UiT The Arctic University of Norway

Faculty of Science and Technology

Department of Chemistry

Computer Modelling of an Enzyme Reference Reaction for Empirical Valence Bond Simulation

Uthman Seyi Abdulyaqeen

Master's thesis in Molecular Sciences, KJE-3900, November 2020

Abstract

It is extensively recognized that an enzyme functions through reducing the energy of activation of its elemental reaction and the catalytic effect of an enzyme is as a result of change in activation free energy obtained relative to a reference reaction in water of the same mechanism. In this study, chorismate mutase which is an enzyme that catalyzes biochemical reactions for chorismate conversion to prephenate has been used as the model enzyme and its catalytic effect was investigated using Empirical Valence Bond method with employment of Density Functional theory calculation to obtain its reference state reaction. It was found in this study that DFT has been able to produce good activation and reaction free energies compared to experimental values that can serve as a reference to investigate enzymatic reaction of chorismate mutase and the enzyme's catalytic effect was well analysed with EVB as the method was able to reproduce the enzyme's activation and reaction free energies.

Acknowledgements

My profound gratitude goes to my supervisor Professor Bjørn Olav Brandsdal for all the necessary help and feedback throughout the writing process and for giving me the opportunity to be part of his group thereby introducing me to the field of biomolecular modelling. Special thanks to my co-supervisor Geir Villy Isaksen for the support, kick-start and feedback of my master project. I would also like to thank my fellow master students in persons of Bente and Magnus for the warm welcome and support during my master program.

My sincere appreciation also goes to the entire department of chemistry, UiT particularly Hylleraas Centre for Quantum Molecular Sciences.

Lastly, my heartfelt gratitude goes to my dear wife, Ruqayah and my brother Ibrahim, his dear wife Maryam (Iya Gani) and their lovely kids Babawale and Haleemah, I thank you all for giving me enormous contributions and unforgettable supports during my down moments.

Table of Contents

List of Abbreviations

BsCM	Bacillus subtilis chorismate mutase
CM	Chorismate Mutase
DFT	Density functional theory
E	Energy
E_{bbs}	Energy from big basis set
EcCM	Escherichia coli chorismate mutase
E_{gas}	Electronic energy in gas phase
E_{solv}	Solvation energy
EVB	Empirical valence bond theory
FEP	Free energy perturbation
Fs	Femtosecond
IRC	Intrinsic reaction coordinate
LFER	Linear Free Energy Relationships
LIE	Linear Interaction Energy
MD	Molecular dynamics
MM	Molecular mechanics
PS	Product state
PDB	Protein Data Bank
QM	Quantum mechanics
RS	Reactant state
TS	Transition state
ZPE	Zero-point energy
$\Delta G^{\ddagger}$	Activation free energy
ΔG^{0}	Reaction free energy
$\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$	Activation enthalpy and entropy

1 INTRODUCTION

1.1 Enzymes and their functions

Enzymes are macromolecules that act as catalysts in biochemical reactions. In general term, they serve as biocatalyst that speed up the rate of metabolic processes.[1]
The basic reaction is:

$$S + E \rightarrow [ES] \rightarrow [EP] \rightarrow E + P \qquad (1.1)$$

Where: S = substrate; E = enzyme catalyzing the reaction; P = product of the reaction.

Computational modelling of enzymes and simulation of enzymatic reactions have gradually become fundamental in drug design. As many three-dimensional structures of enzymes have been constructed, one of the central encounters in this is to create and clarify the association among enzymes arrangement and catalytic action. The conversion rate of an enzyme can be analyzed with different types of methods, and the essential problem is not only about the stability of the transition state, but to find the contribution to this stability and in what ways it relates to the enzyme structure as well as its catalytic energy.[2]
Specificity of the enzymes make them very much important aspect in the field of research and diagnostic. There are several factors that can affect their activity, these include substrate concentration, enzyme concentration, effect of inhibitors, pH and lastly, temperature effect.[1,3]
In addition, enzymes have their known functional characteristics which include lowering the activation energy, enhancing biochemical reactions and can be reused and also recycled.
Furthermore, there are various classification of enzymes based on their functions, some of which include oxidoreductases which are enzymes that involve in electron transfer from one molecule to another. e.g. dehydrogenases, pyruvate, oxidase etc.; transferases that involve transfer of groups like acetyl, amino etc. among acceptors and donors in the biochemical reaction for example kinase, deaminase etc.; isomerases are enzymes responsible for transferring of functional groups and rearranging of the atoms in any structure and help in

changing of the shape of molecules in biochemical reaction for example chorismate mutase. Others are hydrolases, lyases and ligases.

It has been established that the effect of enzyme catalysis can be related to an appropriate reference chemical reaction in solution of the same mechanism. Quantitatively, catalytic effect of an enzyme is as a result of change in activation free energy relative to a reference reaction, this means, catalytic effect of an enzyme is mainly relative to an uncatalyzed reaction. [1,4,5]

1.1.1 **Chorismate mutase and its mechanism of catalysis**

In enzymology, chorismate mutase is an enzyme that catalyzes chemical reactions for chorismate conversion to prephenate, also known as the Shikimate pathway in the construction of the two important amino acids, phenylalanine as well as tyrosine in vegetation and living organisms.[6] The chorismate mutase is found at a subdivision place in the metabolic pathway. The enzyme in living cells plays an important role in been responsible for the biosynthesis of tyrosine and phenylalanine in order to balance the cell aromatic amino acids.[6,7] Chorismate mutase is naturally occurring and it has been recommended that inhibitors of bacteria containing chorismate mutase may act as antibiotics (e.g. against tuberculosis), the basis is that in the occurrence of these inhibitors, the bacteria cannot synthesize phenylalanine. Although it is known as a promising antibacterial target based on metabolic role. The common crystallized structures of CM are the ones obtained from bacteria Escherichia Coli and Bacillus subtilis and sometimes from yeast. CM is among the few common catalyst for biochemical pericyclic reactions and it belongs to the family of isomerases which are functional group transferring enzymes.[6,8,9]

The first step in its mechanism of action is the conversion of the chorismate to prephenate. This is done to produce the amino acids - tyrosine and phenylalanine. It has been established that reactions catalyzed by CM has only one substrate as the RS which is chorismate and only one product which is prephenate. The common study of the intramolecular rearrangement of the enzyme is the one that catalyzes Claisen rearrangement of chorismate to prephenate.[6,10] The rearrangement process of the enzyme for bond forming and bond breaking start from making the substrate to be in a diaxially reactive form, then formation of carbon – carbon bond between carbon 1 and carbon 9, then the ether bond breaking between oxygen 7 and carbon 5 of the substrate.[6] It has been suggested that the rearrangement occurring in the active site of the

enzyme with stability of the enzyme might be due to interaction between the two carboxylate group in the substrate and two arginine residues in the active site of the enzyme, the sequence number of the two arginine residues might be different depending on the source of the enzyme.[6,9] CM is known to have more affinity towards the TS than the substrate and does not transform itself or and does not have effect on the surrounding solvent during the process of isomerization of the substrate to form the product state. But if the catalyst is not present in this reaction then it will proceed in a concerted step, and in an asynchronous way. The mechanism given below is adapted from Advances in Protein Chemistry and Structural Biology[11]

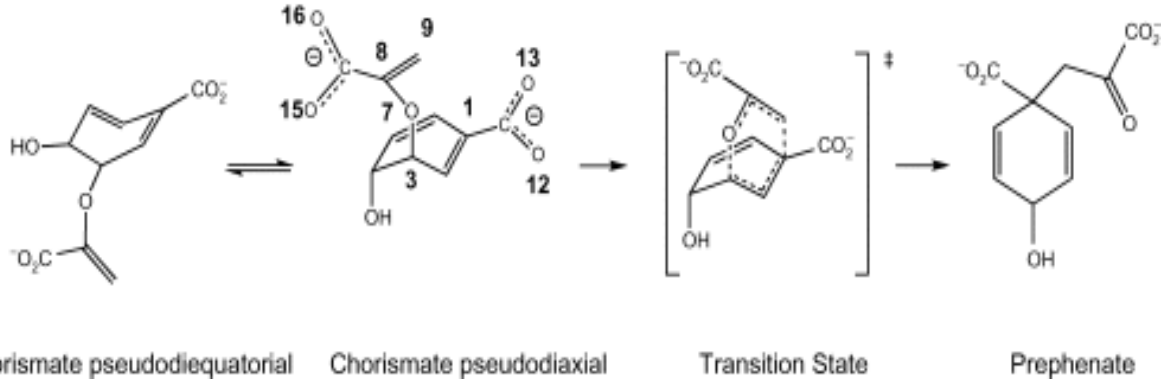

Figure 1: Mechanism of conversion catalyzed by chorismate mutase.[11]

1.2 Free Energies

The concept of Gibbs free energy in chemistry is understood to be a change in the energy that occurs when all the products and reactants are in the standard conditions along with the pH value of 7.[12–14] This can be expressed as:

$$\mathbf{G = H - TS} \qquad (1.2.1)$$

where G = Gibbs free energy, H = enthalpy, T = temperature, S = entropy

At standard temperature and pressure, almost all systems want to attain a minimum of free energy. So, this will increase the entropy that reduces the Gibbs energy. Likewise, if the availability of heat is excess then it will reduce the free energy by reducing the enthalpy. Hence the cells will have to work with the laws of thermodynamics and chemical reactions too. [12–14]

For the change in energy the equation (1.3.1) above changes to

$$\mathbf{\Delta G = \Delta H - T\Delta S} \quad (1.2.2)$$

From the above equation, there will be three possibilities:

- $\Delta G < 0$, this means reaction will proceed
- $\Delta G = 0$, equilibrium conditions;
- $\Delta G > 0$, reverse reaction

For any reaction, ΔG can be determined using

$$\Delta G=\Delta G°+RT\ln\left([\frac{[B]b}{[A]a}]\right) \quad (1.2.3)$$

Where [A] and [B] are concentration of reactants and products respectively.

The above equation can be modified into

$$\Delta G=\Delta G°+RT\ln\left([\frac{\text{products}}{\text{reactants}}]\right) \quad (1.2.4)$$

For any reaction $\Delta G°$ and T are constant, ΔG will always depend on ratio of products and reactants. If the ratio of the product and reactant increases, then the value of ΔG also increases in positive direction.[12,14]

1.3 Transition State Theory

Transition state theory gives an insight about the rates of chemical reaction. This theory is used to predict the Gibbs free energy of activation, standard enthalpy of activation, and standard entropy of activation of the reaction if the constant rate has been evaluated experimentally or by quantum mechanical calculation. The main assumption of this theory is that, a chemical reaction follows a distinct reaction path that involves breaking and forming of bonds. Moreover, energy is needed in breaking and forming of bonds, therefore, a reaction must have an activation energy to overcome, this implies that the collision between the reactant molecules will result in formation of a product. The occurrence of the activated complex is the point at which the reaction potential energy is the highest.[4,13] Since the kinetic energy of the colliding molecules will have necessary energy and orientation which is now transferred as potential energy. Hence the definition of activation free energy is the difference between the transition

state of reaction and the ground state of reactants. Transition state theory is widely used in the field of computational chemistry where it is mostly used to model the reactions which are influenced by the catalyzers.[4]

1.4 Computational Modelling Tools

1.4.1 Quantum Mechanics / Molecular Mechanics (QM/MM)

The biochemical system which includes enzymes are very large data, and it is difficult to predict their properties and catalytic action with the use of QM technique only. Also, the available force fields of the molecular mechanics are not so much flexible to process the model that contains the chemical bonds (addition and deletion). In order to overcome the limitation, then comes the combination of full QM rules which is at one end and full MM which is at the other end, certain other QM/MM methods are also developed which can consider the small part of the whole system as minute particle as QM in chemistry and it has to retain the MM part which is needed for the computational purpose.[15–17]

The purpose of dividing the system into regions might be that some parts are described with different theory and have character in a condensed phase. The timescale for the computation is also reduced in order to save the time to predict the solution. Hence the QM/MM has a potential to lead the chemical reaction to a better understanding. Hence, it can be differentiated in a way like reaction center and spectator which are directly responsible for the chemical reaction and which are not responsible for the chemical reaction directly. A small region in which a chemical reaction occurs and therefore cannot be described with a force field is treated at a sufficiently high level of QM part and the remainder of the system is modelled at the MM level. [15–17]

1.4.2 Molecular Mechanics

A branch of classical mechanics which involves application of Newtonian mechanics that are used to describe the physical basis of any given model. Any molecule consists of an atom which has protons and electrons and are considered as point charge with some mass. [15–19] The interaction between the neighboring atoms can be compared to a spring action which tells about

the chemical bonds and also about the van der Waal's forces. The interaction of the atom which are due to electrostatic forces can be determined using the equation of Coulomb's law. The velocity at which the atom rotates are inside the molecule and depends upon the temperature, and other macroscopic quantity. The potential function that describe a molecular system is related to the internal energy and all the thermodynamic quantities. The potential function can be defined as: [15–19]

$$E_{total} = E_{bonds} + E_{angle} + E_{dihedral} + E_{non\text{-}bonded} \quad (1.3.3.1)$$

$$E_{non\text{-}bonded} = E_{electrostatic} + E_{van\text{-}der\ Waals} \quad (1.3.3.2)$$

Where E is total energy in the system which is the summation of bonded energy, energy of angle, dihedral energy and non-bonded energy. The non-bonded energy is composed of the electrostatic energy and Van der Waal's energy. But these energies which form total energy of potential are commonly known as force fields. Different molecular approaches are used to find the different mathematical results through the force field.[18]

The following are some of the principles which molecular mechanics (MM) are based upon:

- The bonds present in between the particles are treated as harmonic oscillators.
- Nuclei and electrons are both treated as single particle like a round structure.
- Different potential functions are required to describe the different interactions like bond stretching, angle bending, torsional etc.
- The potential function is mainly dependent upon the empirically derived data like force constants of the equilibrium etc. which can tell the user that atom or set of atoms are in the ideal state or the natural state.
- The summation of the particle interaction of the spatial distribution of atoms are equal to the measure of the intramolecular strains.[19]

The MM approach is used extensively in order to extract the result which has structure, dynamics, properties of surface, also the thermodynamics of organic and it extends to biological and polymeric systems too. In molecular science, the main activity that has been used as MM method is that of protein folding, catalysis of the enzyme, stability of the protein and changes which are conformational and associated with the function of the biomolecule etc.[19] From a figure adapted from Odegard et al, showing the MM interactions between atoms.[20]

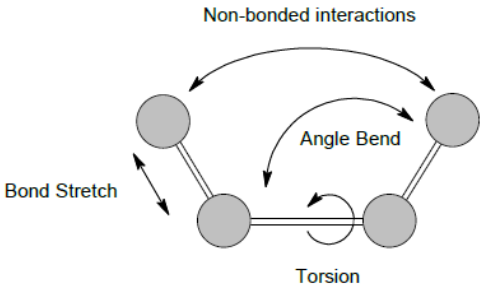

Figure 2: MM different interactions between the atoms.[20]

The above figure gives an insight about the necessary forces which are present inside the molecule and between the atoms which are connected to the single central atom. The potential energy of a molecular system is the summation of the individual components which are potential enough to get the energy and this includes, bond length, stretching, angle bending, torsion, and remaining van der waals interaction forces. And is given by: [16–18,21]

$$E = \sum E_{stretch} + \sum E_{bend} + \sum E_{torsion} + \sum E_{VDW} + \cdots. \qquad (1.3.3.3)$$

1.4.3 **Force fields**

Force fields are the mathematical expression which describe the energy dependency of a system on the coordinates of the constituent particles.[18] The interatomic potential energy are written in analytical form like U(**r**1, **r**2, . . . , **r**N) and also the parameters along with it. The parameters are typically obtained either from ab initio or semi-empirical quantum mechanical calculations or it can also be obtained by fitting the data of the experiments such as neutron or X ray etc. Force fields are always designed purposefully to describe molecular properties accurately. They play fundamental role in molecular modeling and are used to describe all the parameters that contribute to potential energy of molecular systems. Molecules can be the atoms which are held together by a force like simple elastic and the true potential with a simple model in that region are simulated. A typical force field can be expressed as: [16,18]

$$U(r^N) = \sum_{bonds} \frac{1}{2} k_b (l - l_0)^2 + \sum_{angles} \frac{1}{2} k_a (\theta - \theta_0)^2 + \sum_{torsions} \frac{V}{2} n[1 + \cos(n\emptyset - \delta) + \sum_{improper} V_{imp} + \sum_{LJ} 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}} \right) + \sum_{elect} \frac{q_i q_j}{r_{ij}} \qquad (1.3.4.1)$$

In the above equation, the first four terms refer to the local contribution of the total energy or also called intramolecular forces. This may include bond stretching, angle bending, dihedral etc. and last two terms indicate the repulsive and Van der Waal's interaction forces.

Bond Stretching Energy

Stretching of the atoms takes place between two neighboring atoms. This follows the Hook's law in nature where it says stress is directly proportional to the strain.[17,18] In the molecular mechanics, the atoms are treated as spheres, and bonds which are present between the atom are treated as springs. Hence the hooks law can be applied between them and equation is as follows:

$$F = -kx \qquad (1.3.4.1)$$

where k=stiffness and x is the displacement due to force F. [17,18]

Expressing the above equation further,

$$V = -\int_{x1}^{x2} F(x)dx \quad = \int_{x1}^{x2} (kx)dx \; = 1/2kx^2 \qquad (1.3.4.2)$$

In the above equation (i), V = potential energy and f(x) is displacement function. The limits are given from x1 to x2 and that represents the system x1 and x2.

Hence for the bond stretching the equation above changes to:

$$\boldsymbol{E_{stretch}} = 1/2k_s\ (l - l_0)^2 \qquad (1.3.4.3)$$

K_s is the force the stretching and is the constant value which are determined empirically, l is length of the bond, which is presented in the chemical equation, lo is the natural bond length.

Angle Bending Energy

Energy occurs as a result of change in angle between neighboring atoms that are bonded to the same central atom. [17,18] The equation is given as:

$$E_{bend} = 1/2k_b(\theta - \theta_0)^2 \qquad (1.3.4.4)$$

In the above equation, k_b is the stiffness constant and obtained empirically. θ is the actual bond angle and θ_0 is the natural bond angle. For example: optimal bond angle for H-Csp2-Csp3 is 122°, then any change in angle (θ-θ_0), either wider or narrower, will increase the energy of the molecule. The angle-bending term is summed over all bond angles in the molecule. [17,18]

Torsion Energy

There is energy contribution that arises due to angle twisting between rotating atoms which are connected to the same central atom. The equation which is required to find the energy is given by: [17,18]

$$E_{torsion} = 1/2V_0(1 + \cos n\omega) \qquad (1.3.4.5)$$

In the above equation:
V_0 is the barrier which is responsible for the natural bond to rotate freely, n is the number of the period of the rotation or number of cycles for 360-degree turn.

Non-bonded interactions:

The atoms in a molecule which are present as far as two bonds can interact with each other through many of the attraction forces like, van der Waal's force, steric repulsion, electrostatic repulsive or attraction which will be depending on the distance between them. For the non-bonded atoms, which are far away from each other, the interaction followed in order to describe is London dispersion force and once it reaches near other atom then it will start to repel each other and Van der Waal forces or strain or steric strain is used. [16–18,21]

1.4.4 **Molecular Dynamics**

The principle of the molecular dynamics (MD) simulation technique is to integrate the equation of motion by giving the trajectory for the n number of particles in any molecular system. MD

simulations is of pivotal importance for conformational sampling of the chemical reaction coordinates in solution and in the active site of the corresponding enzymes. This also uses newton's equation of motion.[22–24] To get solution using this approach, it requires the initial conditions which includes mainly the position and velocity at which the particles are moving along with the position and valid model which represents the force field between the particles and the boundary conditions must also be employed. With the Newton's equation of motion (1.6.6.1) and recall that the force acting on a particle is the potential energy, which is related to the force field, so the equation can be further expressed as equation (1.6.6.2)

$$F = ma \tag{1.4.6.1}$$

$$m_i \frac{d^2 r_i}{dt^2} = f_i = -\frac{\partial}{\partial r_i} U(r_1, r_2, \ldots., r_N) \tag{1.4.6.2}$$

In the above equation, the term U($\mathbf{r}_1$, $\mathbf{r}_2$, . . . , $\mathbf{r}_N$), refers to the potential energy which are written in the analytical form and also it is the dependency of the energy for n particles. The above-mentioned equation is a coupled second order differential equation which cannot be solved analytically and can take numerous times to extract solution, hence numerical approach is followed in order to get the solution. The use of MD techniques such as FEP, EVB and LIE to predict properties and energies of biomolecules have been well established.[22–24]

1.4.5 **Density Functional Theory (DFT)**

DFT is a computational simulation method that derives properties of the molecules based on determination of the electron density of the molecule.[25,26] Practical approach on DFT are coded into efficient computer programs and been simulated to assist in solving chemical problems and to know the properties of molecules.

It has been established that, to find a suitable approximate way for solving the Schrödinger equation to get information about a molecule, the components of that system are described by functions. These functions give all the information of a system.[25,27] Solving Schrödinger equation to get the functions of a simple system is possible but difficult for molecules having more than one atom. Starting from Born-Oppenheimer Approximation which assumes that the motion of nuclei and electrons in a molecule can be separated, it is necessary to find

approximate solution to solve the equation to get information about molecules. Approximate simple representation of the functions as electron density form the basis of DFT.[25,27] This method was developed from the theorem of Hohenberg-Kohn which describes that all the ground-state properties of a system can be determined by the density of that system.[25] Therefore, total energy of a molecule can be obtained from functions of its electron density since all ground-state properties are functions of electron density.[25,28] In DFT, Functional can be defined as the function to the function and here the functional means the energy of the molecule to form an electron density. The electron density is the function of three axis and its variables are taken as x, y and z position of the electrons.[25,26]

There are some forms of approximate corrections to DFT, namely: Local density approximation, Gradient corrected and hybrid methods. In LDA, the assumption is that the density of the molecule in the local is equal. Gradient corrected takes care about the non-uniformity of the density in the molecule. The most common method used in the hybrid method is B3LYP which means Becke 3-term correlation functional (Lee, Yang, and Parr exchange function), others are BPV86, CAM-B3LYP, PBEPBE.[28–30] The most important advantage of DFT is the more increase in computational accuracy without the additional increase in the run time of the tools or package. B3LYP/6-31G(d) are oftentimes considered to be a standard model in chemistry for many applications. [28–30] One of the main disadvantages of DFT methods is the challenge in determining the most appropriate method for a particular application. The practitioner should, prior to choosing a DFT method, consult the literature to determine the suitability of that choice for that particular problem or application. As such, DFT usage tends to favor the more sophisticated user. [28–30]

B3LYP: it is one of the old hybrid theories developed early 80's**.** B3 is a function which combines three functions of the Hartree-Fock exchange correlation and LYP which is a function that recovers the dynamic electron correlation. [28–31]

LDA: a type of hybrid theory of approximation which are for the exchange correlation of the energy for the potential function in the DFT and only depend upon the density of the electron locally in the molecule.[31]

CAM-B3LYP: Coulomb-attenuating method-B3LYP is hybrid method which combines the B3LYP hybrid qualities present in it with the long range correlation to get desired output.[31]

1.4.6 **Free Energy Perturbation (FEP)**

It is an alternative binding free energy extraction approach which are associated with the conversion of the two states of QM and MM. These two are used in the thermodynamic integration (TI) and the Hamiltonian is represented as interpolation between the states and the coupling parameter is taken as λ.[32,33]

$$H(q,p,\lambda) = (1-\lambda)\mathrm{Ha(q,p)} + \lambda\,\mathrm{Hb(q,p)} \qquad (1.4.6.1)$$

$$\Delta G = \int_0^1 \langle \frac{\partial H}{\partial \lambda} \rangle \lambda\, d\lambda \qquad (1.4.6.2)$$

Where "q and p" are position and momentum of the atom which is present in the system. To find out the free energy difference between the state A and state B when λ =0 and λ = 1 respectively, the system must be sampled for the fixed values of λ between 0 and 1 which is followed by the integration over a system.

This method uses application of statistical mechanics to calculate change in binding free energy as a relative between two states. This is obtained from Zwanzig formula: [32,33]

$$\Delta G = -KTln\{e^{-(U_B-U_A)/KT}\}_A \qquad (1.4.6.3)$$

This method categorises a state as the reference system and the other state as target, therefore an intermediate state can be introduced to properly describe these states.[34]

$$U_m = (1-\lambda_m)U_A + \lambda_m U_B = U_A + \lambda_m \Delta U \qquad (1.4.6.4)$$

Where λ_m is intermediate state introduced[34]

$$\Delta U = U_A - U_B \qquad (1.4.6.5)$$

$$\Delta G = G_B - G_A = -\beta^{-1} ln\,\{exp(-\beta \Delta U)\}_A \qquad (1.4.6.6)$$

Since $\Delta U = U_A - U_B$, therefore,

$$\Delta G = G_B + G_A = -\beta^{-1} \sum_{m=1}^{n-1} ln\,\{exp\,[-\beta(U_{m+1} - U_m)]\}_m \qquad (1.4.6.7)$$

Where $\beta = K_bT$

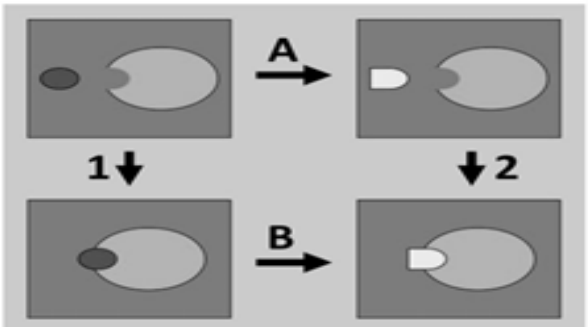

Figure 3: The change in binding free energy is explained in the thermodynamics cycle below:[32,33]

$$\Delta\Delta G_{bind} = \Delta G_1 - \Delta G_2$$

Or $$\Delta\Delta G_{bind} = \Delta G_A - \Delta\Delta_B \quad (1.4.6.8)$$

1.4.7 **Empirical Valence Bond Theory**

EVB is a computational technique based on an approximation that enzymes reaction free energy can be determined via a calibrated Hamiltonian which is compared to transfer of electrons that occur during chemical reaction in a solution.[35] The Empirical-Valence-Bond (EVB) technique can deliver consistent predictions about enzymatic action and mutual mutations. It is a semi-empirical QM/MM method developed by Warshel to study enzymatic reactions energy surface.[36] The main and fundamental feature of this procedure is its use of exclusive standardization probabilities to insert calibrated realistic examined data into Hamiltonians, especially information about relevant solutions that often deliver a method to sort out the "errors pertinent to the model".[16,21,35,37] Starting with the simple reference reaction of the solution, this fragmentation process is required because quantum mechanics approaches cannot grasp a phase where the enzymatic reactions calculation can be trusted upon, especially in a case of potential reaction. This process follows the example of the transfer of the proton followed by the evaluation of the surface free energy.[16,21,35,37]

In combination with classical valence-bond constructions, EVB initiates with resonance states (or more exactly diabatic states). These root states are assorted to define the reactive arrangement.

$$\varphi = c_1\emptyset_1 + c_2\emptyset_2 \qquad (1.4.7.1)$$

Where $\emptyset_1$ corresponds to the RS and $\emptyset_2$ corresponds to the PS and c_1 and c_2 are linear combination balancing coefficient.

The potential strong suit of the diabatic states and the terminologies are characterized via Hamiltonian matrix features.[16,21,35,37]

$$H_{ij} = \varepsilon_i = \alpha_{i\,(gas)} + U_{i(int)}(R;Q) + U_{i(int)}(R;Q;r;q) + U_{i\,(solvent)}(r;q) \qquad (1.4.7.2)$$

$$H_{ij} = A\exp(-a|\Delta R|) \qquad (1.4.7.3)$$

Here R as well as Q signify the molecular coordinates and potential charges of the participating products as well as reactants ("designated as solutes") respectively, in the diabetic state, and the coordination and charge ("solvent") of the water or protein is designated as r as well as q. $\alpha_{i\,(gas)}$ is the energetic entity of the diabetic stage at the gas level at “ith” stage; Where all fragments value is considered as infinite; $U_{i(int)}$ (R; Q) is the intrinsic potential (potential of inter molecular particle) of the solvent structure in this state (comparative to the minimum). The $U_{i(int)}$(R; Q; r; q) characterizes the contact among the solvent atom as well as the nearby atom of solvent; and $U_{i(solvent)}$ (r,q) signify the energy-potential of the solvent.[16,21,35,37]

For a two states EVB, the ΔG of moving from the ground state potential in the path of a reaction coordinate in which $\Delta G^{\ddagger}$ of the potential energy surface is obtained.

$$\mathrm{E_g} = \frac{1}{2}\mathrm{H_{11}} + \mathrm{H_{22}} - \sqrt{(}\mathrm{H_{11}} - \mathrm{H_{22}})^2 4\mathrm{H_{12}^2}) \qquad (1.4.7.4)$$

Where E_g is the ground state potential, H_{11} and H_{22} are modelled potentials with force field description corresponding to RS and PS valence bond configurations.

The main feature of the EVB method is the calibration of α and H_{ij} by a reference reaction, obtaining the calibrated values as a comparison to transfer of electrons that occur during

chemical reaction in water which in turn can be suitable to analyse potential energy surface of an enzymatic reaction.

Additional essential improvement of this method is its capability to afford a diabatic free-energy surface of the pertinent structure of resonance. This permits anyone to inspect, e.g., the infinitesimal length of time of Linear Free Energy Relationships (LFERs) inside the protein. In addition to EVB Simulation and experimental study suggests these LFERs may be also effective at enzymatic activated locations (if they simply imply the strength of the inner structure of resonance relatively than the energy of the products and reactants). Adopting the inherent concept of the process of LFERs to protein molecules can, at minimum as an approximation, permit one to categorize the effect of catalyst in modest and influential ways.[35] EVB is less time-consuming and allow cost-effective conversion design which is fundamental in enzyme study. The shortfall of this method is getting the reference reaction to calibrate its Hamiltonian, the reference can be obtained from experimental analysed data, but for some reaction which has no available experimental data, then the EVB reference state can be obtained from DFT calculation.

1.4.8 **Molecular Docking**

Docking is used in ligand binding to detect the binding mode and identify the best binding site of a receptor.[24] Docking plays an important role in predicting structures of enzymes since the binding pocket of protein is completely surrounded by amino acids, suggesting that the protein undergoes a significant conformational change in ligand binding and product release change. With this scenario, docking has a scoring function that rank poses which are generated. The scoring function can be empirical based, force field based obtained from existing sets of parameters or a kind of design that recreates structures from experiments. There are different established programs used for docking, common among them is GLIDE[38] program which is grid-based that allows receptor grid generation, this generated grid allows molecules to be properly docked. GLIDE docking program has two different precision set which are Standard precision (SP) and Extra precision (XP).[34,38,39]

1.5 Aims of the Project

The overall aim of this project is to initiate the development of a small database with calibrated EVB Hamiltonians for enzymes catalyzed reactions. The enzyme, chorismate mutase has only one substrate which is chorismate and only one product which is prephenate. DFT calculation was carried out for the uncatalyzed conversion of chorismate to prephenate and the activation and reaction free energies obtained were used as reference for calibrating the EVB Hamiltonian for EVB simulation to investigate catalytic effect of chorismate mutase. The calibrated Hamiltonian was used in uncatalyzed conversion of chorismate to prephenate in water solution, then the calibrated EVB Hamiltonian was tested in the chorismate mutase to verify catalysis. The uncatalyzed reactions was carried out at five different temperatures. The EVB simulations were also able to give activation and reaction enthalpy and entropy for the reaction. The tasks in this project are divided into three parts to analyze free energies of an enzymatic system. The first task is the use of a quantum mechanical based technique (DFT) to obtain a reference state reaction while the second and third tasks are employment of MD that treats the molecular system majorly with MM techniques. For the MD parts, EVB method was first used to reproduce the uncatalyzed $\Delta G^{\ddagger}$ and ΔG_0 then after protein preparation and docking of the substrate and the enzyme, the method was later used to test for catalysis in chorismate mutase (enzyme catalyzed reaction). More so, Arrhenius simulation was carried out to construct Arrhenius plot that analyze the effect of each temperature on the activation free energy of the reaction.

Finally, carrying out all these procedures make it easier to be more familiarized with state-of-the-art computational modelling techniques that are used to model the action of biological macromolecules and their interactions with inhibitors and substrates.

2 METHODS

2.1 DFT calculation

DFT calculation was carried out to predict free energy of activation and reaction for the uncatalyzed conversion of chorismate to prephenate achieving the desired transition state. The $\Delta G^{\ddagger}$ and ΔG_0 obtained served as reference state to calibrate EVB Hamiltonian. The molecular structure of the only substrate in chorismate mutase (chorismate) was prepared with few water molecules around it, then the relevant bonds were frozen except the bonds that are directly involved in the reaction to make a transition state guess. All the DFT calculations were carried out using a computational program package called GAUSSIAN 09.[40] GAUSSIAN is an electronic software which is capable of making prediction of the properties of the atoms like molecular energies, structure, electron density, etc. by using incorporated theories like density functional theory, semi empirical theory, molecular mechanics and also hybrid methods.[40] All the DFT calculations were carried out with each of the different functionals of DFT which are B3LYP (with empirical dispersion gd3), CAM-B3LYP, and PBEPBE using 6-31g(d,p) as basis set except for Single Point Energy calculation of big basis set in which 6-311+G(2d,2p) was used. MOLDEN[41] application was used for structural visualization throughout the DFT calculations.

2.1.1 Geometry Optimization

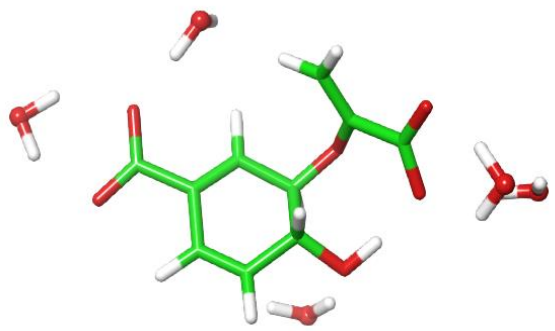

Figure 4: Structure of modelled chorismate

Starting from the chorismate structure above, the following were carried out for the DFT calculation. All these calculations were done in gas phase. Each of the steps were visualized with MOLDEN program to be sure of achieving the desired structures of each of the reactant state (RS), product state (PS) and transition state (TS).

- To get the fair approximation of the transition state, Geometry optimization of TS guess was done on the chorismate with all the relevant bonds frozen except the atoms that are involved in the chemical reaction coordinate.
- Then, a TS calculation was done on the molecule with no bond frozen
- After the TS optimization converges, frequency calculation was carried out to check for desired transition state vibrations which must be attained, if not, then must be repeated till it is achieved.
- To trace the path of the chemical reaction from TS to PS and RS, Intrinsic Reaction Coordinate calculation were carried out in two paths (forward and backward) respectively.
- After the IRC calculations have been able to trace PS and RS paths, geometry optimizations were carried out on the PS and RS.

The reaction from RS to PS is shown below in the figure.

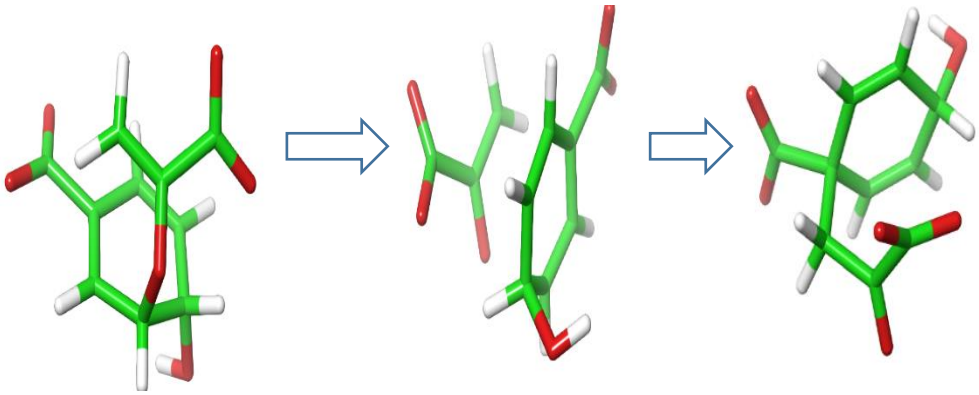

Figure 5: Conversion from RS, TS and PS.

2.1.2 Single Point Energy and Free Energies Calculations with Corrections

The following calculations were carried out to obtain the $\Delta G^{\ddagger}$ and ΔG^{0} values.:

- Frequency calculations were carried out on each of the geometrically optimized TS, PS and RS. The frequency calculations were able to give the values of free energies, enthalpy, entropy and the correction to the Gibb's free energy which is the Zero-point energy (∂G_{ZPE}).
- SPE calculation was carried out in gas phase for each of the TS, PS and RS to get the electronic energy of each of the state.
- Then SPE calculation of bigger basis set of 6-311+G(2d,2p) was carried out on each RS, TS and PS.
- In other to correct for solvation effect, solvation energies were also calculated for each state using water as solvent with two separate Universal Solvent Models[42] (firstly SMD model, then with CPCM model).

To get the activation and reaction free energies, there is need to correct for energies of solvation, big basis set and Zero-point energy (ZPE). Corrections to energy from big basis set and solvation for each state were obtained in relative to electronic energies in gas phase. All the energy values from the TS and PS that correspond to energies of activation and reaction respectively were obtained in relative to the RS.

The $\Delta G^{\ddagger}$ and ΔG^{0} values obtained from the B3LYP functional of the DFT were used further in this study and served as reference for calibrating EVB Hamiltonian.

2.2 MD Simulation Tool

An MD software package called "Q" was employed for the EVB simulation. Q is used for predicting free energies of biomolecules based on the approaches of free energy perturbation, empirical valence bond and linear interactions energy calculation of receptor and ligand binding properties. The program is being used to predict accurate binding free energies in comparison to experimental values.[43]

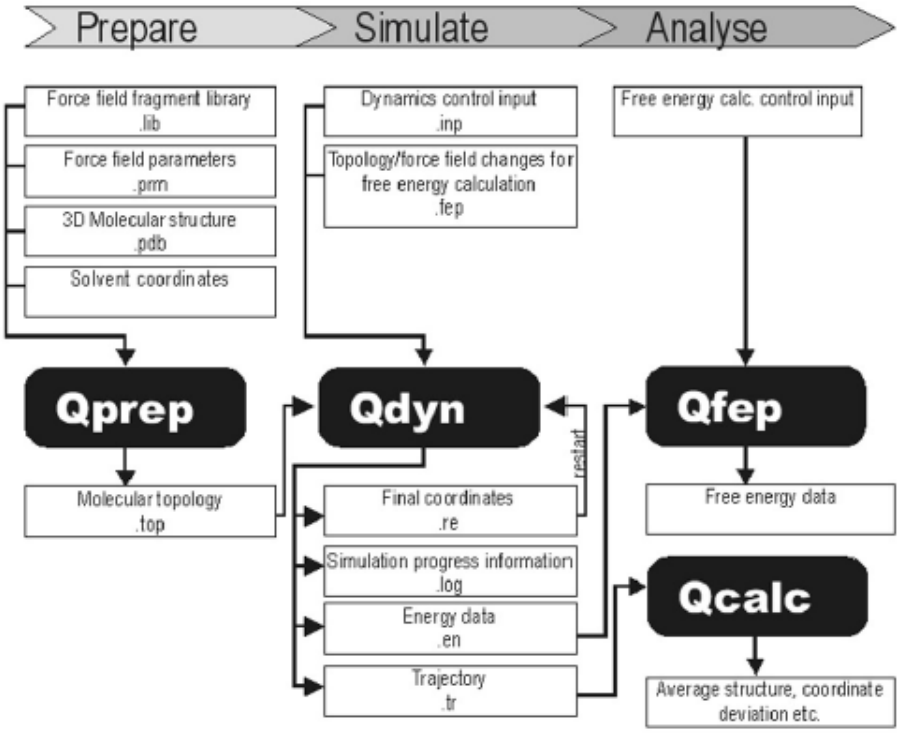

Figure 6: Flowchart for calculating free energy in Q adapted from manual for the Q MD package[44]

There are three main steps in order to solve the problem in Q. first is to prepare, then simulate and analyze the files. In the preparation step, the force field must be given in the model followed by the solvent coordinates which will give the molecular topology. In the simulation part, control of the dynamics input has to be given as input followed by the topology or the force field as input for the calculation of the free energy. Next step the free energy output obtained in the last step must be fed into this step and calculation must be done based on this. The force fields are directly available in the software directory. AMBER95, AMBER/OPLS, OPLS-AA etc. are some of the available force field in this software which can be used directly. The program has PyMOL[45] incorporated in it for visualization of modelled structures and also uses OPLS2005[46] force field server to generate force field parameters and library files that are missing during the simulation.

The Q program consists of four functional part which contains all the libraries and essential parameters to carry out MD simulations. They are Qprep, Qdyn, Qfep and Qcalc. Qprep part is used in the program to prepare the molecule by creating a topology which describes the connectivity of the molecules and their interatomic potentials (e.g. bonds, atom types, atomic

coordinates, atomic charges) and other force field parameters. The Qdyn part specifies the simulation data (steps, step-size, temperature, trajectories, energies) and carries out the simulation. The Qfep part analyses and uses the input file created from Qdyn part to make the free energies to be calculated. The Qcalc part helps to analyze the structure and give more details of the simulation. Q uses Qgui as its graphical user interface.[47]

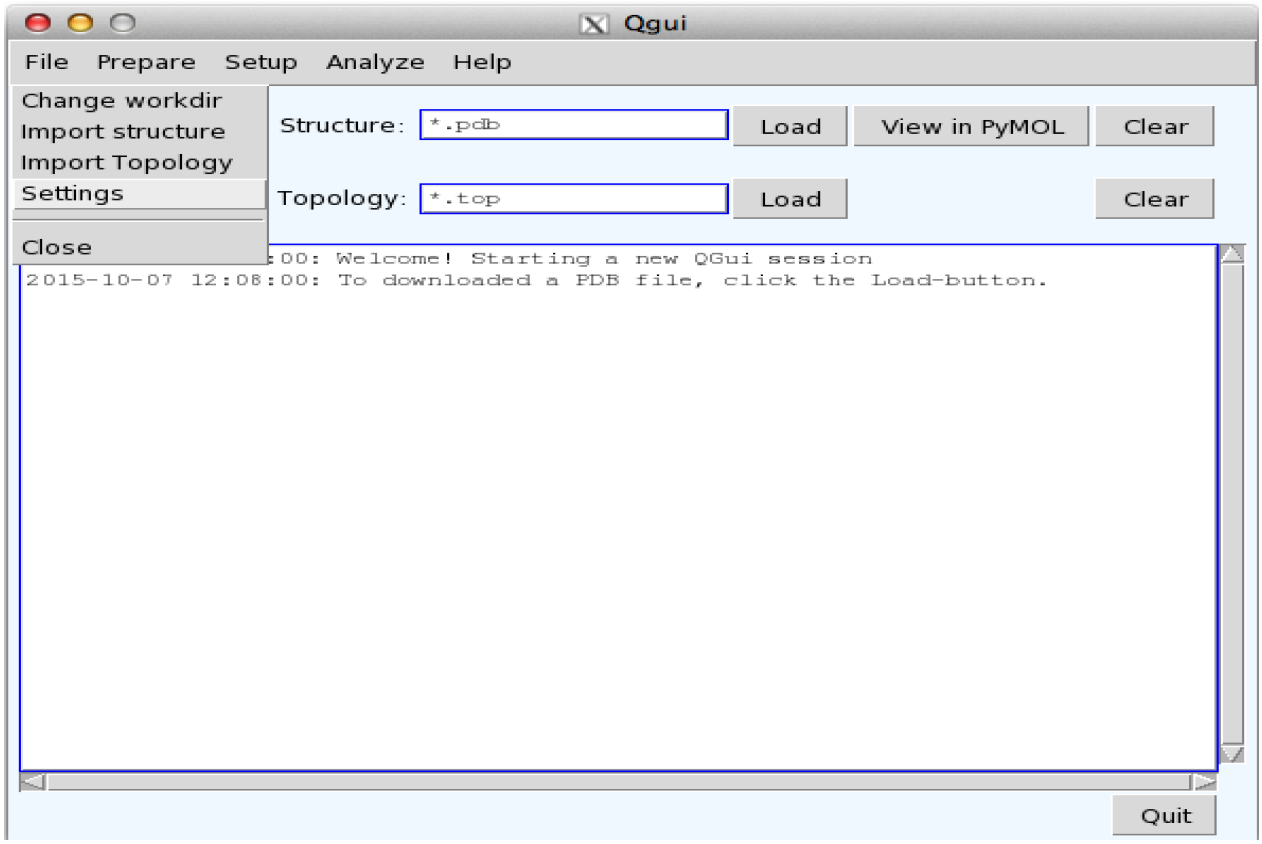

Figure 7: Illustration of the Qgui main window[47]

The Qgui main window contains all the menu for Q to work. Under the prepare menu, topology of the molecule to work with can be prepared and parametrization of the force field and library files. The setup dropdown menu contains the four methods of calculation which are MD, LIE, FEP and EVB in which the user perform transformation of the FEP state and choose PyMOL to visualize modelled structure. Under the analyze menu, the user works on parametrizing the reference reaction and calibration to give the target free energies, the EVB simulation is analyzed to give all the thermodynamic details and their comparison. In this project, the Q software has been employed for the EVB simulations.

2.2.1 Uncatalyzed Reaction with EVB

It has been established that EVB method requires a reference state to function, the $\Delta G^{\ddagger}$ *and* ΔG^{0} values obtained from DFT calculation were used to parametrize the reference state of the EVB. In this study, EVB simulation was firstly done to reproduce activation and reaction free energies for uncatalyzed and catalyzed conversion of chorismate to prephenate. Analyzing the EVB simulated energy files further also give other thermodynamic parameters of the reaction.

Preparing Topology

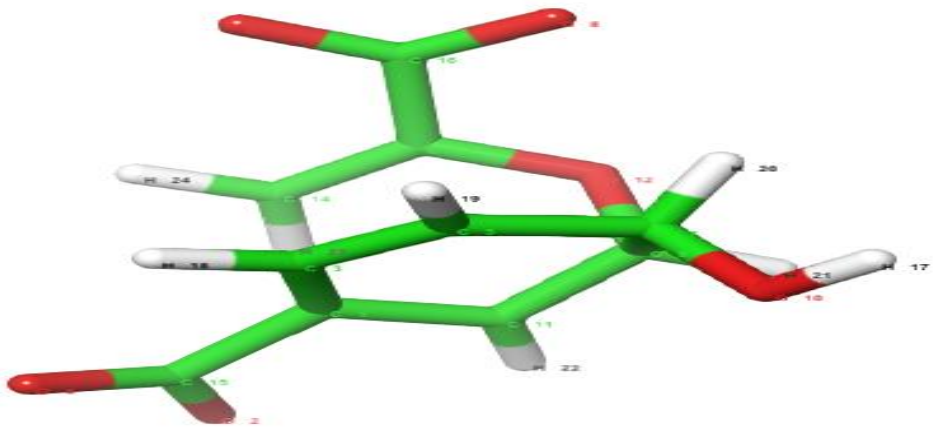

Figure 8: Structure of optimized chorismate

Q program automatically assign Force field parameters and library files, the structure above was firstly uploaded in the Q graphical user interface Qgui to assign chorismate force field parameters and library files. The generated parameters were set in the relevant section of the Qgui setting. For the structure to be readable by the program (Q-atoms), its topology was prepared during which the molecule was set to simulation sphere of 15Å, the sphere was solvated with water molecules using TIP3P[48] solvation model and the total charge of the residue was confirmed to be -2 and the coordinate of the molecule was also confirmed.

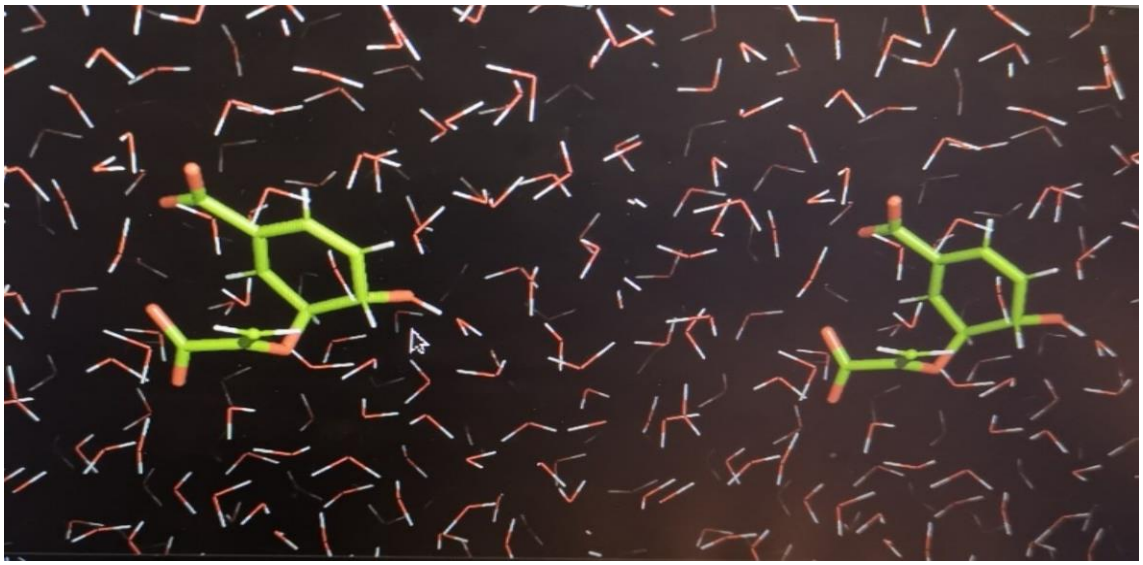

Figure 9: Two EVB states to be defined

EVB Setup for the Uncatalyzed Reaction

The 24 constituent atoms of the chorismate in the water solution were defined, as explained under section 1.2 above, the breaking ether bond which occurs between atom O7 and C5 were defined and the forming C-C bond between C1 and C9 were defined. For a two state EVB, going from one state to another, the ether bond was defined as unbonded in state 2 and the C-C bond was defined as bonded in state 2 as shown below.

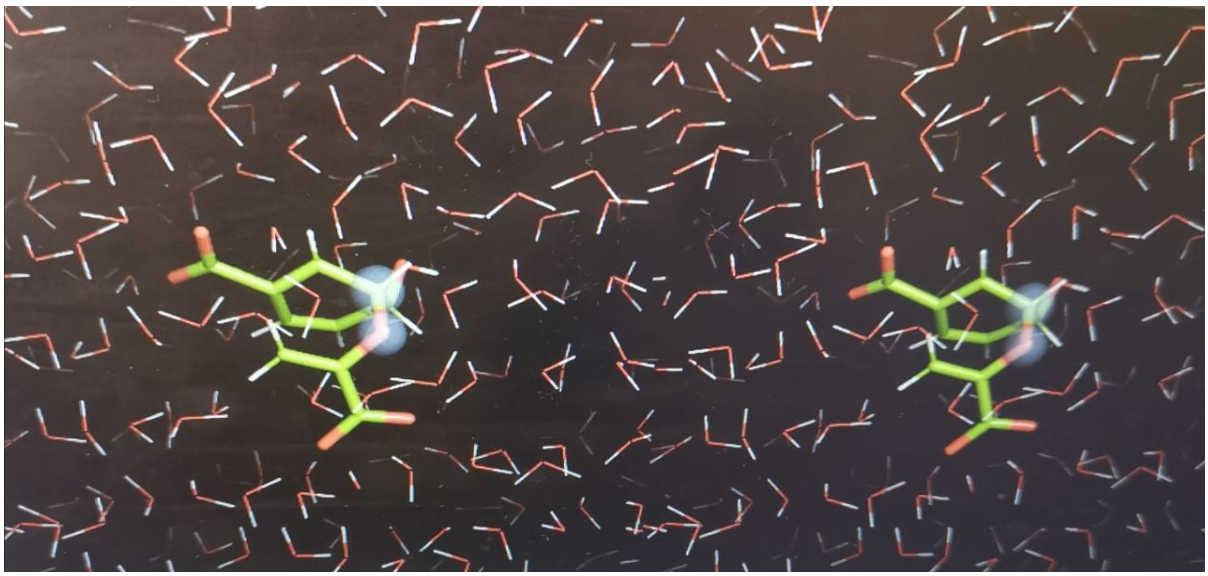

Figure 10: Defining the ether bond between atoms O7 and C5

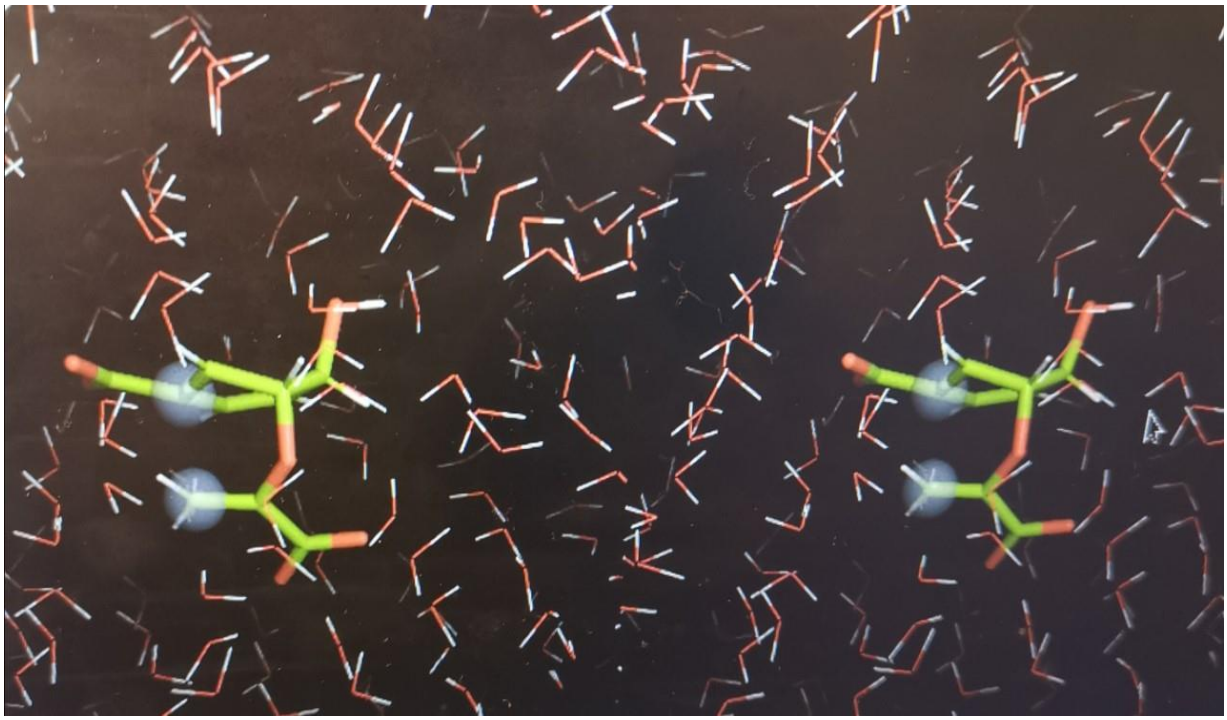

Figure 11: Defining the C-C bond between C1 and C9

Since Q automatically assign force field parameters such as charges, atom types, bonds, angles, Torsions, impropers, couplings etc., all these parameters were properly checked, and the missing ones were assigned manually.

EVB Simulation Details

The simulation was done setting the system temperatures 298K, total simulation time was set at 0.51ns with a total step of 510000, temperature bath coupling was set to 100fs and SHAKE[49] was used to constrain the solvent. The cut off for non-bonded interaction between solute-solute, solute-solvent and solvent-solvent was set at 10 Å each and polarization restraint was used. Sequence restraint with a force of 0.5A was added to the 24 Q-atoms in each of the two states represented with "i" and "j". The output recording interval was set at 10 25 for non-bonded list with a total energy file of 10 summarized in 5 steps and trajectory of 1000. The simulation was carried out in replica of 10 for each of the temperature set.

Analyzing EVB Simulation

The energy files from the simulation were analyzed firstly by setting the reference reaction followed by calibrating the EVB Hamiltonian. This was done by setting the target $\Delta G^{\ddagger}$ *and* ΔG^{0} values to be the values obtained from DFT calculation. For a two state EVB, Qgui automatically calibrates α and H_{ij}, after the reference reaction has been parametrized, then the α and H_{ij} kept adjusting till the target $\Delta G^{\ddagger}$ *and* ΔG^{0} values are reproduced. This was done for each replica of the 10 runs. The average values of α and H_{ij} were used further to calculate other thermodynamic parameters at different temperatures of the reaction system.

Arrhenius Simulation

Arrhenius simulation was carried out on the uncatalyzed conversion of chorismate to prephenate at five different temperatures which are 288K, 283K, 298K, 303K and 308K. this was done also to construct Arrhenius plot that analyze the effects of the five different temperatures on the free energy of activation. The simulation was performed in replica of 5 runs for each temperature.

Analyzing Thermodynamic Parameters

The average values of α and H_{ij} obtained during calibration of EVB Hamiltonian was used to analyze the thermodynamic parameters of the reaction at the five different temperature sets. From the analysis of the energy files, Arrhenius plot, Regression parameters which give the activation enthalpy and entropy and values of model free energies against the computed values and average value of the activation free energy were all obtained.

2.3 Protein Preparation and Docking

The crystal structure of chorismate mutase used in this project was obtained from the protein data bank[50] with PDB ID: 1COM (CM from Bacillus Subtilis). The structure contains 12 chains

of 4 homotrimers complex with prephenate in each homotrimer. Using a computational modelling program called PYMOL[45], the crystal structure was made simpler by removing 3 homotrimer chains with their prephenate complexes. With the use of PyMOL, the remaining homotrimer structure was aligned with another homotrimer functional structure of chorismate mutase (PDB ID: 1DBF)[50], this is to compare the structure with a functional protein of the same molecule.

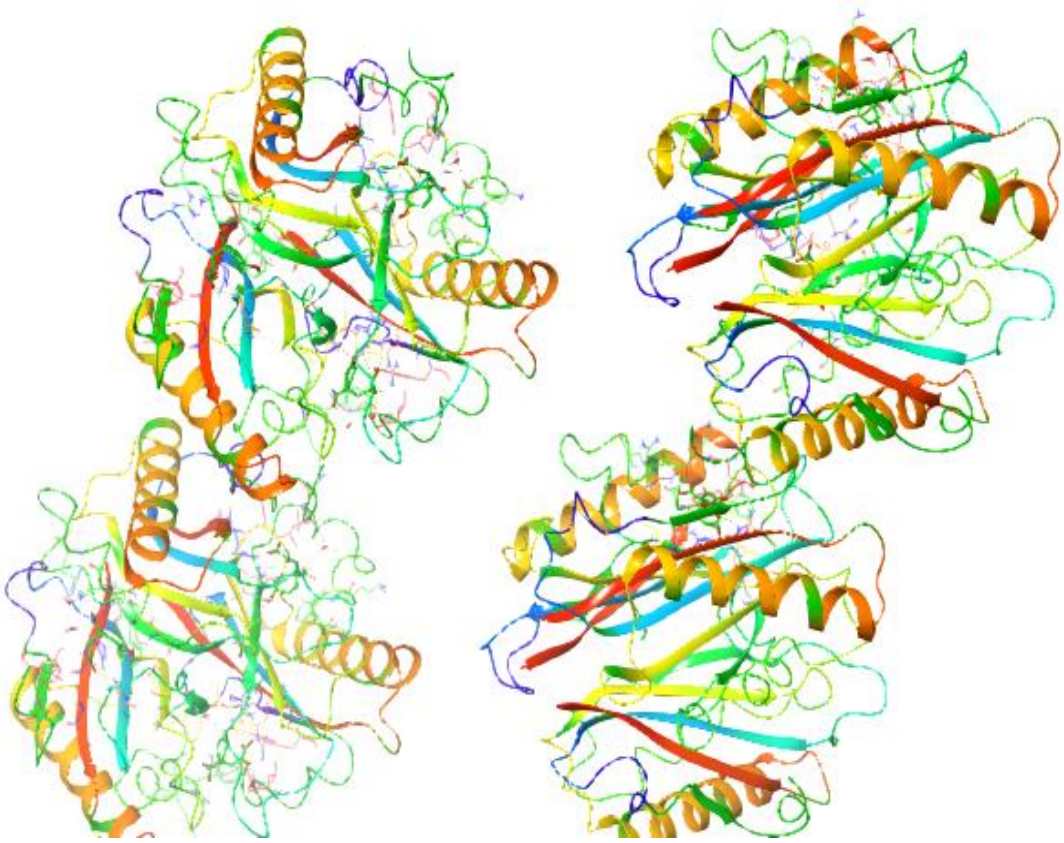

Figure 12: Chorismate mutase from Bacillus subtilis (PDB ID: 1COM)[50]

Starting from the trimmed structure of 1COM, the protein was prepared using Protein Preparation Wizard in MAESTRO[51]. The protein was preprocessed by adding hydrogen atoms, assigning bond orders and removing water molecules beyond 5Å beyond prephenate molecule in the protein. The preprocessed protein was then optimized using JAGUAR[52] with B3LYP-D3/631G**.

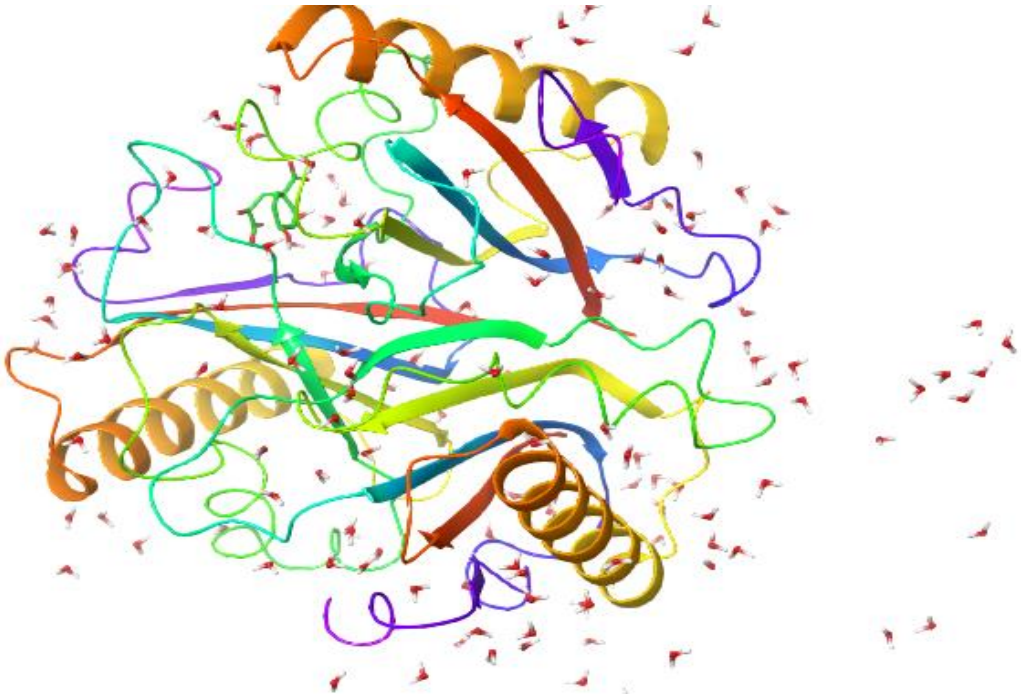

Figure 13: Structure of the trimmed chorismate mutase

Docking of the DFT calculated chorismate (RS) to the CM was carried out using Glide Application[38], this was achieved by initially generating a receptor grid in the CM. The generated grid truncated the prephenate ligand in the CM thereby making the CM ready to accept the RS as ligand molecule. XP precision[53] was employed in performing the ligand docking, the Van der Waals radii scaling factor was at 0.8 with partial cut-off of 0.15, ligand sampling was set to flexible with sample ring conformations, bias sampling of torsions was made for all predefined functional groups, addition of Epik state penalties[54] to docking score and post-docking minimization was also performed. The docked protein was used further for EVB simulation to obtain free energies for catalyzed conversion of chorismate to prephenate in enzymatic reaction.

2.3.1 **Catalyzed Reaction with EVB**

EVB simulation was carried on chorismate mutase catalyzed conversion of chorismate to prephenate. The calibrated EVB Hamiltonian (α and H_{ij}) obtained from the uncatalyzed reaction was then used to parametrize the enzyme catalyzed reaction, this was done in order to test for catalysis in the enzyme.

Preparing Topology and EVB Setup for Catalyzed Reaction

Starting from the docked protein as explained from section 2.3 above, the protein was uploaded in the Q graphical user interface Qgui to assign force field parameters and library files. The generated parameters were set in the relevant section of the Qgui setting. For the structure to be readable by the program (Q-atoms), its topology was prepared during which only the constituting atoms of the protein were selected without water molecules and set to simulation sphere of 15Å, the sphere was solvated with water molecules using TIP3P solvation model. For the EVB setup, all procedures carried out under section 2.2.1 above on EVB setup for the uncatalyzed reaction were followed without adding restraints on the substrate. The EVB simulation details were the same for the uncatalyzed reaction except for running the catalyzed reaction at a single temperature of 298K.

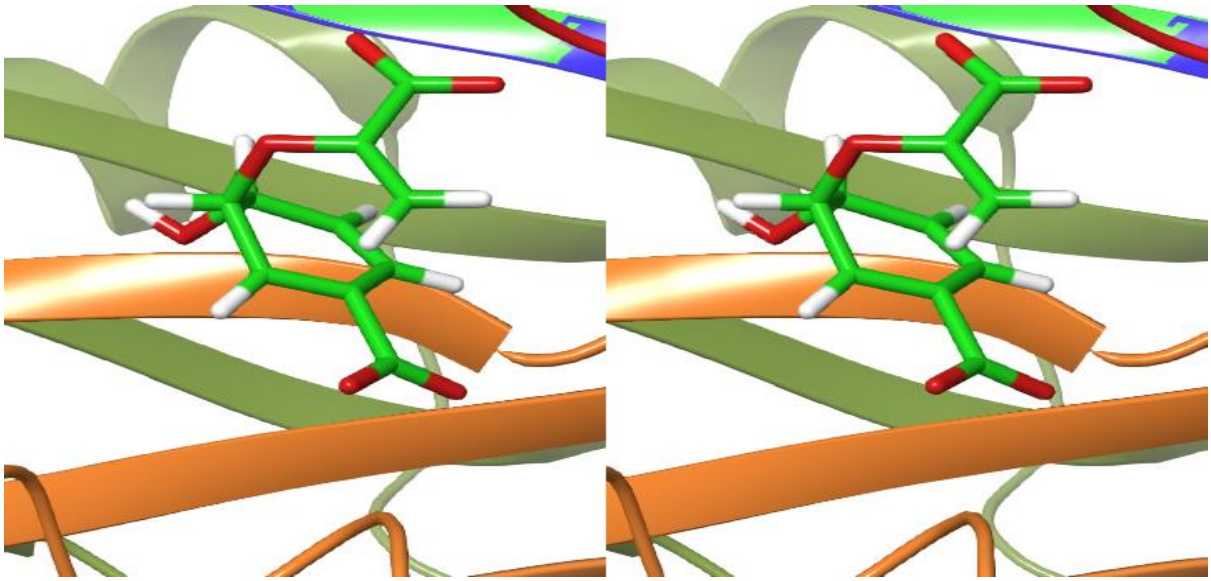

Figure 14: Defining the two EVB states for the catalyzed reaction.

3 RESULTS AND DISCUSSION

In any research work, modelling and simulation play important roles since the validation of the experiments has to do with either numerical or analytical approaches. But in the reality analytical method is quite a difficult task compared to numerical analysis. Hence, numerical investigation become popular in recent times involving software packages which will have definite and set of required tools that can solve the equations and predict the solution. In any chemical reaction, each state has a potential free energy, an activation complex is formed going from one state to another, this is the transition state of the reaction and this is the point where the potential energy is highest. Enzymes catalyzed reactions make biological reaction to occur faster by lowering the activation energy. In this project, the first thing was to predict an approximate conversion of chorismate (RS) to prephenate (PS) in which the free energies of activation and reaction for the conversion then served as reference state for calibrating EVB Hamiltonian to calculate the activation and reaction free energies for uncatalyzed conversion of chorismate to prephenate in water solution and the calibrated Hamiltonian was used to test for catalysis in chorismate mutase. The DFT calculation was firstly employed to predict the activation and reaction free energies for the uncatalyzed conversion of chorismate to prephenate. A protein preparation was done to obtain a refined crystal structure of the chorismate mutase which served as the model enzyme in this project. The EVB simulations were run at five different temperatures giving the calculated activation and reaction free energies and also the activation enthalpy and entropy were also obtained for the reaction.

3.1 DFT Geometry Optimization

Following all the procedures as explained in section 2.1.1 above, a structure of chorismate was built as the starting molecule and a TS guess carried out on the molecule. A geometry convergence of the TS guess was attained using the DFT method of computational modelling which was able to predict the RS and PS. Then the achieved RS and PS were also geometrically optimized. The following figures show geometrically optimized structures obtained from B3LYP/d3 functional.

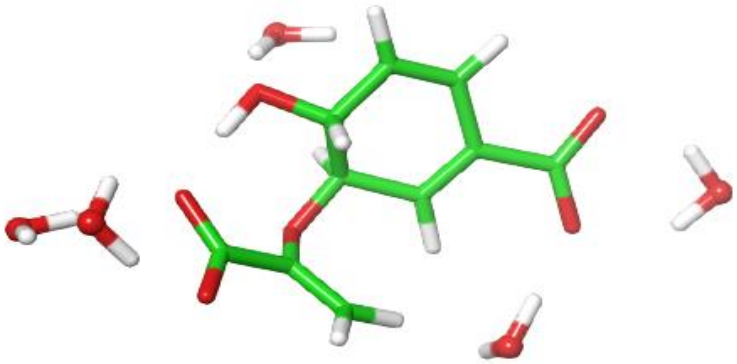

Figure 15: DFT geometry optimized reactant state (chorismate)

Figure 16: DFT geometry optimized transition state

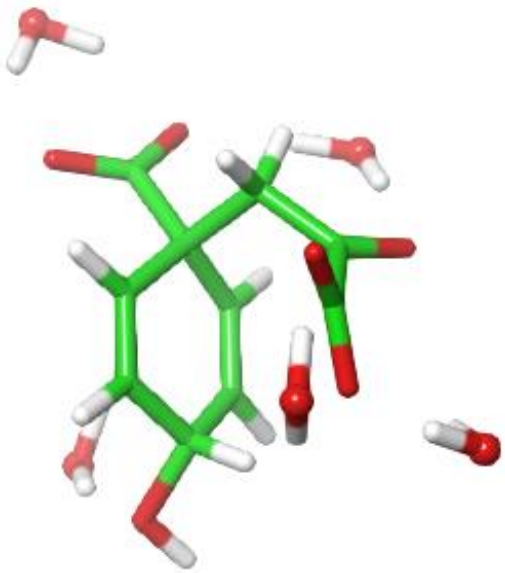

Figure 17: DFT Geometry optimized product state (prephenate)

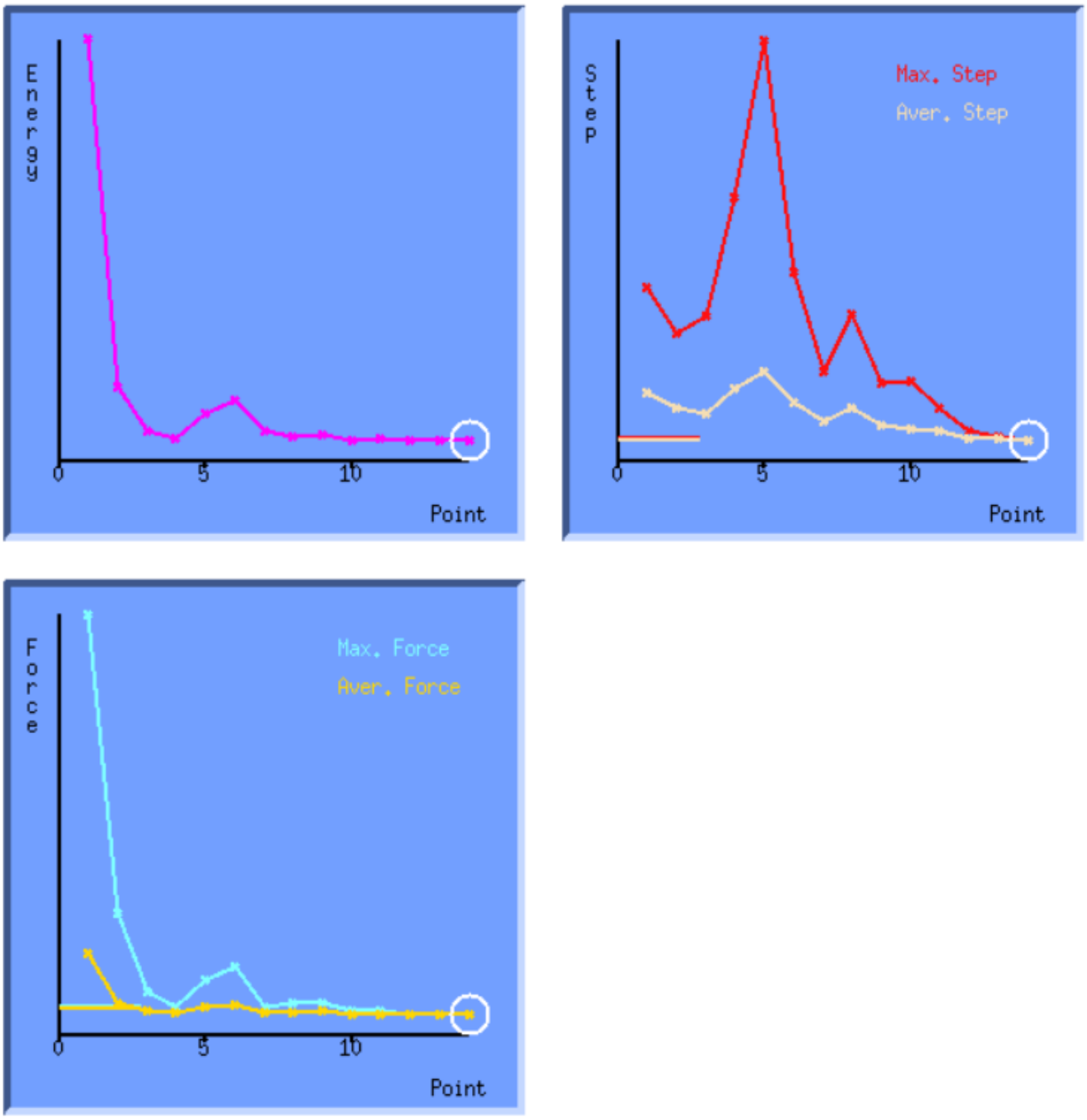

Figure 18: Geometry convergence of the transition state.

TS Vibrational Frequency

The frequency calculation on the TS gave a total of 111 frequencies and attained one true vibrational frequency which is -293.53cm^{-1} with intensity of 14.52. Prediction to RS and PS was started from this vibrational frequency.

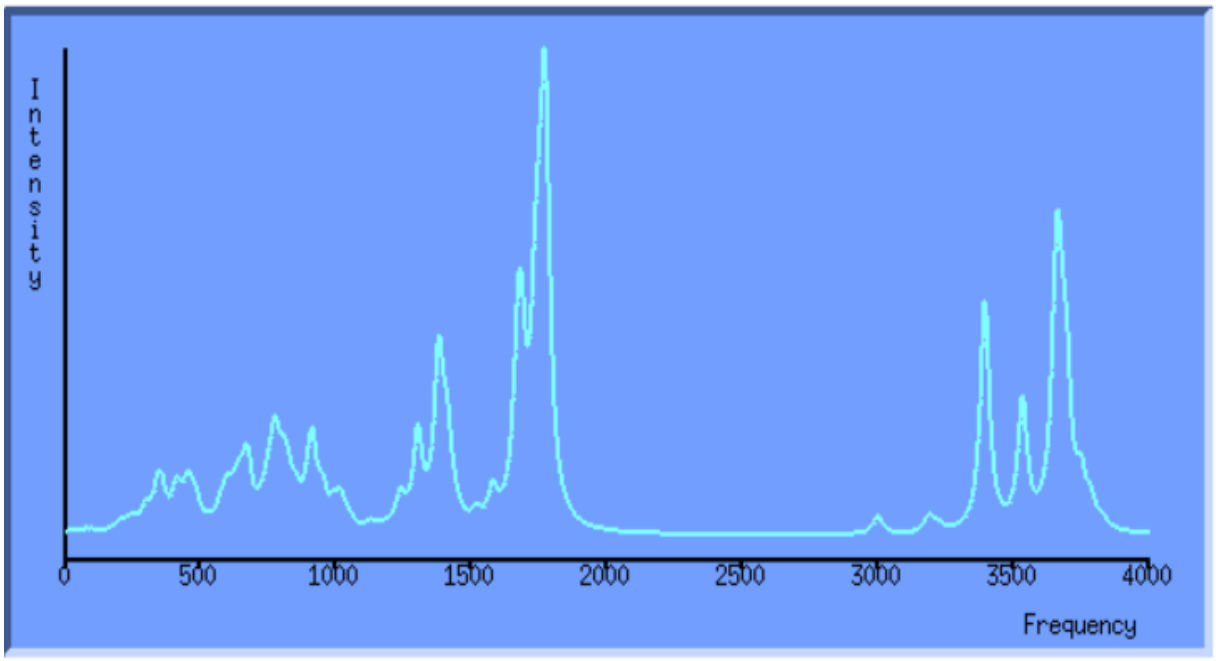

Figure 19: Spectrum from the frequency calculation of the TS

3.2 Activation and Reaction Energies from DFT Calculation

After obtaining the electronic energies for each of the RS, TS and PS in gas phase, from big basis set, from solvation and correction to Gibb's free energy, the following equations are used to obtain the $\Delta G^{\ddagger}$ *and* ΔG^{0} values.

Correction to energy from big basis set:

$$\Delta E^{\ddagger} = E_{bbs}^{TS} - E_{gas}^{RS}$$

$$\Delta E^{0} = E_{bbs}^{PS} - E_{gas}^{RS}$$

Solvation energies of the RS, TS and PS:

$$\Delta E_{sol}^{\mathrm{RS}} = E_{sol}^{RS} - E_{gas}^{RS}$$

$$\Delta E_{sol}^{\mathrm{TS}} = E_{sol}^{TS} - E_{gas}^{TS}$$

$$\Delta E_{sol}^{\mathrm{PS}} = E_{sol}^{PS} - E_{gas}^{PS}$$

Electronic energy from solvation:

$$\Delta\Delta E_{sol}^{\ddagger} = \Delta E_{sol}^{TS} - \Delta E_{gas}^{RS}$$
$$\Delta\Delta E_{bbs}^{0} = \Delta E_{bbs}^{PS} - \Delta E_{gas}^{RS}$$

Correction to energy from solvation:

$$\Delta E_{sol}^{\ddagger} = E_{sol}^{TS} - E_{sol}^{RS}$$

$$\Delta E_{sol}^{0} = E_{sol}^{PS} - E_{sol}^{RS}$$

Thermal correction to Gibb's free energy:

$$\partial G_{ZPE}^{\ddagger} = \partial G_{ZPE}^{TS} - \partial G_{ZPE}^{RS}$$

$$\partial G_{ZPE}^{0} = \partial G_{ZPE}^{PS} - \partial G_{ZPE}^{RS}$$

Correction to ZPE,

$$\Delta\partial G_{ZPE}^{\ddagger} = \partial G_{ZPE}^{TS} - \partial G_{ZPE}^{RS}$$

$$\Delta\partial G_{ZPE}^{0} = \partial G_{ZPE}^{PS} - \partial G_{ZPE}^{RS}$$

From all the equations above, the activation and reaction free energies equations are given below:

$$\Delta G^{\ddagger} = 627.5 * \Delta \boldsymbol{E}_{\boldsymbol{bbs}}^{\ddagger} + \Delta\Delta \boldsymbol{E}_{\boldsymbol{sol}}^{\ddagger} + \Delta \boldsymbol{\partial G}_{\boldsymbol{ZPE}}^{\ddagger}$$
$$\Delta G^{0} = 627.5 * \Delta \boldsymbol{E}_{\boldsymbol{bbs}}^{0} + \Delta\Delta \boldsymbol{E}_{\boldsymbol{sol}}^{0} + \Delta \boldsymbol{\partial G}_{\boldsymbol{ZPE}}^{0}$$

Where 627.5 is a conversion factor from Hatree to kcal/mol since all the energy values of the application used are given in atomic unit.

Table 1: Results from B3LYP functional of the DFT calculation with CPCM solvent model. All the values are given in Hatree except values from entropy and ΔS which are given in cal/mol/kelvin and also $\Delta G^{\ddagger}$ and ΔG^{0} values given in kcal/mol.

	RS	**TS (‡)**	**PS (^{0})**
E (gas)	-1219.504	-1219.449	-1219.508
ZPE σdg (Hatree/particle)	0.235746	0.234369	0.234881
SMD E(solv)	-1219.720	-1219.675	-1219.733
E(bbs)	-1219.979	-1219.928	-1219.989
ΔE (solv) = E(solv)-E(gas)	-0.215	-0.226	-0.225
ΔΔE(solv)	Values obtained in relative to the Reactant state	-0.0105	-0.001
ΔE (bbs)		0.051	-0.001
ΔE (σΔG)		-0.0013	-0.000865
ΔG (kcal/mol)		24.4	-12.6

From the table 3.1 above, $\Delta G^{\ddagger}$ *and* ΔG^{0} are 24.4 kcal/mol and -12.6 kcal/mol respectively.

Table 2: DFT calculated $\Delta G^{\ddagger}$ and ΔG^{0} with SMD solvent model compared with Experimental $\Delta G^{\ddagger}$

	$\Delta G^{\ddagger}$ (kcal/mol)	**ΔG^{0} (kcal/mol)**
B3LYP	20.6	-15.8
CAM-B3LYP	30.4	-17.8
PBEPBE	21.3	-19.2
Experimental[6]	24.5	

Table 3: DFT calculated $\Delta G^{\ddagger}$ *and* ΔG^{0} with CPCM solvent model.

	$\Delta G^{\ddagger}$ (kcal/mol)	ΔG^{0} (kcal/mol)
B3LYP	24.4	-12.6
CAM-B3LYP	35.0	-14.1
PBEPBE	25.1	-16.4

DFT is a central technique in computational chemistry and it is sensitive to the functional used, these density functionals give balance between accuracy and speed. Same chemical reactions give different results depending on the functionals used, a suitable result can be obtained in comparison with experimental data or known calculated data of smaller models. In this project, other density functionals such as LSDA, BPV86, M08HX, M06 and B3PW91 were also tried but due to non-convergence of the molecule structure optimization, the functionals were limited to three.

Comparing the experimental $\Delta G^{\ddagger}$ which is 24.5 kcal/mol with the values obtained from each of the DFT functionals above. $\Delta G^{\ddagger}$ and ΔG^{0} from B3LYP functional with CPCM solvent model which are 24.5 and -12.7 kcal/mol respectively have been used further in this project, since the values match the experimental data. These values served as the target free energies for calibrating EVB Hamiltonian.

The difference in the calculated energy barrier for the three functionals show the sensitivity of the barrier to functionals of DFT and how the water models vary in allowing explicit of freedom water. From the results above, the smaller exchange correlation functional B3LYP gives barrier of 20.6 and 24.5 kcal/mol for the two water models used, CAM-B3LYP with long range exchange correlation gives 30.5 and 35.0 kcal/mol and the exchange-free functional PBEPBE gives 21.3 and 25.1 kcal/mol. The DFT predicted reaction free energies for the two solvent models ranges between -12 and -19 kcal/mol.

To know the most reliable DFT functional to use, a projector-based embedding calculation can be employed. The projector-based embedding eliminates the barrier dependence on the density functional used.[55]

3.3 Free Energies of Uncatalyzed Reaction with EVB

3.3.1 Defined EVB states

Figures 3.6 and 3.7 below show the defined state in water from the reactant state (chorismate) to the transition state and finally to the product state (prephenate).

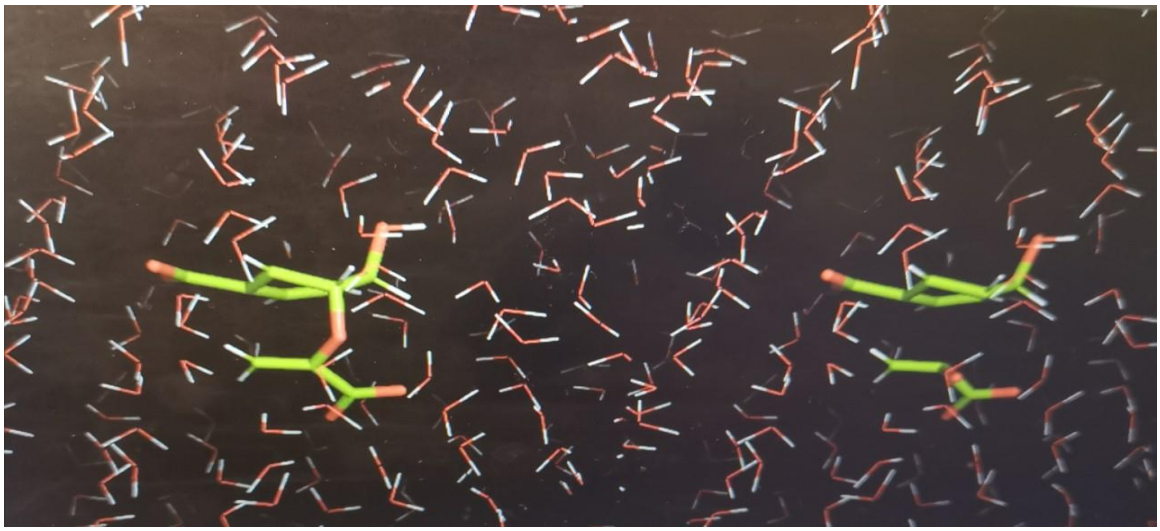

Figure 20: Defined unbounded ether bond in state two (Transition state)

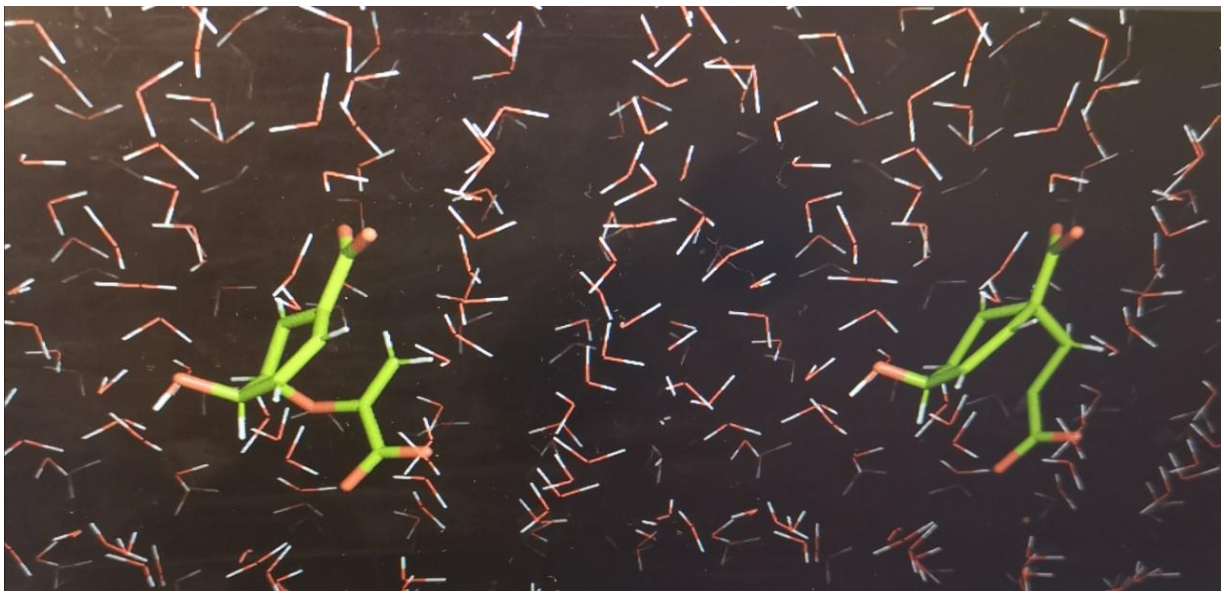

Figure 21: Defined state one (reactant state) and two (product state)

3.3.2 **EVB Reference Reaction Calibration for the Uncatalyzed reaction**

The EVB simulation produced α and H_{ij} values when parametrized with target $\Delta G^{\ddagger}$ *and* ΔG^{0} values obtained from the DFT calculation and kept adjusting until the target free energies were reproduced.

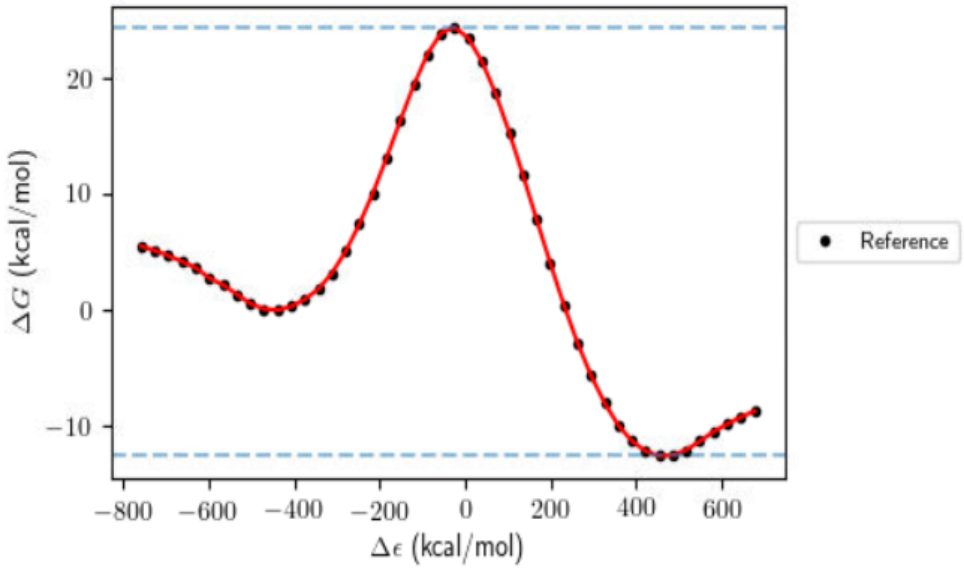

Figure 22: Converged calibration of $\Delta G_{\ddagger}$ towards the target reference activation free energy.

Table 4: α and H_{ij} values for the replica of 10 runs at 298K.

α	H_{ij}
-129.248	87.122
-129.182	87.173
-129.191	87.173
-129.189	87.169
-129.187	87.196
-129.248	87.122
-129.248	87.122
-129.248	87.122
-129.182	87.122
-129.248	87.122
-129.217	87.144

Average

The average values of α and H_{ij} which are -129 and 87 respectively correspond to the EVB calibrated Hamiltonian. These values were then used to parametrize the calculation of the reaction free energy profiles at five different temperatures and also to test for catalysis in chorismate mutase catalyzed reaction.

3.3.3 **EVB Computed Free Energies for Uncatalyzed Reaction**

Table 5: The reproduced values of $\Delta G^{\ddagger}$ and ΔG^{0} for each of the 10 runs at 298K

	$\Delta G_{\ddagger}$	ΔG_0
	24.4	-12.6
	24.4	-12.5
	24.4	-12.6
	24.4	-12.6
	24.4	-12.6
	24.4	-12.6
	24.4	-12.6
	24.4	-12.6
	24.4	-12.5
	24.4	-12.6
Average	24.4	-12.6

At 298K, the EVB computed activation and reaction free energy are 24.4 kcal/mol and -12.6 kcal/mol respectively. The computed activation free energy is in accord with the experimental estimate which is 24.5 kcal/mol for uncatalyzed conversion of chorismate to prephenate in water.

The Arrhenius simulation for the conversion of chorismate to prephenate in water solution was performed at different temperatures of 288K, 293K, 298K, 303K and 308K. The calibrated Hamiltonian values α and H_{ij} were used to parametrize the calculation. The simulation gave the reaction free energy profiles at different temperature. The simulation was able to construct Arrhenius plot that shows the effect of temperature on the activation free energy.

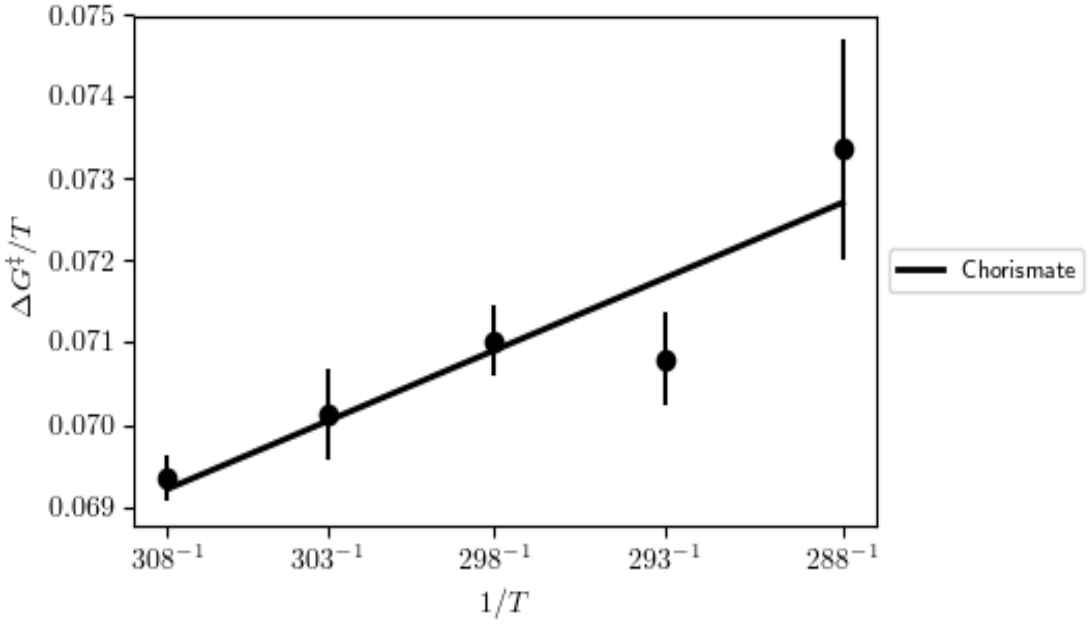

Figure 23: Arrhenius plot showing the plot of the activation reaction rate constant $\Delta G^{\ddagger}/T$ against the reciprocal of each of the reaction temperature.

Table 6: The average free energies at different temperature.

Temperature, K	$\Delta G^{\ddagger}$	+/-	ΔG^{0}	+/-
288	24.6	0.35	-12.9	0.57
293	24.2	0.18	-12.9	0.29
298	24.4	0.18	-12.6	0.26
303	24.7	0.22	-12.5	0.36
308	24.7	0.09	-12.3	0.2

3.3.4 EVB Activation Enthalpy and Entropy

The EVB computed activation enthalpy and entropy are in excellent agreement with the experimental values for uncatalyzed transformation of chorismate to prephenate.

Table 7: From the Arrhenius calculation, the regression parameters gave the activation enthalpy and entropy.

	$\Delta H^{\ddagger}$ (kcal/mol)	$T\Delta S^{\ddagger}$ (kcal/mol) at 298K
	20.5	-4.1
Experimental[6]	20.7	-3.8

3.3.5 EVB Model vs Computed Activation Thermodynamic Parameters

Table 8: The difference between the model and computed activation free energies at different temperatures.

T (K)	**Computed $\Delta G^{\ddagger}$ (kcal/mol)**	**$\Delta H^{\ddagger}$**	**$T\Delta S^{\ddagger}$ (kcal/mol)**	**Model $\Delta G^{\ddagger}$ (kcal/mol)**
288	24.6	20.5	-3.9	24.4
293	24.2	20.5	-4.0	24.5
298	24.4	20.5	-4.1	24.5
303	24.7	20.5	-4.1	24.6
308	24.7	20.5	-4.2	24.7

3.4 Structure of Chorismate Mutase and Docking Score

The modelled and optimized RS served as the substrate molecule. The enzyme, chorismate mutase structure was retrieved from Protein Data bank (PDB) and was refined, then the substrate was docked with it. The superposition of the prepared structure of the CM used on another structure of a functional protein gave RMSD score of 0.376, this is to ascertain the functionality of the structure of the prepared CM.

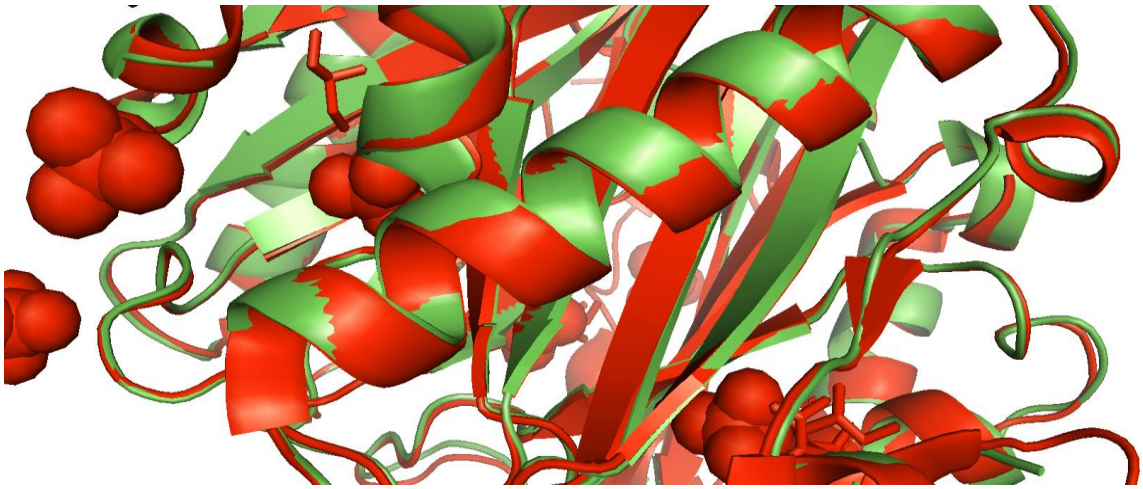

Figure 24: structure of superimposed CM. Structure in green is the prepared CM, red is the reference functional CM.

The RS was docked with the prepared CM as shown in the figure 3.9 below with a docking score of -7.4 kcal/mol.

Figure 25: structure of the docked CM with chorismate

3.5 Test for Catalysis with EVB

The energy files from the EVB simulation for the catalyzed reaction were parametrized with the calibrated EVB Hamiltonian (α and H_{ij}**)** from the uncatalyzed conversion of chorismate to prephenate. This was done to test for catalysis in chorismate as the calibrated Hamiltonian from the uncatalyzed reaction served as reference reaction in water. The value of 18.2 kcal/mol was obtained as the activation free energy at 298K. This is a good value compared to experimental value of 15.4 kcal/mol for BsCM and 17.2 kcal/mol for EcCM.[6]

4 CONCLUSION

Clear understanding of the catalytic effect of enzymes has to do with thermodynamic parameters obtained from that particular enzymatic reaction. These parameters are the activation free energy that must be attained for conversion of a RS to PS and then the reaction free energy which is the free energy of the PS obtained in relative to the RS. Moreover, enzymatic operation is linked with lowering the activation energy of a biological reaction, this makes an enzyme catalyzed reaction to be independent of reaction free energy but solely dependent on activation free energy.

The activation and reaction free energies have been reproduced for uncatalyzed and catalyzed conversion of chorismate to prephenate at five different temperatures, also $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ of this reaction have been obtained. With the employment of both DFT and EVB method of simulation, it is shown that for a reaction lacking experimental data, DFT can be used to predict the activation and reaction free energies that will serve as reference state for calibrating EVB Hamiltonian. The EVB method can be very suitable for predicting free energies in enzymatic reactions for it to have reproduced $\Delta \mathbf{G}^{\ddagger}$ of 24.4 kcal/mol at 298K with a reference state $\Delta \mathbf{G}^{\ddagger}$ of 24.4 kcal/mol for uncatalyzed conversion of chorismate to prephenate. From the experimental data, $\Delta \mathbf{G}^{\ddagger}$ of uncatalyzed conversion of chorismate to prephenate is 24.5 kcal/mol Comparing these values, the EVB method has been able to reproduce the reaction free surface of the transforming molecules. The activation barrier is reduced to the value of 18.2 kcal/mol from the EVB simulation of the catalyzed reaction, which is in accord with experimental value, the catalytic effect of the enzyme chorismate mutase has been verified.

The EVB simulation in this project has also been able to produce respectively activation enthalpy and entropy of 20.4 kcal/mol and -4.1 kcal/mol at 298k compared to experimental values of 24.5 kcal/mol and -3.8 kcal/mol at 298k. Based on these results, it can be concluded that EVB is a powerful method to generate all the thermodynamic properties necessary for transforming a reaction from one state to another, this makes the method to reckon with for calculating free energy profiles for complex biomolecules.

Future Work

The overall aim of this project is to initiate the development of a small database with calibrated EVB Hamiltonians for enzymes catalyzed reactions. Since EVB shortfall is reference reaction and to obtain this, DFT can be employed for reactions lacking experimental data. Further work can be done on getting the most reliable functionals to use for DFT calculations on other molecules with similar mechanism of action. For this project, several functionals have been tried and further work on this can lead to deeper understanding on sensitivity of the functionals to energy barrier.

For deeper understanding of reactions in solution and in enzymes, further study on this project might be to carry out studies on other enzymes with similar mechanism of action and to check for how conformational changes in the active site residue of the enzyme as it catalyzes conversion from chorismate to prephenate affect the free energies of conversion and to study the effect that arises from change in binding free energies between the enzyme and its RS, PS and TS substrates. This can give more insight into the transition state stability of chorismate mutase.

5 REFERENCES

(1) Petsko, G.; Ringe, D. *Protein Structure and Function*; Primers in Biology; Oxford University Press: Oxford, New York, 2008.

(2) Cunningham, M. A.; Ho, L. L.; Nguyen, D. T.; Gillilan, R. E.; Bash, P. A. Simulation of the Enzyme Reaction Mechanism of Malate Dehydrogenase. *Biochemistry* **1997**, *36* (16), 4800–4816. https://doi.org/10.1021/bi962734n.

(3) Tahergorabi, R.; Hosseini, S. V. Chapter 2 - Proteins, Peptides, and Amino Acids. In *Nutraceutical and Functional Food Components*; Galanakis, C. M., Ed.; Academic Press, 2017; pp 15–38. https://doi.org/10.1016/B978-0-12-805257-0.00002-8.

(4) Cook, P. F. Steady-State Enzyme Kinetics. In *Encyclopedia of Biophysics*; Roberts, G. C. K., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2013; pp 2466–2470. https://doi.org/10.1007/978-3-642-16712-6_55.

(5) Roskoski, R. Enzyme Structure and Function XPharm: 2007, Pages 1-7. *The Comprehensive Pharmacology Reference* **2007**, 1–7.

(6) Hiersemann, M.; Nubbemeyer, U. *The Claisen Rearrangement: Methods and Applications*; John Wiley & Sons, 2007.

(7) Kim, S.-K.; Reddy, S. K.; Nelson, B. C.; Vasquez, G. B.; Davis, A.; Howard, A. J.; Patterson, S.; Gilliland, G. L.; Ladner, J. E.; Reddy, P. T. Biochemical and Structural Characterization of the Secreted Chorismate Mutase (Rv1885c) from Mycobacterium Tuberculosis H37Rv: An *AroQ Enzyme Not Regulated by the Aromatic Amino Acids. *Journal of Bacteriology* **2006**, *188* (24), 8638–8648. https://doi.org/10.1128/JB.00441-06.

(8) Sasso, S.; Okvist, M.; Roderer, K.; Gamper, M.; Codoni, G.; Krengel, U.; Kast, P. Structure and Function of a Complex between Chorismate Mutase and DAHP Synthase: Efficiency Boost for the Junior Partner. *EMBO J.* **2009**, *28* (14), 2128–2142. https://doi.org/10.1038/emboj.2009.165.

(9) Lyne PD, Mulholland AJ, Richards WG. 1995. Insights into Chorismate Mutase Catalysis from a Combined QM/MM Simulation of the Enzyme Reaction. J. Am. Chem. Soc. 117:11345–50.

(10) Higashi, M.; Hayashi, S.; Kato, S. Transition State Determination of Enzyme Reaction on Free Energy Surface: Application to Chorismate Mutase. *Chemical Physics Letters* **2007**, *437* (4), 293–297. https://doi.org/10.1016/j.cplett.2007.02.044.

(11) Guhe, V.; Soni, B.; Ingale, P.; Singh, S. Autophagy Proteins and Its Homeostasis in Cellular Environment. In *Advances in Protein Chemistry and Structural Biology*; Elsevier, 2020; p S1876162319300975. https://doi.org/10.1016/bs.apcsb.2019.12.002.

(12) *Perrot, Pierre (1998). A to Z of Thermodynamics. Oxford University Press. ISBN 0-19-856552-6.*

(13) Brachman, M. K. (1954). “Fermi Level, Chemical Potential, and Gibbs Free Energy”. The Journal of Chemical Physics. 22 (6): 1152. Bibcode:1954JChPh..22.1152B. Doi:10.1063/1.1740312.

(14) *CRC Handbook of Chemistry and Physics, 2009, Pp. 5-4–5-42, 90th Ed.*

(15) Ranaghan, K. E.; Lonsdale, R. QM/MM Methods. In *Encyclopedia of Biophysics*; Roberts, G. C. K., Ed.; Springer: Berlin, Heidelberg, 2013; pp 2154–2156. https://doi.org/10.1007/978-3-642-16712-6_275.

(16) Kamerlin, S. C. L.; Warshel, A. The Empirical Valence Bond Model: Theory and Applications. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1* (1), 30–45. https://doi.org/10.1002/wcms.10.

(17) Jensen, F. *Introduction to Computational Chemistry*, 2nd ed.; John Wiley & Sons: Chichester, England ; Hoboken, NJ, 2007.
(18) González, M. A. Force Fields and Molecular Dynamics Simulations. *JDN* **2011**, *12*, 169–200. https://doi.org/10.1051/sfn/201112009.
(19) Atkins, P. W.; Friedman, R. *Molecular Quantum Mechanics*, 4th ed.; Oxford University Press: New York, 2005.
(20) Odegard, G. M.; Gates, T. S.; Wise, K. E. Constitutive Modeling of Nanotube-Reinforced Polymer Composite Systems https://www.cs.odu.edu/~mln/ltrs-pdfs/NASA-aiaa-2002-1427.pdf.
(21) *Theory and Applications of the Empirical Valence Bond Approach: From Physical Chemistry to Chemical Biology*; Duarte, F., Kamerlin, S. C. L., Eds.; John Wiley & Sons, Ltd: Chichester, UK, 2017. https://doi.org/10.1002/9781119245544.
(22) Vlachakis, D.; Bencurova, E.; Papangelopoulos, N.; Kossida, S. Chapter Seven - Current State-of-the-Art Molecular Dynamics Methods and Applications. In *Advances in Protein Chemistry and Structural Biology*; Donev, R., Ed.; Academic Press, 2014; Vol. 94, pp 269–313. https://doi.org/10.1016/B978-0-12-800168-4.00007-X.
(23) Berhanu, W. M.; Masunov, A. E. The Atomic Level Interaction of Polyphenols with the Aβ Oligomer Aggregate, A Molecular Dynamic Guidance for Rational Drug Design. In *Polyphenols in Human Health and Disease*; Elsevier, 2014; pp 59–70. https://doi.org/10.1016/B978-0-12-398456-2.00006-2.
(24) Zheng, L.; Alhossary, A. A.; Kwoh, C.-K.; Mu, Y. Molecular Dynamics and Simulation. In *Encyclopedia of Bioinformatics and Computational Biology*; Ranganathan, S., Gribskov, M., Nakai, K., Schönbach, C., Eds.; Academic Press: Oxford, 2019; pp 550–566. https://doi.org/10.1016/B978-0-12-809633-8.20284-7.
(25) Sherrill, C. D. The Born-Oppenheimer Approximation. *http://vergil.chemistry.gatech.edu/notes/bo/bo.pdf* 7.
(26) Harrison, N. M. An Introduction to Density Functional Theory. *Nato Science Series Sub Series III Computer and Systems Sciences* **2003**, *187*, 45–70.
(27) Ravindran, P. Introduction to Density Functional Theory Part-2. **2015**, 89.
(28) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. J Chem Phys 98, 5648-5652, Doi:10.1063/1.464913 (1993).
(29) *Siegbahn, P. E. M. & Himo, F. The Quantum Chemical Cluster Approach for Modeling Enzyme Reactions. Wiley Interdisciplinary Reviews: Computational Molecular Science 1, 323-336, Doi:10.1002/Wcms.13 (2011).*
(30) *Himo, F. Recent Trends in Quantum Chemical Modeling of Enzymatic Reactions. J Am Chem Soc 139, 6780-6786, Doi:10.1021/Jacs.7b02671 (2017).*
(31) Salek, P.; Helgaker, T.; Vahtras, O.; Ågren, H.; Jonsson ⊥, D.; Gauss, J. A Comparison of Density-Functional-Theory and Coupled-Cluster Frequency-Dependent Polarizabilities and Hyperpolarizabilities. *Molecular Physics* **2005**, *103* (2–3), 439–450. https://doi.org/10.1080/00268970412331319254.
(32) Andrew R. Leach. Molecular Modelling. Principles and Applications. Vol 5. Essex: Prentice Hall., Ian. 2001.
(33) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *The Journal of Chemical Physics* **1954**, *22* (8), 1420–1426. https://doi.org/10.1063/1.1740409.
(34) Abel, R.; Wang, L.; Harder, E. D.; Berne, B. J.; Friesner, R. A. Advancing Drug Discovery through Enhanced Free Energy Calculations. *Accounts of Chemical Research* **2017**, *50* (7), 1625–1632. https://doi.org/10.1021/acs.accounts.7b00083.

(35) Åqvist J & Warshel A (1993) Simulation of Enzyme-Reactions Using Valence-Bond Force-Fields and Other Hybrid Quantum-Classical Approaches. Chem Rev 93(7):2523-2544.
(36) Warshel, A.; Weiss, R. M. An Empirical Valence Bond Approach for Comparing Reactions in Solutions and in Enzymes. *J. Am. Chem. Soc.* **1980**, *102* (20), 6218–6226. https://doi.org/10.1021/ja00540a008.
(37) Warshel A, Parson WW. 2001. Dynamics of Biochemical and Biophysical Reactions: Insight from Computer Simluations. Q. Rev. Biophys. 34:563–679.
(38) Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy | Journal of Medicinal Chemistry https://pubs.acs.org/doi/10.1021/jm0306430 (accessed May 18, 2020).
(39) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat Rev Drug Discov* **2004**, *3* (11), 935–949. https://doi.org/10.1038/nrd1549.
(40) *Frisch, M. J. et al. (Wallingford CT, 2009).*
(41) MOLDEN a visualization program of molecular and electronic structure http://cheminf.cmbi.ru.nl/molden/ (accessed May 21, 2020).
(42) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113* (18), 6378–6396. https://doi.org/10.1021/jp810292n.
(43) Marelius, J.; Kolmodin, K.; Feierberg, I.; Aqvist, J. Q: A Molecular Dynamics Program for Free Energy Calculations and Empirical Valence Bond Simulations in Biomolecular Systems. *J. Mol. Graph. Model.* **1998**, *16* (4–6), 213–225, 261. https://doi.org/10.1016/s1093-3263(98)80006-5.
(44) Manual for Molecular Dynamics Simulation Package Q. Http://Xray.Bmc.Uu.Se/~aqwww/q/Documents/Qman5.Pdf.
(45) The PyMOL Molecular Graphics System, Version 2.3.3 Schrödinger, LLC.
(46) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12* (1), 281–296. https://doi.org/10.1021/acs.jctc.5b00864.
(47) Isaksen, G. V.; Andberg, T. A. H.; Åqvist, J.; Brandsdal, B. O. Qgui: A High-Throughput Interface for Automated Setup and Analysis of Free Energy Calculations and Empirical Valence Bond Simulations in Biological Systems. *J. Mol. Graph. Model.* **2015**, *60*, 15–23. https://doi.org/10.1016/j.jmgm.2015.05.007.
(48) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [Optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110* (6), 1657–1666. https://doi.org/10.1021/ja00214a001.
(49) *Ryckaert, J-P; Ciccotti G; Berendsen HJC (1977). "Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes". Journal of Computational Physics. 23 (3): 327–341. Doi:10.1016/0021-9991(77)90098-5.*
(50) Protein Data Bank, R. P. D. RCSB PDB: Homepage https://www.rcsb.org/ (accessed Nov 7, 2019).
(51) Schrödinger Release 2020-2: Maestro, Schrödinger, LLC, New York, NY, 2020.
(52) Schrödinger Release 2020-2: MS Jaguar, Schrödinger, LLC, New York, NY, 2020.

(53) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. *J. Med. Chem.* **2006**, *49* (21), 6177–6196. https://doi.org/10.1021/jm051256o.

(54) Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. Epik: A Software Program for PK a Prediction and Protonation State Generation for Drug-like Molecules. *J Comput Aided Mol Des* **2007**, *21* (12), 681–691. https://doi.org/10.1007/s10822-007-9133-z.

(55) Ranaghan, K. E.; Shchepanovska, D.; Bennie, S. J.; Lawan, N.; Macrae, S. J.; Zurek, J.; Manby, F. R.; Mulholland, A. J. Projector-Based Embedding Eliminates Density Functional Dependence for QM/MM Calculations of Reactions in Enzymes and Solution. *J. Chem. Inf. Model.* **2019**, *59* (5), 2063–2078. https://doi.org/10.1021/acs.jcim.8b00940.