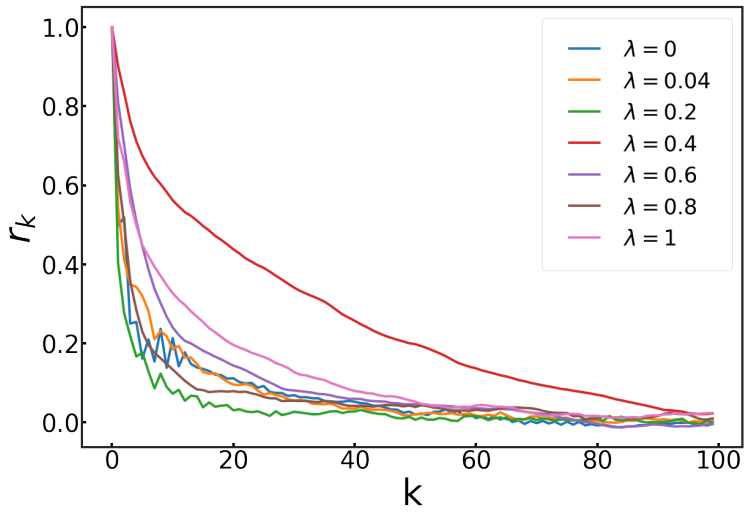



**Figure 3.4:** Autocorrelation values for all $\frac{\partial H}{\partial \lambda}$ values in TIP3P water model simulation

The exponential decay nature of the above plot motivated the equation of form $e^{-\frac{t}{\tau}}$ where $\tau$ is defined as the autocorrelation time which measures how long it takes for one point to "forget" what happened previously. From this definition, $\tau$ can be defined mathematically as

$$\frac{1}{\tau} = -\frac{\partial log(r_k)}{\partial t} \tag{3.11}$$

Using **equation 3.11**, if $\log(r_k)$ is plotted against $k$, the slope of the linear line generated will have a value of $-\frac{1}{\tau}$. Plots of $\log(r_k)$ vs $k$ are shown below in **figure 3.5**

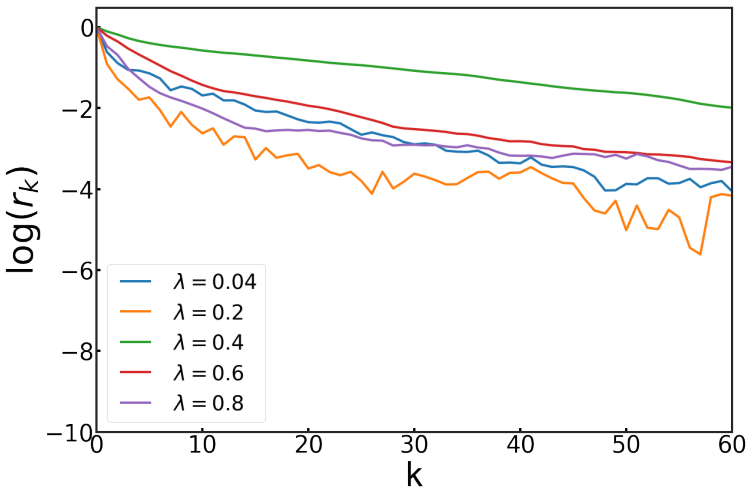



**Figure 3.5:** Linearized autocorrelation curve

Linearized autocorrelation curve for all $\lambda$ are fitted to a linear equation where the slope is equal to $-\frac{1}{\tau}$. The number of independent trials $n$ in a time series can then be calculated by dividing the total number of points in that time series by the lag time $\tau$ calculated. By using the number of independent trials, the standard deviations of the $\langle \frac{\partial H}{\partial \lambda} \rangle$ can be calculated using **equation 3.9**.

To compute the error bars, the integral is approximated as a sum as shown in **equation 3.12**

$$\int_a^b f(x)dx = \sum_{i=1}^{n-1}(x_{i+1} - x_i)f(x_i) \tag{3.12}$$

Error bars of the integral curve can then be calculated using the formula for error propagation as shown in **equation 3.13** where $\sigma$ represents the standard deviation

$$\sigma(\sum_i a_i x_i) = \sqrt{\mathrm{a}_i^2 \cdot \sigma(x_i)} \tag{3.13}$$

Finally, the integrand shown in **figure 3.3** is integrated numerically and the following integral curve with error bars is constructed shown in **figure 3.6**

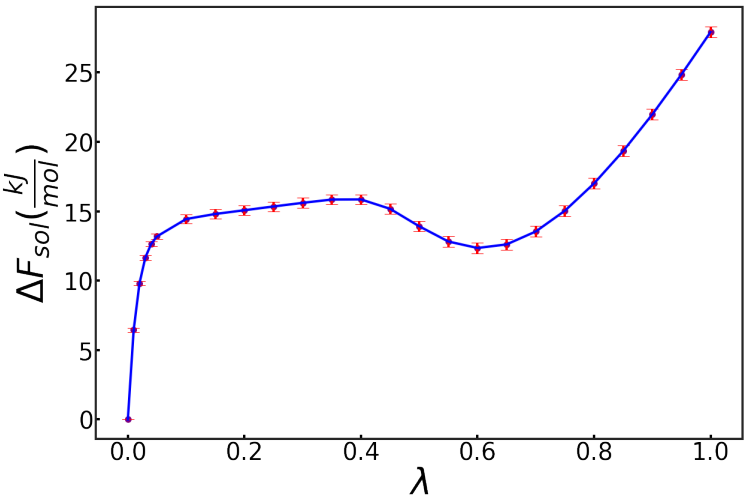



**Figure 3.6:** Integral curve for water into water simulation using Tip3P water model

The $\Delta F_{sol}$ from this simulation is determined to be $27.92 \pm 0.388$ $\frac{kJ}{mol}$ This result might seem puzzling initially since a positive $\Delta F_{sol}$ for water into water implies that the process isn't favorable. The reason for this seemingly confusing result lies in the definition of $\lambda$ for this work. Since $\lambda = 0$ is defined as full interaction and $\lambda = 1$ is defined as no interaction, the physical phenomena represented by this simulation is the process of removing a water molecule from water solvent. As a result, the negative of the result calculated represents the solvation of water into water since the path is reversible. The $\Delta F_{sol}$ of water from this simulation is finally determined to be $-27.92 \pm 0.388$ $\frac{kJ}{mol}$.

A final step of this work is to make sure that the sampling of this simulation is sufficient. Histogram curves are generated for this work to show the number of states at different energy levels for various $\lambda$. In theory, we expect to see two things from the histogram curve. Firstly, we expect to see Gaussian shaped histogram curves as energies in partition function $Q$ as mentioned in **equation 3.5** is in theory normally distributed. Secondly, there should be significant overlaps between the

histogram curves of neighboring $\lambda$ as the process is reversible. The plot of the histogram of selected $\lambda$ is shown in **figure 3.7**. As shown in **figure 3.7**, the two criterion were met proving that the sampling of this simulation is sufficient.

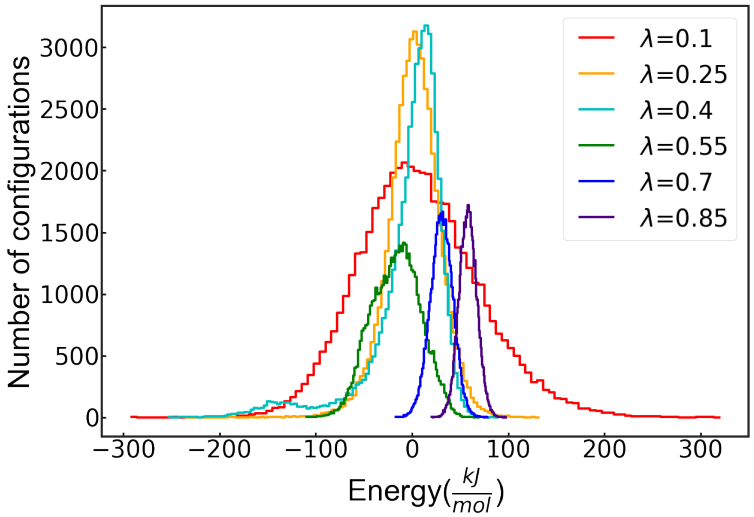



**Figure 3.7:** Histogram plot of number of configurations at different energies for $H_2O$ simulation

### 3.3.2 SPC/E water model

SPC/E model is another widely used water model in molecular simulations alongside TIP3P water model. The same thermodynamic integration procedure is performed for water into water simulation using SPC/E water model. The results between the two water models are, however, not equivalent but very close. The integrand curve for SPC/E water model has a similar shape to the integrand curve of TIP3P because the same reversible path was taken for both simulations. However, $\langle \frac{dH}{d\lambda} \rangle$ value at full interaction is much higher for SPC/E water model than TIP3P water model. As a result, the absolute value of the final integral value for the simulation using SPC/E water model is higher than the TIP3P water model. The integral curve for both TIP3P and SPC/E water model are shown in **figure 3.8**

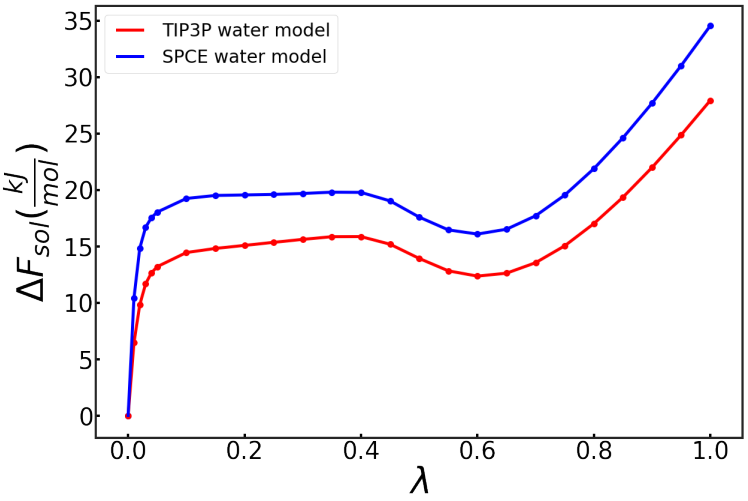



**Figure 3.8:** Integral curve for SPC/E and TIP3P water model simulations

The $\Delta F_{sol}$ value for SPC/E water model simulation is determined to be $-34.54 \pm 0.41 \frac{kJ}{mol}$ which is lower than (higher in absolute value than) the value for TIP3P water model.

The results from the two water models were not expected to agree as they were parameterized and designed for different purposes. To determine the suitable water model for this simulation, the results from the two water models were compared to the experimental result. The next logical question is find the experimental value that corresponds to the solvation of water. Dissolving a water molecule into water is the reverse process of removing a water molecule from water. The latter represents the physical phenomenon of water evaporating which creates vapor pressure. Therefore, if relationship can be determined between the $\Delta F_{sol}$ calculated in this section and vapor pressure, the results from this work can be compared with the vapor pressure of water which is well-known experimentally.

### 3.3.3 Experimental solvation energy of water

Necessary theory needs to first be developed to calculate the solvation energy of water from experimentally determined water vapor pressure. First, the water molecules of interest (water molecules vaporizing) are "tagged" which is a fraction $f$ of the total water molecules. In the liquid, the tagged molecules form an ideal solute with molar volume equal to the molar volume of

liquid water divided by $f$. In the vapor, the tagged molecules contribute a fraction $f$ to the ideal gas vapor pressure. When a tagged molecule is removed from vapor to liquid phase, its molar volume decreases considerably, thereby decreasing its ideal-gas free energy. In equilibrium, the total free energy change of solvation can be calculated as shown in **equation 3.14**.

$$\Delta F_{sol} = -\frac{1}{\beta} log(\frac{V_{vap}}{V_{sol}}) \tag{3.14}$$

Using the above equation, the molar volume in the vapor $V_{vap}$ can be represented as

$$V_{vap} = V_{sol} e^{-\beta \Delta F_{sol}} \tag{3.15}$$

Finally using the ideal gas equation of state, the vapor pressure can be predicted as

$$P_{vap} = \frac{RT}{V_{sol}} e^{\beta \Delta F_{sol}} \tag{3.16}$$

Since solvation free energy $\Delta F_{sol}$ is in the exponential term of **equation 3.16**, the systematic error of $\Delta F_{sol}$ propagates exponentially. To avoid the exponential error propagation, the experimental $\Delta F_{sol}$ is calculated from the vapor pressure of water at ambient conditions (T=298.15K, P=1 bar) using **equation 3.16**. The vapor pressure of water at room temperature is found to be $3.171$ kPa and $\Delta F_{sol,exp}$ is determined to be $-26.7 \frac{kJ}{mol}$.

**Table 3.1:** Theoretical $\Delta F_{sol}$ compared to experimental $\Delta F_{sol}$ for different water models

| **Water model** | $\Delta F_{sol}$ ($\frac{kJ}{mol}$) | **Percent difference from experimental value** |
|---|---|---|
| TIP3P | $-27.92 \pm 0.39$ | 4.6 % |
| SPCE | $-34.54 \pm 0.41$ | 29.4 % |

As shown in the results from **table 3.1**, TIP3P water model has proven to be a suitable water model for the purpose of this research work. The results also show that $\Delta F_{sol,exp}$ is slightly more positive than $\Delta F_{sol,TIP3P}$ indicating less favorable solvation of water into water. From a simulation point of

view, the Lennards Jones potentials can be weakened by tuning the $\epsilon$ and $\sigma$ parameters in **equation 3.20**. By weakening the Lennard Jones potential, the water molecules do not interact as strongly causing solvation of a water molecule into a water solvent to become less favorable.

## 3.4 $H_3O^+$ into water simulation

### 3.4.1 Parameter Fitting

Simulation for $H_3O^+$ is a slightly more complex task to perform since $H_3O^+$ molecule does not have an existing well defined molecular description. It is necessary to first generate a relatively accurate molecular description of the molecule. As a non-reactive force field[4] is used in this research, the bond stretching as well as angle stretching of a molecule can be modelled by a harmonic potential as shown in **equation 3.17** where $x_0$ is the optimal position.

$$U_x = -\frac{1}{2}k(x - x_0)^2 \tag{3.17}$$

The spring constants $k$ can be found through a scan of energy around the optimal length/angle and fitting the energy versus length/angle data to harmonic potential as shown in **equation 3.17**.

The energies at different configurations of the molecule are calculated by *ab initio* calculation at the **6-311g** level theory in Gaussian 09. Firstly, a scan of the O-H bond around its optimal value of $0.098$ nm generated the following curve

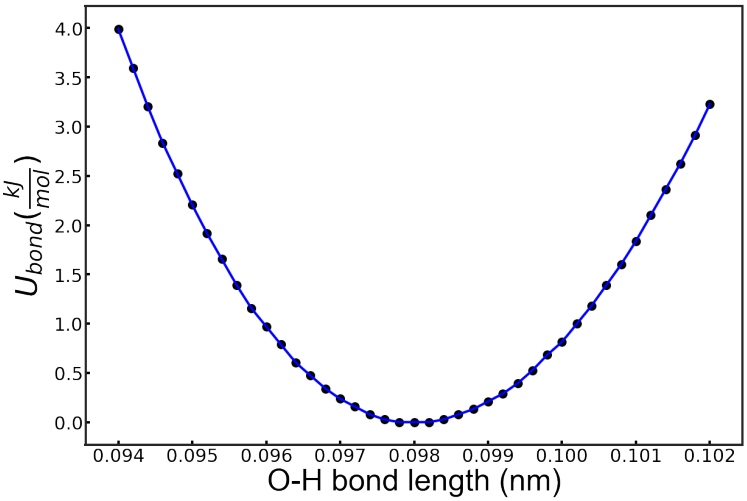



**Figure 3.9:** Scan of O-H bond around its optimal value of 0.098 nm in a $H_3O^+$ molecule

From the curve shown in **figure 3.9** alongside **equation 3.17**, the spring constant for the O-H bond is calculated to be 451460 $\frac{kJ}{mol \cdot nm^2}$. Secondly, the angular force constant was calculated for H-O-H angle. The angular harmonic potential of a $H_3O^+$ molecule can be written in the following way in **equation 3.18** with **a,b,c** denoting the three H-O-H bonds present in $H_3O^+$ molecule.

$$U_\theta = -\frac{1}{2}k(\theta_a - \theta_{a0})^2 - \frac{1}{2}k(\theta_b - \theta_{b0})^2 - \frac{1}{2}k(\theta_c - \theta_{c0})^2 \tag{3.18}$$

The scan for this angle stretching proves to be difficult to perform because the H-O-H angles in $H_3O^+$ molecule are all dependent on each other. Optimized $H_3O^+$ molecule is shaped like an ammonia molecule such that the three H-O-H angles do not add up to 360 degrees. To resolve this complication, $H_3O^+$ molecule is assumed to be planar. Therefore, the dependence of the three H-O-H angles can be written such that the a little change of $\delta\theta$ in angles **B** and **C** cause a change of $2\delta\theta$ in angle **A**. As a result, **equation 3.18** can be rewritten as

$$U_\theta = -3k\delta\theta^2 \tag{3.19}$$

The resulting curve of $H_3O^+$ angular energy is shown below in **figure 3.10**

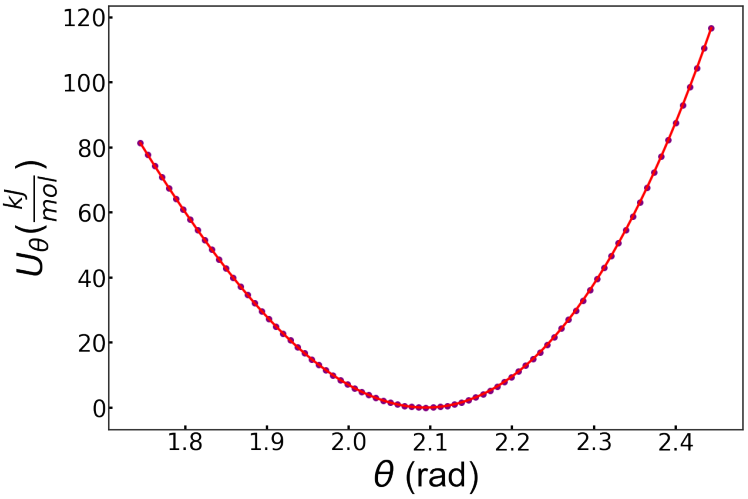
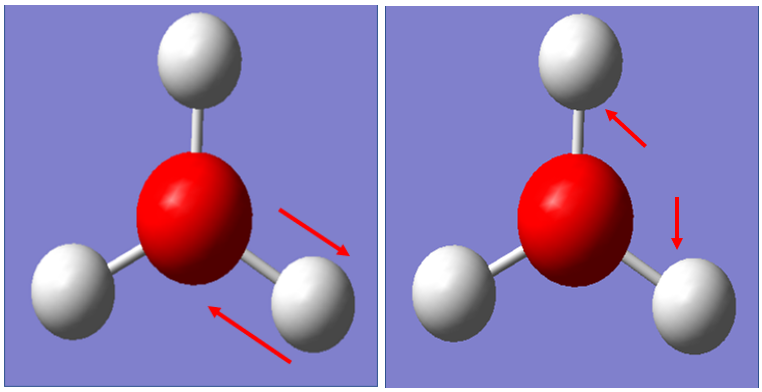



**Figure 3.10:** Scan of O-H-O angle around its optimal value of 2.1 rad in a $H_3O^+$ molecule

The force contant of the O-H-O angle in a $H_3O^+$ molecule is determined to be $275\frac{kJ}{mol \cdot rad^2}$. A pictorial representation of the bond stretching and angle stretching is shown below in **figure 3.11**.



**(a)** $H_3O^+$ bond scan pictorial representation

**(b)** $H_3O^+$ angle scan pictorial representation

**Figure 3.11:** $H_3O^+$ bond and angle scan pictorial representation

Other information such as the the charge distribution within a $H_3O^+$ molecule and the $\sigma$ and $\epsilon$ terms in the Lennard Jones potential as shown in **equation 3.20** are obtained from the work of Kusaka.[3]

$$\mathrm{V}_{LJ} = 4\epsilon[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^{6})] \tag{3.20}$$

### 3.4.2 MD simulation for $H_3O^+$ into water

With all the parameters for $H_3O^+$ determined, the same thermodynamic integration scheme can be performed. TIP3P water model is used for the solvent model as it was determined previously that solvation free energy of water obtained using TIP3P water model closely match the experimental result.

Firstly, the time series plot of $\frac{\partial H}{\partial \lambda}$ is plotted to ensure that the system has reached equilibrium in that it is oscillating around an average value. The time series is shown below in **figure 3.12**

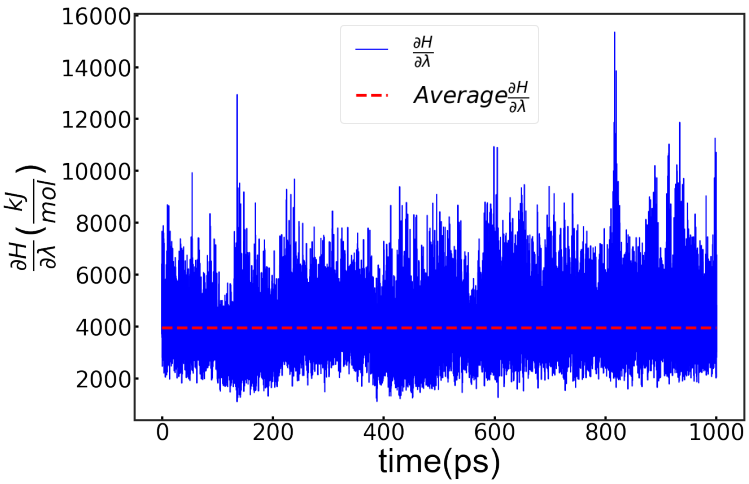



**Figure 3.12:** Time series data of $\frac{\partial H}{\partial \lambda}$ for $H_3O^+$ solvating into $H_2O$ simulation

The integrand of the $H_3O^+$ solvation simulation is shown below in **figure 3.13**

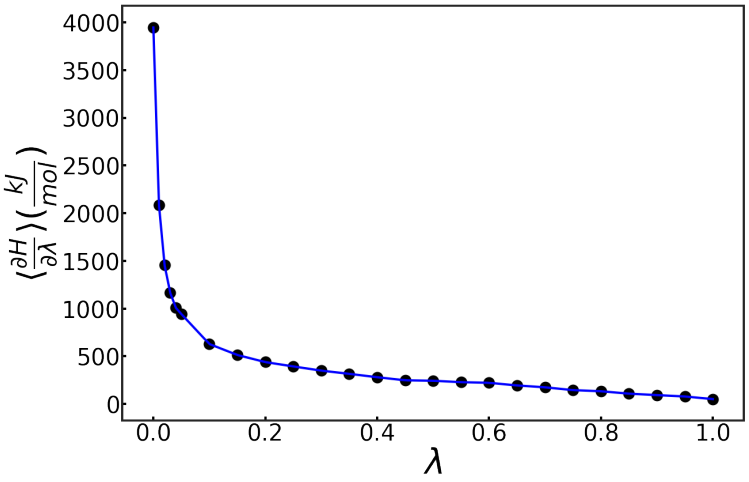



**Figure 3.13:** Integrand curve of $\langle\frac{\partial H}{\partial \lambda}\rangle$ vs $\lambda$ for $H_3O^+$ into $H_2O$ simulation

By integrating this curve, the $\Delta F_{sol}$ of $H_3O^+$ is calculated to be $-346.16 \frac{kJ}{mol} \pm 1.39$. The resulting integral curve is shown below in **figure 3.14**.

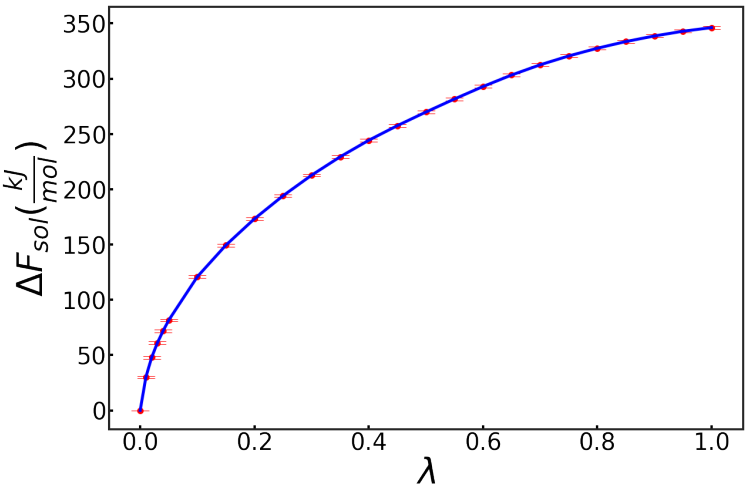



**Figure 3.14:** $H_3O^+$ integral curve

## 3.5 $H_5O_2^+$ into water

### 3.5.1 Parameter fitting

$H_5O_2^+$, similar to $H_3O^+$, lacks a well-defined molecular description. The optimized structure of $H_5O_2^+$ as shown in **figure 2.2** has two water molecules with a hydrogen atom in the middle that is hydrogen bonded to both water molecules. There are a lot of similar parameters between $H_3O^+$ and $H_5O_2^+$ such as the O-H bond stretching and the H-O-H angle stretching in the two water molecules. As a result, the O-H bond stretching and H-O-H angle stretching parameters developed for $H_3O^+$ molecule can be directly used for $H_5O_2^+$. The differences between the two molecules are that $H_5O_2$ has a dihedral potential and bond/angle stretching of the central O-H-O hydrogen bond.

To fit the dihedral potential, an appropriate dihedral potential model must first be chosen. The potential used in this work is shown in **equation 3.21**

$$\mathrm{V}_d(\phi_{ijkl}) = k_\phi(1 + \cos(n\phi - \phi_s)) \tag{3.21}$$

where $\phi_{ijkl}$ is defined as the angle between the planes $ijk$ and $jkl$. A scan on different dihedral configurations of the $H_5O_2^+$ molecule was performed and a pictorial representation viewed from the front of the molecule is shown in **figure 3.15**

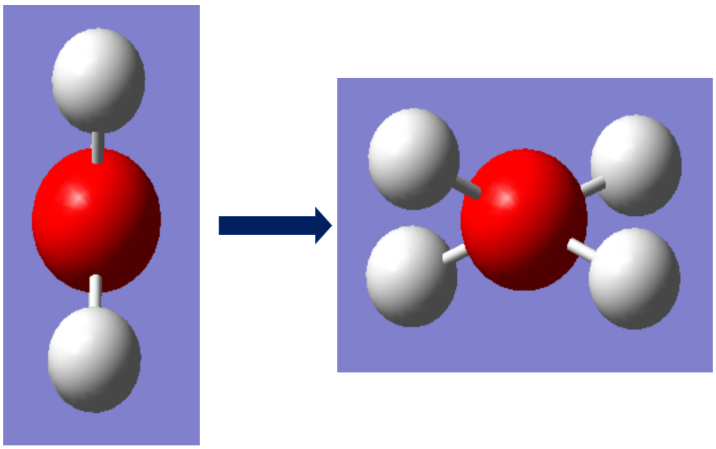


**Figure 3.15:** Scan of $H_5O_2^+$ dihedral

It is important to note that as the dihedral is changing, the distances between the four hydrogens on the two water molecules are changing as well resulting in a change in the Columbic energies. Therefore, in order to capture just the energy contribution from the dihedral potentials, the Columbic energies of the 4 hydrogen atoms interacting with each other must be subtracted from the total energy obtained from the *ab initio* calculations. The Columbic interactions between the four hydrogen atoms are shown pictorially in **figure 3.16**

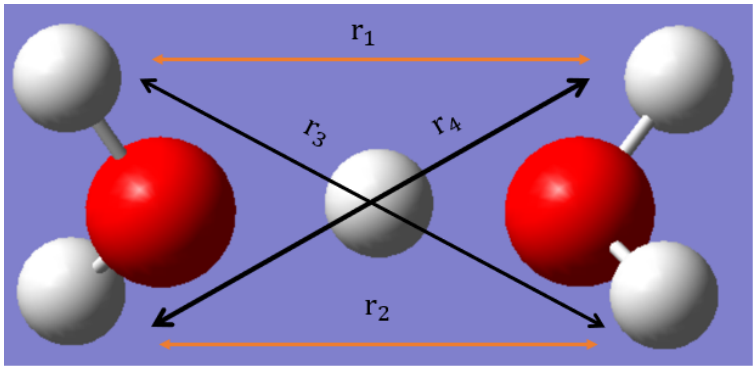



**Figure 3.16:** Columbic interaction in dihedral scan for $H_5O_2^+$

The Columbic potential is modelled by **equation 3.22**

$$U_{columb} = \frac{kQ_1Q_2}{r} \tag{3.22}$$

The sum of the Columbic energy between the four hydrogen atoms is not symmetric as the dihedral angle for the optimized $H_5O_2^+$ molecule changes. This is caused by the fact that the optimized central H-O-H bond is not linear but at an angle of about $175^o$. As a result, the energy contribution from dihedral interaction is unsymmetrical. It was very difficult to fit a dihedral potential equation to an unsymmetrical curve, thus it is assumed that the central bond is $180^o$ to obtain a symmetric dihedral curve.

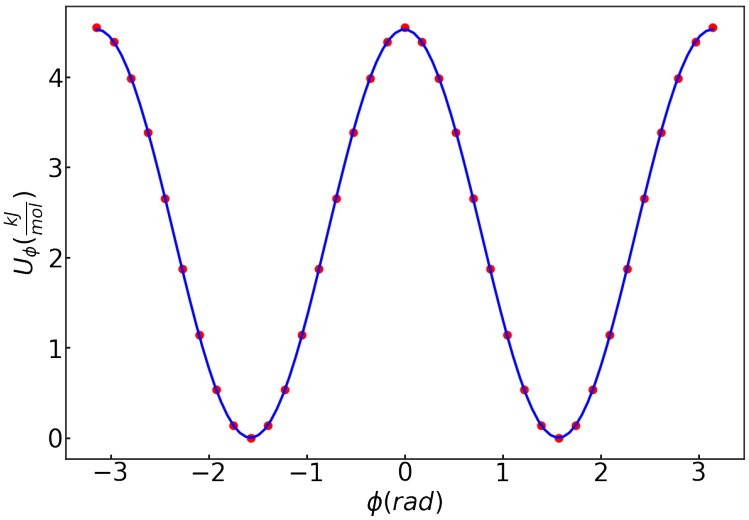



**Figure 3.17:** $H_5O_2^+$ dihedral plot

The resulting curve from assuming linear central bond is shown in **figure 3.17**. As shown, the curve is symmetrical and can be fitted to an equation of the form **equation 3.21** and $k_\phi$ is found to be $2.26\frac{kJ}{mol \cdot rad}$

The force constant for the central bond stretching is then found. The procedure to find the central bond stretching is similar to finding bond stretching of O-H bond in a $H_3O^+$ molecule which involves calculating the energy of the molecule around its optimal bond length. A pictorial representation of the $H_5O_2^+$ central bond stretching is shown in **figure 3.18**

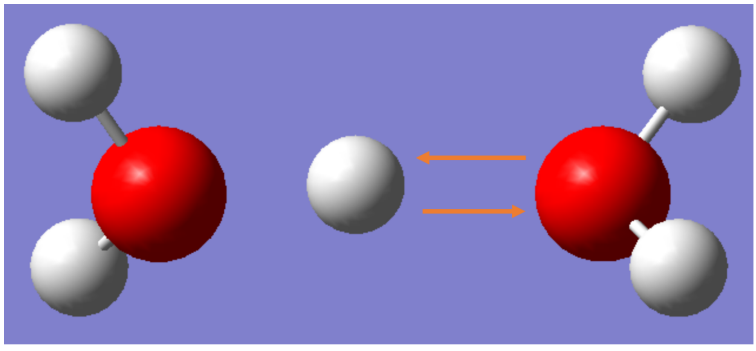


**Figure 3.18:** Central bond scan for $H_5O_2^+$ molecule

The curve will then be fitted to a harmonic potential which is shown in **equation 3.17**. The fitted bond constant for O-H-O central bond is around $121720\frac{kJ}{mol \cdot nm^2}$ and is much lower than the covalent O-H bond constant fitted in the last section for $H_3O^+$. This result is expected since hydrogen

bonding is much weaker than covalent bonding. Finally, the force constant for the central O-H-O angle is fitted. The pictorial representation of the scan is shown in **figure 3.19**

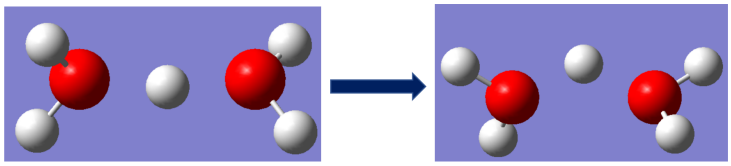


**Figure 3.19:** $H_5O_2^+$ central O-H-O angle scan

The results of the scan is then fitted to the harmonic potential and the angular force constant is found to be $150 \frac{kJ}{mol \cdot rad^2}$. This weak force constant on the central O-H-O angle is expected as the hydrogen bond is much weaker than a covalent bond.

### 3.5.2 MD simulations for $H_5O_2^+$

Similar to $H_3O^+$ and $H_2O$ into water simulations, $H_5O_2^+$ into water simulation follows the same procedure. The time series for $\frac{\partial H}{\partial \lambda}$ is shown below in **figure 3.20**. Since the $\frac{\partial H}{\partial \lambda}$ is oscillating around some average value, the simulation is determined to have reached equilibrium.

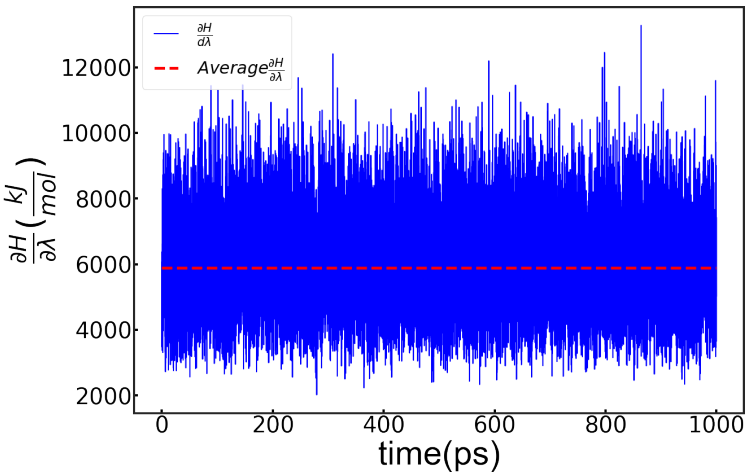



**Figure 3.20:** Time series for $\frac{\partial H}{\partial \lambda}$ of $H_5O_2^+$ solvation into $H_2O$ simulation

Using the $\langle \frac{\partial H}{\partial \lambda} \rangle$ at each $\lambda$, the integrand curve can be constructed as shown below in **figure 3.21**

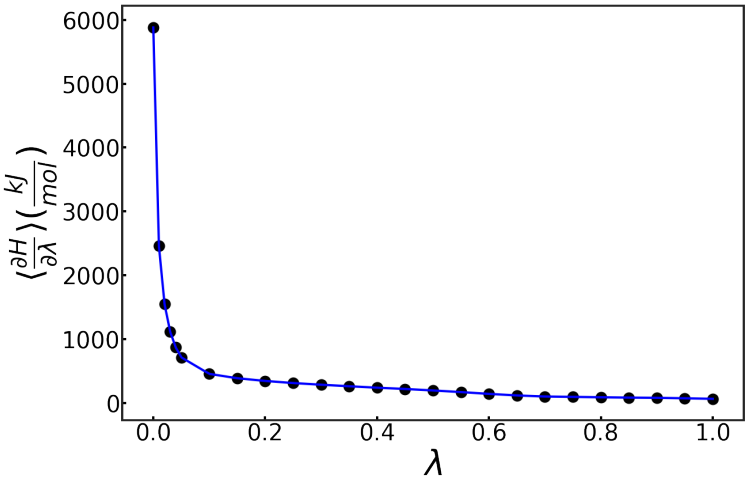



**Figure 3.21:** Integrand curve of $\langle \frac{\partial H}{\partial \lambda} \rangle$ vs $\lambda$ for $H_5O_2$ into $H_2O$ simulation

Using the integrand shown above, the integral curve for the $H_5O_2^+$ into $H_2O$ simulation can be constructed as shown in **figure 3.22** and the solvation free energy is determined to be $-295.77 \frac{kJ}{mol} \pm 0.658$

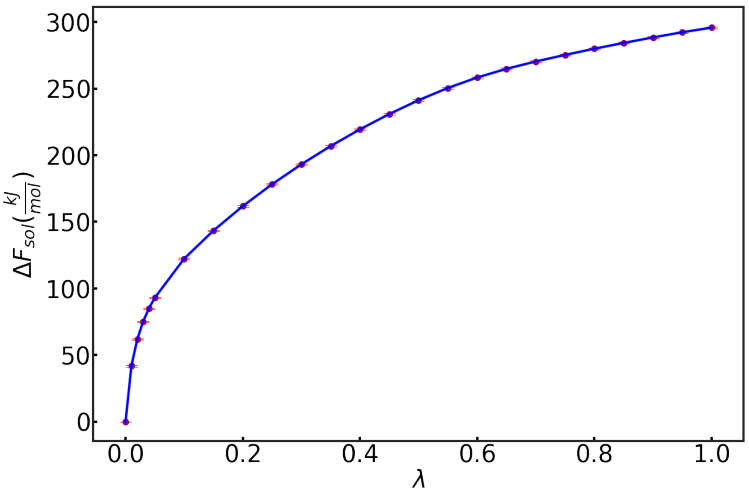



**Figure 3.22:** Integral for $H_5O_2^+$ solvation into $H_2O$

## 3.6 Hess's Law calculation for reaction free energy in aqueous phase

Hess's law states that the free energy change of a reaction is the summation of all free energy changes. In this research work, the reaction free energy is first calculated in the vacuum phase (DFT). Then the reactants as well as the product are solvated into aqueous phase with corresponding $\Delta F_{sol}$. Therefore, using the Hess's Law, the aqueous free energy of reaction is simply the summation of the $\Delta F_{react}$ in vacuum and the $\Delta F_{sol}$ for each of the compound. A pictorial representation of the application of Hess's Law is shown below in **figure 3.23** where $v$ represents vacuum phase and $aq$ represents aqueous phase.

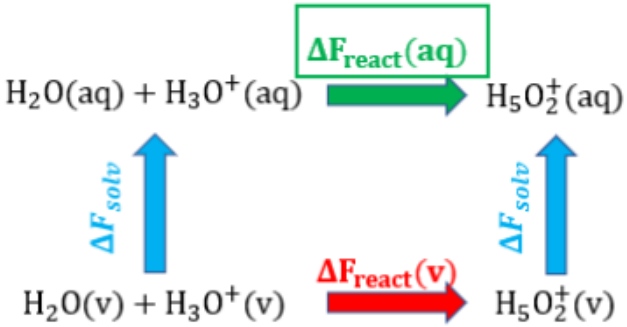



**Figure 3.23:** Hess's Law applied to calculating the aqueous reaction free energy

The results from the simulation with calculation from Hess's Law can then be summarized into the following table

**Table 3.2:** Results from Hess's Law calculation

| | $\Delta F_{formation}$ H2O($\frac{kJ}{mol}$) | $\Delta F_{formation}$ H3O($\frac{kJ}{mol}$) | $\Delta F_{formation}$ H5O2($\frac{kJ}{mol}$) | $\Delta F_{rxn}$ ($\frac{kJ}{mol}$) |
|---|---|---|---|---|
| **Vacuum** | -200736.3521 | -201420.45 | -402268.95 | -112.15 |
| **Solvation** | -27.91 | -346.16 | -295.77 | 78.3 |
| **Aqueous** | -200763.91 | -201766.61 | -402564.72 | -33.85 |

There are few things that are of interest in **table 3.2**. Firstly, the final aqueous reaction free

energy of the chemical reaction $H_2O + H_3O^+ \longrightarrow H_5O_2^+$ is negative indicating that the reaction is favorable. This has proven that proton in aqueous phase prefers to be in a water cluster formation rather than simply on a water molecule to form hydronium ion. The reason that the proton prefers to be in a water cluster form is that it can hydrogen-bond to more oxygen atoms which have partial negative charges and thereby reducing the energy of proton. Secondly, the solvation energy accounted for almost three fourths of the reaction free energy calculated from DFT proving that the solvation correction to reaction free energy is not trivia.

It has been shown in this work that the general structure of a water cluster with two water cluster has a proton hydrogen-bonded to two water molecules. By the same logic, the optimized structure for a water cluster with three water molecules is hypothesized to look like the figure shown in **figure 3.24**

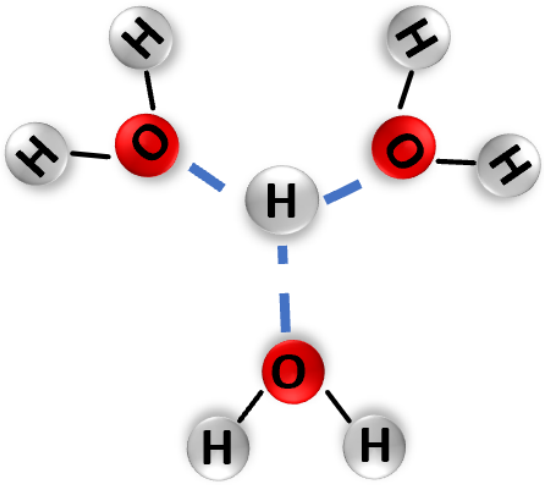



**Figure 3.24:** Three water cluster with proton

However, it is important to note that there is a trade-off between the lowering of energy of the proton and the steric hindrance of the water molecules. Therefore, it is hypothesized that there exists an optimal water cluster size for different systems and will be explored in the future using the method developed in this work.

# Bibliography

[1] Christophe Chipot. *Free Energy Calculations in Biological Systems. How Useful Are They in Practice?*, pages 185–211. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[2] Clara D. Christ, Alan E. Mark, and Wilfred F. van Gunsteren. Basic ingredients of free energy calculations: A review. *Journal of computational chemistry*, 31 8:1569–82, 2010.

[3] I. Kusaka, Z.-G. Wang, and J. H. Seinfeld. Binary nucleation of sulfuric acid-water: Monte carlo simulation. *The Journal of Chemical Physics*, 108(16):6829–6848, 1998.

[4] Adri C. T. van Duin, Siddharth Dasgupta, Francois Lorant, and William A. Goddard. Reaxff: a reactive force field for hydrocarbons. *The Journal of Physical Chemistry A*, 105(41):9396–9409, 2001.

[5] WilfredF. vanGunsteren, Xavier Daura, and AlanE. Mark. Computation of free energy. *Helvetica Chimica Acta*, 85(10):3113–3129.

[6] U. von Barth. Basic density-functional theory an overview. *Physica Scripta*, T109:9, 2004.

[7] Wenlin Zhang, Enrique D. Gomez, and Scott T. Milner. Predicting flory-huggins $\chi$ from simulations. *Phys. Rev. Lett.*, 119:017801, Jul 2017.

# Yusheng Cai

## EDUCATION

**Relevant Courses:** Mathematical Statistics, Statistical Mechanics, Graduate level perturbation analysis, Thermodynamics, Introduction to C++, Introduction to data science using Python

- **The Pennsylvania State University, University Park, PA** 09/2015—05/2019
- **Major:** Chemical Engineering
- **Minor:** Mathematics
- **Penn State Schreyer Honors College**

## RESEARCH&TEACHING EXPPERIENCE

***Undergraduate Researcher,* PSU Chemical Engineering Scott Milner's Lab 01/2018 -present**

- Performing Molecular Dynamics calculations to obtain accurate solvation free energy results.
- Developed Python script to process data output from Molecular Dynamics program.
- Presented research at the undergraduate poster session of 2018 AICHE conference in Pittsburgh.

***Undergraduate Researcher,* PSU Chemistry Xin Zhang's lab 01/2017-12/2017**

- Developed novel chemical fluorescent tools to help understand the mechanism of protein misfolding.
- Worked in a dynamic environment that requires constant collaboration with professors and graduate students.

**Instructional Aid, Chemical Engineering Material Balance course 01/2019- present**

- Collaborated with professor to teach a class of 150 sophomore college students
- Organized and lead two-hour office hour twice every week.
- Designed and created challenging homework problems

***Learning Assistant,* PSU organic chemistry I 09/2017-12/2017**

- Organized and lead 2 Question & Answer based study sessions every week.
- Collaborated with professor to write exam/practice exam questions as well as exam questions.

## Skills

IT: Unix/Linux Systems
Programming: Python, C++, Matlab, Mathematica

## Conferences attended

- 2018 Annual AIChE student conference.